14TH INTERNATIONAL SYMPOSIUM ON PROCESS SYSTEMS ENGINEERING (PSE 2021+)

Edited by
YOSHIYUKI YAMASHITA
MANABU KANO

COMPUTER-AIDED CHEMICAL ENGINEERING, 49

ELSEVIER

PSE 2021+
KYOTO

14th INTERNATIONAL SYMPOSIUM ON PROCESS SYSTEMS ENGINEERING

VOLUME 1

COMPUTER-AIDED CHEMICAL ENGINEERING, 49

14th INTERNATIONAL SYMPOSIUM ON PROCESS SYSTEMS ENGINEERING

VOLUME 1

Edited by

Yoshiyuki Yamashita
Professor and Chair of Department of Chemical Engineering
Tokyo University of Agriculture and Technology (TUAT), Tokyo, Japan
yama_pse@cc.tuat.ac.jp

Manabu Kano
Professor, Department of Systems Science
Kyoto University, Kyoto, Japan
manabu@human.sys.i.kyoto-u.ac.jp

ELSEVIER

Amsterdam – Boston – Heidelberg – London – New York – Oxford
Paris – San Diego – San Francisco – Singapore – Sydney – Tokyo

Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, USA

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

ISBN (Volume 1): 978-0-443-18724-7
ISBN (Set) : 978-0-323-85159-6
ISSN: 1570-7946

Publisher: Susan Dennis
Acquisition Editor: Anita Koch
Editorial Project Manager: Lena Sparks
Production Project Manager: Paul Prasad Chandramohan
Designer: Greg Harris

Typeset by STRAIVE

Contents

Contributed Papers: Process and Product Design/Synthesis

Contributed Papers: Process Dynamics and Control

Contributed Papers: Supply Chain Management and Logistics

Contributed Papers: Process Intensification

Contributed Papers: Integration of Process Operation and Design/Synthesis

Preface

This book contains papers presented at the 14th International Symposium on Process Systems Engineering (PSE 2021+), held at Kyoto University in Kyoto, Japan, June 19-23, 2022. The PSE series is a triennial conference which has been held since 1982, organized on behalf of the international PSE Executive Committee with representation from countries in Asia-Pacific, Europe, and the Americas. The goal is to create an academic and industrial dialogue, a critical assessment of existing enabling technologies, a discussion on research, education, and industrial needs, and an international forum for new directions, challenges, and opportunities in process systems engineering.

The PSE symposium bring together researchers, educators, and practitioners to discuss the latest developments in the field of Process Systems Engineering (PSE), including applications of methods, algorithms, and tools to solve a wide range of problems as well as provide the venue for discussion of new scientific challenges in our field. This symposium will feature more than 371 presentations including invited plenary and keynote lectures, as well as contributed papers (both oral and poster) encompassing a large number of core and cross-cutting PSE themes.

Among the PSE series, PSE 2021+ is special, because the first PSE symposium was held in Kyoto, Japan, in 1982. For this reason, and in light of the challenges facing our society, we chose the main theme of PSE 2021+ as "PSE for Smart & Sustainable Society: perspectives from the origin."

The PSE themes include:

- Process and Product Design/Synthesis
- Process Dynamics and Control
- Scheduling and Planning
- Supply Chain Management and Logistics
- Process Intensification
- Integration of Process Operations and Design/Synthesis
- Modeling, Analysis, and Simulation
- Optimization Methods and Computational Tools
- Process Monitoring and Safety
- Cyber-Physical Systems and Security
- Machine Learning and Big Data
- Energy, Food and Environmental Systems
- Pharma and Healthcare Systems

This book includes 10 invited papers and extended abstracts as well as 361 contributed papers. All papers have been reviewed by the International Programming Committee (IPC). We are very grateful to the IPC members for their assistance and constructive feedback during the review process. We would also like to thank the Elsevier editorial team, particularly Ms. Lena Sparks and Ms. Anita Koch for their support on this project, which provides an archival and fully indexed record of the conference.

The 14th Symposium on Process Systems Engineering (PSE 2021) was originally planned for July 2021. In view of the situation associated with the COVID-19 pandemic, the National Organizing Committee has approved the recommendation of the Executive Committee to postpone the symposium to June 2022. To clarify the postponement, the symposium is written as "PSE 2021+".

We hope PSE 2021+ will foster constructive interaction among thought leaders from academia, industry, and government and that this book will serve as a useful reference for the latest research in all areas of process systems engineering.

Yoshiyuki Yamashita and Manabu Kano

PSE 2021+ General Chairs

International Program Committee

Conference Chairs

Yoshiyuki Yamashita	Tokyo University of Agriculture and Technology
Manabu Kano	Kyoto University

Member

Thomas Adams	McMaster University
Claire Adjiman	Imperial College London
Rakesh Agrawal	Purdue University
Iftikhar Ahmad	National University of Sciences and Technology
Sharifah Rafidah Wan Alwi	Universiti Teknoogi Malaysia
Suttichai Assabumrungrat	Chulalongkorn University
Parisa Bahri	Murdoch University
Michael Baldea	University of Texas at Austin
Ana Paula Barbisa-Póvoa	Universidade de Lisboa
Lorenz Biegler	Carnegie Mellon University
David Bogle	University College London
Fani Boukouvala	Georgia Institute of Technology
Giulia Bozzano	Politecnico di Milano
Arief Budiman	Universitas Gadjah Mada
Ian Cameron	The University of Queensland
Pedro Castro	University of Lisbon
Tianyou Chai	Northeastern University
Chuei-Tin Chang	National Cheng Kung University
Nishanth Chemmangattuvalappil	University of Nottingham – Malaysia
Cheng-Liang Chen	National Taiwan University
Xi Chen	Zhejiang University
Leo Chiang	Dow Inc.
I-Lung Chien	National Taiwan University
Min-Sen Chiu	National University of Singapore
Shoukat Choudhury	Bangladesh University of Engineering and Technology
Luis Cisternas	Universidad de Antofagasta
Selen Cremaschi	Auburn University
Prodromos Daoutidis	University of Minnesota
Maria Soledad Diaz	Universidad Nacional del Sur
Alexander Dowling	University of Notre Dame
Helen Durand	Wayne State University
Mario Richard Eden	Auburn University
Mahmoud El-Halwagi	Texas A&M University

Fadwa Eljack	Qatar University
Sebastian Engell	TU Dortmund University
Antonio Espuña	Technical University of Catalonia
Dominic Foo	University of Nottingham – Malaysia
Ferenc Friedler	The University of Győr
Rafiqul Gani	PSE for SPEED
Furong Gao	The Hong Kong University of Science and Technology
Salvador García-Muñoz	Eli Lilly and Company
Michail Georgiadis	Aristotle University of Thessaloniki
Chrysanthos Gounaris	Carnegie Mellon University
Ignacio Grossmann	Carnegie Mellon University
Martin Guay	Queen's University
Ravi Gudi	IIT Bombay
Takashi Hamaguchi	Nagoya Institute of Technology
Renanto Handogo	Institut Teknologi Sepuluh Nopember
Iiro Harjunkoski	ABB Power Grids Research
M.M. Faruque Hasan	Texas A&M University
Gabriela Henning	Universidad Nacional del Litoral
Seongmin Heo	Dankook University
Diane Hildebrandt	University of South Africa
Mohamed Azlan Hussain	University Malaya
Marianthi Ierapetritou	University of Delaware
Nevin Gerek Ince	AVEVA
Arturo Jimenez	Instituto Tecnológico de Celaya
Xavier Joulia	Université de Toulouse
Niket Kaisare	Indian Institute of Technology Madras
Iftekhar Karimi	National University of Singapore
Yoshiaki Kawajiri	Nagoya University
Soorathep Kheawhom	Chulalongkorn University
Christopher Kieslich	Auburn University
Yasunori Kikuchi	The University of Tokyo
Sanghong Kim	Tokyo University of Agriculture and Technology
Naoki Kimura	Kyushu University
Paisan Kittisupakorn	Chulalongkorn University
Zdravko Kravanja	University of Maribor
Jay H Lee	KAIST
Jong Min Lee	Seoul National University
Moonyong Lee	Yeungnam University
Grégoire Léonard	University of Liège
Jinghai Li	Chinese Academy of Sciences
Zukui Li	University of Alberta
Hankwon Lim	UNIST

Fernando V. Lima	West Virginia University
Patrick Linke	Texas A&M University – Qatar
Liu Liu	Jiangnan University
Vladimir Mahalec	McMaster University
Thokozani Majozi	University of the Witwatersrand
Zainuddin Abd Manan	Universiti Teknologi Malaysia
Flavio Manenti	Università degli Studi di Milano
Seyed Soheil Mansouri	Technical University of Denmark
Christos Maravelias	Princeton University
Mariano Martin	University of Salamanca
Hideyuki Matsumoto	Tokyo Institute of Technology
Ruth Misener	Imperial College London
Alexander Mitsos	RWTH Aachen University
Ludovic Montastruc	INP-ENSIACET
Il Moon	Yonsei University
Iqbal Mujtaba	University of Bradford
Cláudio Augusto O Nascimento	University of São Paulo
Masaru Noda	Fukuoka University
Costas Pantelides	Imperial College London & Process Systems Enterprise Ltd.
Stavros Papadokonstantakis	Chalmers University of Technology
Lazaros Papageorgiou	University College London
Stratos Pistikopoulos	Texas A&M University
Joe Qin	City University of Hong Kong
G. V. Rex Reklaitis	Purdue University
Lakshminarayanan Samavedham	National University of Singapore
John Siirola	Sandia National Laboratories
Sigurd Skogestad	Norwegian University of Science and Technology
Ken-Ichiro Sotowa	Kyoto University
Rajagopalan Srinivasan	Indian Institute of Technology Madras
Hongyu Su	Zhejian University
Hirokazu Sugiyama	The University of Tokyo
Christopher Swartz	McMaster University
Kazuhiro Takeda	Shizuoka University
Raymond Tan	De La Salle University
Ana I. Torres	Universidad de la República
Anjan Tula	Zhejiang University
Metin Turkay	Koc University
Petar Varbanov	Brno University of Technology
Dimitrios Varvarezos	Aspen Technology
Xiaonan Wang	National University of Singapore
Yuan Yao	National Tsing Hua University
Fengqi You	Cornell University

Xigang Yuan	Tianjin University
Lei Zhang	Dalian University of Technology
Jinsong Zhao	Tsinghua University
Teng Zhou	Max Planck Institute, Magdeburg
Edwin Zondervan	University of Twente
Danielle Zyngier	Hatch

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50001-4

Actions toward carbon-neutral society with fuel cell technology

Yoshihiko Hamamura

Fuel Cell Business Field., *Toyota Motor Corporation, Aichi 471-8571, JAPAN*
yoshihiko_hamamura@mail.toyota.co.jp

Abstract

In recent years, many countries have announced the declarations and policies toward a carbon-neutral society in response to global warming. The entire industries must tackle with the issues and it is necessary not only to develop zero CO_2 emission vehicle, but also to reduce CO_2 emission during the process of vehicle manufacturing, disposal, and fuel production in case of the automotive industry. Hydrogen is considered to be an important energy toward carbon-neutral society because of portability, storability, and producibility from various renewable energies such as solar and wind. Toyota is challenging toward the wide expansion of fuel cell applications by communicating with the customers and utilizing the fuel cell system for the various applications around the world as well as the 2nd-generation MIRAI, the fuel cell electric vehicle (FCEV). Toyota is also proposing the fuel cell system modules for the efficient packaging so that every customer can deploy them to their system products without a large effort and time. Further collaboration and communication from a wide range of perspectives will be essential with every stakeholder in the fuel cell industry and research field because there are many technical hurdles toward the promotion of the application of the fuel cell systems. The strength of PSE, which specializes in system integration and optimization, is strongly demanded to solve such interdisciplinary and complex issues surrounding the fuel cell technologies.

Keywords: Carbon neutrality; Hydrogen; Fuel cell; System application

1. The world trend of carbon neutrality and hydrogen energy

Carbon neutrality and decarbonization are gathering attention, especially after the Kyoto Protocol (United Nation, 1997) and the Paris Agreement (United Nation, 2015), for the prevention of global warming. Table 1 shows the recent world trend surrounding carbon neutrality and hydrogen energy. In 2018, 1st. International Conference on Hydrogen was held in Japan and the roadmaps for the promotion of hydrogen utilization were announced. In 2019, many countries declared high numerical targets toward carbon neutrality and decarbonization as follows: deployment of 8 million fuel cell electric vehicles (FCEVs) and 400 - 1000 hydrogen fueling stations by 2050 in E.U.; 10 million FCEVs and hydrogen fueling stations in every 10 km on the roads within next 10 years in Japan; 1.2 million FCEVs and hydrogen fueling stations by 2030 in U.S.A. In 2020, large scale investments were announced from many countries all over the world as follows: Investment of 750 billion Euro in total for a 'Green Deal' was announced by the E.U. targeting the recovery from the COVID-19 crisis, in which 30 % is shared to climate control related purposes; U.K. set the target of decarbonization by 2050 and the prohibition of sales of gasoline and diesel internal combustion engine vehicles after 2030; and Japan declared the achievement of carbon neutrality by 2050. In 2021, The U.S.A government announced the recovery to the Paris Agreement, the investment of 2 trillion

U.S. Dollars for the environmental and infrastructural industries, and the target of carbon neutrality by 2050. The features of hydrogen in storability and portability of energy allow it to function well with a variety of renewable energy sources such as solar and wind, thus making hydrogen an essential energy source for achieving carbon neutrality (Daud et al., 2017).

Table 1. The world trend surrounding carbon neutrality and hydrogen energy

Year	Country	Policy / Event
2018	Netherlands	Hydrogen Roadmap
	U.S.A	DOE Hydrogen and fuel cell Program Overview
	E.U.	Hydrogen initiative
	Japan	International Conference on Hydrogen
	Australia	National Hydrogen Roadmap
2019	Republic of Korea	Hydrogen Economy Promotion Roadmap
	E.U.	FCH-JU Hydrogen Roadmap
	Japan	Hydrogen and Fuel Cell Roadmap
		International Conference on Hydrogen
	France	Hydrogen Roadmap
	Australia	Renewable Hydrogen Strategy
	U.S.A	FCHEA Roadmap
		California FC-Bus Roadmap
	Saudi Arabia	Renewable Energy Strategy
2020	E.U.	European Green Deal
	Republic of Korea	Green New Deal
	Japan	Carbon Neutrality in 2050
	U.K.	Green Industrial Revolution
2021	U.S.A	Green revolution
	Japan	Decarbonization Supply-Chain among Japan, Australia, and India

2. Fuel cell development for various applications and usages

The 1st-generation MIRAI was launched in 2014 as the world's first mass-production fuel cell electric vehicle (FCEV). In addition to the improvement in many aspects of the system performance, the 2nd-generation MIRAI was designed to achieve a significant improvements in fundamental vehicle features. It has not only high environmental performance, but also responsive acceleration, superior handling and vehicle control, reliable cruising range, an increased number of passengers, and a stylish exterior design as shown in Fig.1. It is expected that the role of the 1st-generation MIRAI is the 'starting point for popularization' and the 2nd-generation MIRAI is 'accelerating to popularization'.

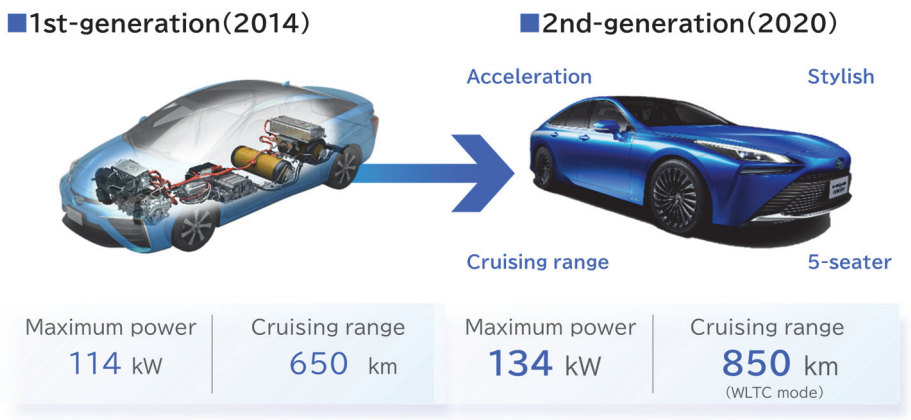


Fig. 1. MIRAI, fuel cell electric vehicle (FCEV) from TOYOTA

MIRAI's fuel cell system was not developed exclusively for MIRAI. The implementation of the 1st-generation fuel cell system to a variety of applications in addition to passenger vehicles, such as commercial vehicles of buses and tracks, stationary power generators, forklifts, and ships, has been investigated since 2014. In the process of developing these applications, it became obvious that significant time and effort were required until the completion of system integration for each application.

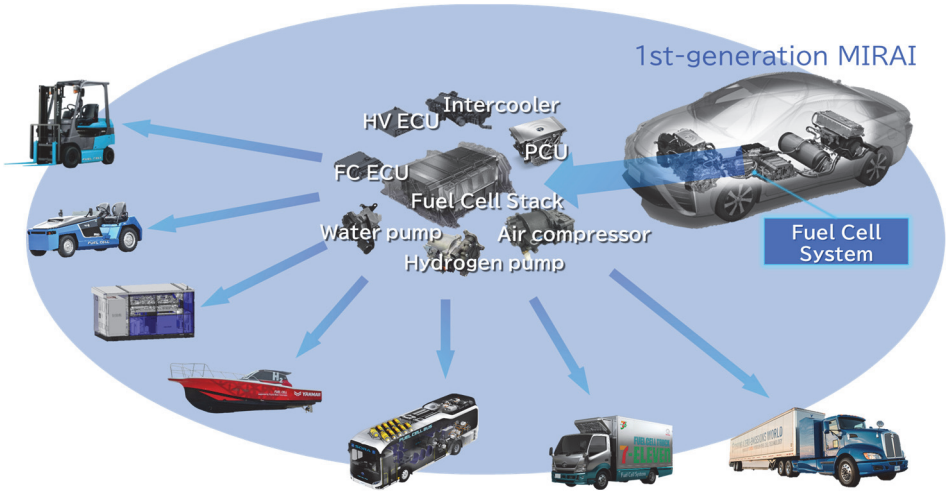


Fig. 2. The lineup of the different geometries of the fuel cell system modules

To overcome the problem described above, the system-modularization concept was adopted for 2nd-generation fuel cell system development. Vertical, horizontal and compact fuel cell system modules have been developed, where the same 2nd-generation system components such as fuel cell stack, air compressor, hydrogen pump, water pump, intercooler, engine control units (ECUs), and power control units (PCU) are efficiently repackaged. In addition, the interfaces between the fuel cell system and the applied system

are gathered in one mounting surface in a rectangular package for ease of integration. Customers can choose suitable fuel cell system modules from the lineup according to their application requirements. It is expected that these products reduce the barrier to entry into the fuel cell industry with benefits for the customers and users. Actually, our partners are accelerating the development of the fuel cell powered trucks, buses, stationary power generators, ships, and construction machinery by utilizing these fuel cell system modules as shown in Fig. 3.

Fig. 3. The implementation of the fuel cell system to a variety of purposes

3. Collaboration for developing fuel cells and expectations for PSE

Though hydrogen and fuel cell technologies are steadily spreading around the world, many technical issues are still remaining. It is important for every stakeholder in the fuel cell industry to communicate with each other, move in the same direction in addition to a proper competition, and make the fuel cell technology easier to use.

Fig. 4 is a conceptual drawing of the 'FC-Platform' project supported by NEDO, New Energy and Industrial Technology Development Organization of Japan, as an example of the recent research consortium (FC-Cubic, 2021). This consortium is acting as a role of platform where the fuel cell researchers with various important and detailed knowledge can gather from universities and technical colleges across Japan. They are investigating the complex fuel cell reactions and mass transport with the high resolution analysis instruments for the physicochemical phenomena in atomic scale. The government and fuel cell manufacturers can share their specific issues, discuss the research and development roadmap with them, and collaborate and communicate with each other in the consortium.

PSE is the academic field which specializes in the system modeling, process synthesis/aggregation, and decomposition/analysis (Klatt et al., 2009). Since the fuel cell technology is highly interdisciplinary technical fields where the wide range of physics in various scale from nano to meter must be considered, the role of PSE will be more important as an interface of the variety of technical fields. The author encourages PSE engineers and researchers to join the fuel cell industry and research activities.

Fig. 4. The conceptual drawing of the 'FC-Platform' project in Japan as an example of the recent fuel cell research consortium

Conclusions

The role of hydrogen energy and fuel cells were discussed by describing the world trend surrounding carbon neutrality. The 2nd-generation fuel cell electric vehicle MIRAI and a fuel cell system module comprised of the MIRAI fuel cell components were shown. The fuel cell system module facilitates the development of fuel cell applications such as commercial vehicles of buses, tracks, and marine. Finally, the expectation for PSE researchers and engineers from the fuel cell industry were discussed. The author encourages PSE engineers and researchers to join the fuel cell industry and research activities.

References

FC-Cubic, 2021,
https://www.fc-cubic.or.jp/,
Jan. 26th, 2022, accessed

K. U. Klatt, W. Marquardt, Perspective for process systems engineering – Personal view from academia and industry, *Computers and Chemical Engineering*, 33(3), 2009, Pages 536-550

Toyota Motor Corporation, 2022,
https://www.toyota.co.jp/fuelcells/en/technology.html,
Jan. 26th, 2022, accessed.

Toyota Motor Corporation, 2022,
https://www.toyota.co.jp/fuelcells/en/applications.html,
Jan. 26th, 2022, accessed.

United Nations, Kyoto Protocol to the United Nations Framework Convention on Climate Change, 1997,
http://unfccc.int/resource/docs/convkp/kpeng.pdf,
Jan. 26th, 2022, accessed

United Nations, Paris Agreement, 2015,
http://unfccc.int/files/essential_background/convention/application/pdf/english_paris_agreement.pdf,
Jan. 26th, 2022, accessed

W.R.W. Daud, R.E. Rosli, E.H. Majlan, S.A.A. Hamid, R. Mohamed, T. Husaini,
PEM fuel cell system control: A review, *Renewable Energy*, Volume 113, 2017, Pages 620-638,

This page intentionally left blank

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50002-6

Challenges and Opportunities for Process Systems Engineering in a Changed World

Rafiqul Gani[a*], Xi Chen[b], Mario R. Eden[c], Seyed S. Mansouri[d], Mariano Martin[e], Iqbal M. Mujtaba[f], Orakotch Padungwatanaroj[g], Kosan Roh[h], Luis Ricardez-Sandoval[i], Hirokazu Sugiyama[j], Jinsong Zhao[k], Edwin Zondervan[l]

[a]*PSE for SPEED Company, Ordup Jagtvej 42D, DK-2920 Charlottenlund, Denmark*
[b]*Department of Control Science & Engineering, Zhejiang University, Hangzhou, China*
[c]*Department of Chemical Engineering, Auburn University, Auburn, AL 36849, USA*
[d]*Department of Chemical Engineering, Technical University of Denmark, DK-2800 Lyngby, Denmark*
[e]*Dep artment of Chemical Engineering,, University of Salamanca, Plz. Caídos 1-5, 37008 Salamanca, Spain*
[f]*Department of Chemical Engineering, University of Bradford, Yorkshire BD7 1DP, UK*
[g]*PSE for SPEED Company, 294/65 RK Office Park, Romklao Rd., Bangkok, Thailand*
[h]*Department of Chemical Engineering and Applied Chemistry, Chungnam National University, 34141 Daejeon, Republic of Korea*
[i]*Department of Chemical Enginering, University of Waterloo, N2L 3G1, Canada*
[j]*Department of Chemical Systems Engineering,, The University of Tokyo, 113-8656 Tokyo, Japan*
[k]*State Key Laboratory of Chemical Engineering, Tsinghua University, Beijing, China*
[l]*University of Twente, Faculty of Science and Technology, Enschede, The Netherlands*

**rgani2018@gmail.com*

Abstract

Changes have always been taking place on earth. However, the latest changes related to the climate, the COVID-19 pandemic, natural resources, pollution, to name a few, have changed our world and a new normal is emerging. The energy-water-environment-food-health nexus is becoming more complex. These challenges, however, also provide opportunities to tackle them and make scientific and engineering advances. PSE is well-placed through its core and expanding domain as well as its ability to apply a systems approach to meet current and future challenges. Many opportunities exist for the PSE community to take the lead in managing this complexity. This paper will provide an overview on some of the key challenges and opportunities where PSE could make immediate as well as long lasting impacts by developing sustainable and innovative solutions. Focus will be placed on the choice of problems to solve and the solution approaches that could make an impact and help to define the new normal for future generations.

Keywords: Process Systems Engineering; Climate; COVID-19; Resources; Pollution; Opportunities

1. Introduction

The effects of climate change, COVID-19 infections, inefficient resources utilization and uncontrolled pollution, to name a few, have changed the world and now urgent actions are needed to not only minimize their impact but also to find novel and innovative solutions that are environmentally-friendly and allow the sustainable development of society. These problems are global and their solutions no doubt need a multi-disciplinary approach. As defined by Pistikopoulos *et al*. (2020), Process Systems Engineering (PSE) is the scientific discipline of integrating scales and components describing the behaviour of a physicochemical system, via mathematical modelling, data analysis, design, optimization and control. PSE provides the 'glue' within scientific chemical engineering (and other related engineering domains), and offers a scientific basis and computational tools towards addressing contemporary and future challenges such as in energy, environment, the 'industry of tomorrow' and sustainability. As Sargent (2004) pointed out, "PSE is all about the development of systematic techniques for process modelling, design and control - some formulate their synthesis, design and/or control problem, or some useful simplification of it, in precise mathematical terms, and then seek to exploit the mathematical structure to obtain an effective algorithm, while others seek insight on the problem structure from physical intuition". Therefore, to address the challenges currently faced by society, the PSE community has the opportunity to play an important role (Grossmann and Harjunkoski, 2019) by helping to find novel and innovative solutions that can not only arrest the undesired trends but also guide us towards achieving the well-established goals of sustainable development (UN, 2021).

Figure 1 shows plots of effect X under *business as usual* and *controlled actions* as a function of time, where X could be any one of the challenges with respect to climate, COVID-19, resources utilization, pollution, sustainability and many more. Figure 1 also points out Earth's capacity to absorb the negative results of effect X, which means the *business as usual* curve needs to be flattened with *controlled actions* before it is too late. Note that although one plot is shown to highlight the concept, the actual curves and earth's capacity are different for different effects. The energy-water-environment-health-food nexus (Al-Ansari *et al*., 2015; Mujtaba *et al*., 2018; Slorach *et al*., 2020) indicate however, that the individual items cannot be considered in isolation from each other because they have intrinsic interactions. The biocapacity of earth, which is a measure of its natural resources against its activities has reduced to 1.7 (Global Footprint Network, 2016) and must not become negative.

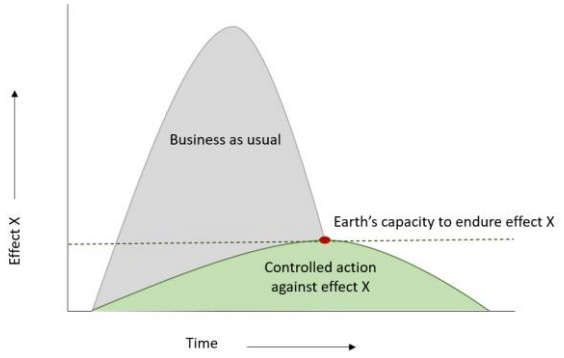


Figure 1: Business as usual versus control action against effect X (climate change, COVID-19 infections, resources utilization, pollution, sustainability, and many more)

Earth's natural resources for energy, water, biomass, minerals, etc., are not uniformly distributed but they are needed everywhere in different forms. According to Gani *et al*. (2020), a core activity of chemical and biochemical engineering is to convert these resources to products (various energy products, food products, health-care products, *etc*.), devices (car, television, phone, *etc*.) and/or services (electricity, fuels, fresh water, *etc*.) that society needs for its sustainability. The recent pandemic due to COVID-19 has shown that while the problem is global, their severity is different at different regions on earth. The same is true for global warming, resources utilization, pollution and major issues that impact society.

The objectives of this paper are to highlight selected focus areas representing current and future challenges on earth, the choices of problems to solve within them, and the systems approach-based solution methods that may be used to tackle them. The paper is organized as follows: a brief overview of the current status of PSE methods and associated computer-aided tools are given following the introduction; that is followed by a discussion on focus areas where PSE methods and tools can be applied together with a selected set of examples; and, ending with perspectives and concluding statements. The contents of the paper are the result of a discussion between the corresponding author and the co-authors on the following issues:

- In which of the problems can the PSE community make a significant contribution?
- What role should the PSE community play in tackling these problems?
- Which PSE methods and tools are best suited to tackle the problems?

This paper will also form the basis for the plenary lecture to be given by the corresponding author at the PSE2021 with material to be added.

2. PSE methods and tools

In the multi-layered view of PSE, Pistikopoulos *et al*. (2022) classified the PSE methods and associated tools in terms of the inner *fundamental* layer that involves process-product related activities where application of the fundamental concepts of PSE help to design, build and operate manufacturing processes that convert specific raw materials to desired products; the middle *expanding* layer that involves resources-efficiency related activities leading to the development of new technologies and more sustainable engineering solutions; the outer *unifying* layer that involves activities related to tackling of societal challenges leading to a more sustainable society. In this paper, some of the PSE methods and associated tools from all three layers are highlighted. See also Stephanopoulos and Reklaitis (2011) for a historic review of PSE activities and Grossmann and Harjunkoski (2019) for academic and industrial perspectives on PSE.

2.1. Methods

Three topics under methods are highlighted: modelling (because it is at the core of almost all PSE methods); numerical methods (which is necessary to solve the equations representing any model and are continuously updated to satisfy the demands from the expanding and unifying layers of PSE) and algorithms (since the schemes for integration and/or combination of models and numerical methods are needed to tackle the societal challenges from the unifying layer of PSE).

2.1.1. Modelling

A model (Hangos and Cameron, 2001) is a pattern, plan, representation, or description designed to show the structure or workings of an object, system, or concept; it could also be a study of a miniature of the actual; the model objectives need to be clearly defined. Modelling is the process of generating abstract or conceptual representation of a physical

system, *i.e*., representing reality in a virtual environment for a purpose. In systems approaches to problem solution, modelling is a core activity in the development of any model-based problem solution tool and requires a very good understanding of the system being investigated. As our understanding of the sub-systems whose problems we would like to solve, is incomplete, models based only on first principles cannot usually be obtained (or it is very time-consuming and resource demanding to obtain them). Therefore, knowledge related to the core fundamental layer of PSE is not sufficient, modelling options from the middle expanding layer of PSE, for example, systems identification or artificial intelligence (including machine learning, deep data learning, nature-inspired, *etc*.) need to be utilized (Lee *et al*., 2018; Venkatasubramanian, 2019). Note that the modelling objectives are related to specific problems that need to be solved, giving rise to models of different complexities and forms. Therefore, to match and/or improve the numerical solver efficiency and reliability, local models (Chimowitz *et al*., 1983), hybrid models (Chaffart and Ricardez-Sandoval, 2018), or surrogate models (Bhosekar and Ierapetritou, 2018) are being used. A few modelling related issues are highlighted below (not ordered in terms of priority):

- Development of predictive property models – more than 10 million chemicals have been identified but measured data are available for less than 50 thousand chemicals; properties such as toxicity, color, smell, *etc*., need to be modelled.
- How to obtain new process (operation) models from generic models when new processes and/or operations do not have all required sub-system details; new process models such as intensified operations, fuel cells, medicine delivering devices.
- How to create plug and play options for links to external databases, solvers, models and/or new theory, data, computational resources in currently available computer-aided tools to expand their application range.

2.1.2. Numerical solvers

Models representing a system of interest, consisting of different combinations of equations (algebraic, ordinary differential, partial differential, symbolic, *etc*.) involving different types of variables (real, integer, Boolean, symbolic, *etc*.) require appropriate numerical solvers. According to Pistikopoulos *et al*. (2021), the two key tasks that PSE have focused on are *i)* optimization methods, comprising a variety of formulations, most notably mixed-integer linear and nonlinear programs, dynamic optimization (including optimal control) and hierarchical optimization (semi-infinite, bilevel, trilevel) and *ii)* simulation/optimization of dynamic systems with hybrid discrete-continuous (or in some case equivalently non-smooth) nature. Kronqvist *et al*. (2019) and more recently Nolasco *et al*. (2021) have reviewed optimization solvers. For a list of selected numerical solvers commonly used by the PSE community, see also Pistikopoulos *et al*. (2021). To improve the convergence and reliability of numerical solvers, symbolic computation methods, which directly use mathematical expressions for operations and derivations to identify the solution have been recently proposed by Zhang *et al*. (2021a, 2021b). Below, a few numerical solver related issues are highlighted (not ordered in terms of priority):

- Which criteria (*e.g*., efficiency, reliability, and/or flexibility) should be used for numerical solver selection, when for similar problems more than one solver could be available.
- How to adapt currently available solvers to emerging modes of computation such as, quantum computing.
- Options to incorporate features such as machine learning, data analytics, etc. into the solver algorithm to make them intelligent.

2.1.3. Algorithms

Although algorithms also refer to the work-flow of numerical solvers, in this section, the term is used to refer to only synthesis, design and/or analysis of a wide range of process (chemical, petrochemical, biochemical, pharmaceutical, food, *etc.*) and/or chemicals-based products (solvents, refrigerants, fuel blends, paints, detergents, *etc*.), including devices (medicine delivery, power-supply, air-purifier, *etc*.). They are further classified in terms of available options (sustainable design, design under uncertainty, reverse design, flexibility and/or controllability analysis, *etc*.); type of approach (rule-based, process groups based, superstructure-based, *etc*.); and, application areas (process technology, product technology, analyser technology, etc.). The following selected articles provide an overview on the challenges and opportunities related to specific topics: Chen and Grossmann (2017) on process synthesis; Schilling *et al*. (2017) and Papathanasiou and Kontoravdi (2020) on product and process synthesis-design; Skiborowski *et al*. (2014) and Tula *et al*. (2017) on sustainable process design; Morari and Lee (1999), and, Yu and Biegler (2019) on process control, Diangelakis *et al*. (2017) and Rafiei and Ricardez-Sandoval (2020) on integrated process design and control optimization; Garcia and You (2015) on supply chain design and optimization, and, Maravelias and Sung (2009) on production planning and scheduling. Some algorithm related issues are briefly highlighted below (not ordered in terms of priority):

- Application range versus reliability versus flexibility.
- Detailed activity diagram (model) needed for software implementation.
- Adoption of hybrid approaches (interfacing of algorithms with computational resources).

2.2. Computer-aided tools

The PSE community continue to develop problem specific computer-aided tools for a wide range of applications. The most well-known PSE computer-aided tool is the process simulator, available in different versions, that is widely used for education as well as industrial practice. However, are the current versions of the various simulators able to solve the problems related to the energy-water-environment-food-health nexus? Also, as pointed out by Tula *et al*. (2019), process simulation is just one out of many tasks that needs to be performed for sustainable and innovative design. For example, tools for modelling are needed if the required model is not available in the simulator model library; or, a product design tool is needed to design-select an appropriate chemical for solvent-based separation; or, a process synthesis tool is needed to generate a flowsheet if a reference flowsheet is not available; or, analysis tools (sustainability, safety and hazards, LCA, economics, etc.) are needed to verify the feasibility of the chemical process. Pistikopoulos *et al*. (2021) provides a list of the above-mentioned tools developed by the PSE community. Some issues related to computer-aided tools are briefly highlighted below (not ordered in terms of priority):

- Application range of the available models in the model libraries – are they problem specific according to application area?
- Can the models, data, algorithms be adopted from one sector to another?
- Are simulation and design (including synthesis and analysis) options available in the same tool?
- Can they serve as virtual reality simulators to provide users with real experience?

3. Challenges and roles of PSE

Table 1 lists selected data to highlight the challenges and issues. The energy-water-environment-food-health nexus is getting more complex in the changing world and an integrated solution approach is necessary to tackle better the interactions among the individual effects. For example, energy in the form of fuels and electricity is needed by industry, transportation and housing sectors, but the type of fuel and the electricity generation defines the amounts of green-house gas (GHG) emissions, which in turn is related to climate change. Supply chain factors as well as waste disposal and therefore, resource utilization and pollution are also related, particularly for the chemical, petrochemical and pharmaceutical sectors. With respect to COVID-19, the pandemic is still not under control (December 31st, 2021) even though the rates of hospitalizations and deaths appear to be slowing down (see Table 1). However, based on data on consumption of energy (see Table 1), the non-renewable resources that emit CO_2 still dominate energy supplies. Capture, utilization, and/or sequestration of CO_2 is a challenge where adopted PSE methods and tools can play an important role. Society's daily needs such as plastics for packaging; chemicals for drugs, cosmetics, detergents, *etc.*; rare earth metals for construction, equipment, cars, etc., are causing pollution of land, water and/or air upon disposal and through their end-of-life properties.

Table 1: Current status of selected effects

Effect	Data	Reference
Global warming	New estimates of the chances of crossing the global warming level of 1.5°C in the next decades indicate that unless there are immediate, rapid and large-scale reductions in greenhouse gas emissions, limiting warming to close to 1.5°C or even 2°C will be beyond reach	IPCC (2021)
COVID-19	Worldwide total infections have reached 285,231,011 with 5,442,088 deaths as of 31 December 2021	Worldometers (2021)
Resources (energy)	CO_2 emitting non-renewable resources still contribute nearly 85% of the energy (not electricity) and around 65% for generation of electricity	Vooradi *et al*. (2017); IEA (2021)
Resources (water)	97% of water on earth is salt water and only 3% is fresh water, out of which, 68.1% is ice-caps and glaciers, 30.1% is ground water, 0.3% is surface water and 0.9% is other); Globally, at least 2 billion people use contaminated drinking water source	WHO (2019); Greenlee *et al*. (2009)
Resources (biomass)	Currently biomass contributes 3.4% of the total transportation energy demand. It would require 2.4 times the amount currently devoted to all energy demands (or, more than 1.3 times earth's current biomass resources) to satisfy only the total energy demand for the transportation sector in 2030.	Energy (2021); WBA (2020); IRENA (2014); IEA (2016)
Pollution (plastic waste, water)	Projected plastic waste generation of the EU-27 is estimated to reach 17 Mt/y in 2030	Fan *et al*. (2022)
Pollution (GHG release)	The GWP of CH_4 and N_2O are around 27~29 and 273 (100-year time period), respectively.	IPCC (2021)

Where should the focus to tackle these challenges be, what should be the choice of problems to solve, and, which solution approaches should be applied that could make an impact and help to define the new normal for future generations are still open questions that need to be addressed. Guillén-Gosálbez *et al*. (2019) recently highlighted process systems engineering thinking and listed tools that could be applied to solve sustainability related problems. Bakshi (2019) reviewed the role of process systems engineering toward sustainable chemical engineering. Burre *et al*. (2020) discussed how process systems engineering can help address common challenges for Power-to-X technologies. Martin *et al*. (2022) the challenges and opportunities related to sustainable process synthesis, design and analysis. Avraamidou *et al*. (2020) highlighted the challenges and opportunities for PSE related to achieving circular economy.

3.1. Sustainable process networks

The production of clean energy is directly linked to avoiding the depletion of natural resources, even if this goal has a more general aim, as well as limiting the production of waste and avoiding polluting the environment. According to the carbon-neutral roadmap presented by each government, chemical and petrochemical industries must reduce their greenhouse gas emissions to achieve zero net emissions by 2050. In order to significantly reduce carbon emissions in such industries while maintaining the current business portfolio, adoption of new technologies that can directly utilize electricity originating from renewable energy resources such as solar and/or wind energy to produce chemical products is necessary (Rangel-Martinez *et al*., 2021). Process integration at the different manufacturing and production levels will play a critical role to ensure efficient and sustainable operation of existing and emerging systems, and their corresponding integration (Burnak *et al*., 2019; Rafiei and Ricardez-Sandoval, 2020). Figure 2 highlights the concept of integration of sub-networks of utilities (energy and water), process for conversion of optimal product(s), integrated with capture and utilization of captured CO_2. Decisions related to individual sub-networks need to be made such that the overall objectives of sustainable design are satisfied. The objective for sustainable design of networks could be, for example, to find a design with zero or negative CO_2 (preferably all GHG) that is economically feasible, operationally safe and environmentally acceptable. Choices of raw materials and products can represent any industrial sector while choices of resources for utilities need to be made such that net zero emission requirement can be satisfied. Note that in Figure 2, the processing of waste utilities (energy, water), waste material, by-products for recycle, re-insertion and/or utilization in the process thus promoting circular economy is not shown.

Using core PSE methods and associated tools Li *et al*. (2022) have developed a conceptual application example of such a superstructure based sustainable network synthesis. Roh *et al*. (2018) have developed a computer-aided tool (called ArKaTAC^3) that allows to perform both superstructure-based process synthesis and multi-dimensional analyses (including techno-economic analysis and life cycle assessment) of carbon capture and utilization systems. Filippini *et al*. (2019) reported design and economic evaluation of solar-powered hybrid multi effect and reverse osmosis system for seawater desalination. Also, Sanchez *et al*. (2019) have shown the utilization of captured CO_2 hydrogenation with green hydrogen for methanol, ammonia and urea productions, while, Guerras *et al*. (2021) proposes that biomass as a renewable resource should be devoted for the production of added value products (for example, in the pharma, food additives, health sectors) and only wastes should be used for energy production. Evaluation of these new technologies should also consider operational flexibility (Mitsos *et al*., 2018) as well as safety (Eini *et al*., 2016).

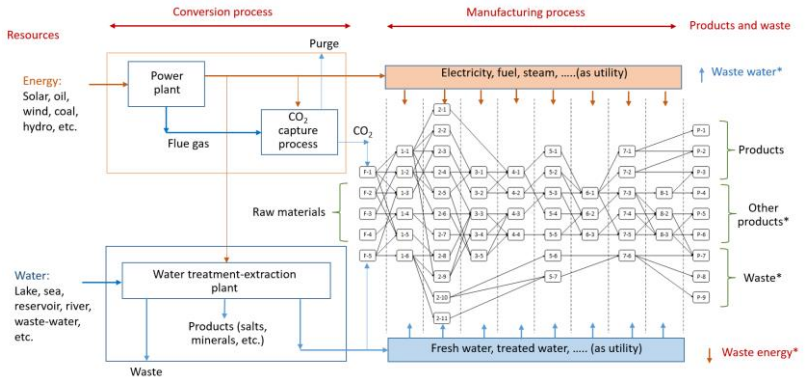


Figure 2: Superstructure for optimal integrated net zero emission network to achieve circular economy (note: * indicates they will need additional processing for recycle, re-insertion and/or utilization).

3.2. Chemical process safety

Even though the chemical industry has achieved a very impressive improvement in occupational safety, the reduction in major process accidents (on a global basis) has been less impressive and the insured losses due to major accidents in the chemical industry have not reduced in the last 30 years. It is estimated that 70% of the chemical accidents were caused by human errors. These incidents also point to inefficient resource utilization. PSE should be able to play a key role in developing and deploying advanced artificial intelligence-based technologies that assist operators in estimating and/or identifying all potential risks in the complex and dynamic chemical industrial operations and to make correct and consistent decisions. Interesting developments that could be evaluated for potential deployment are method for fault detection and diagnosis (see Fig. 3) based on transfer learning (Wu and Zhao, 2020), automatic frequency estimation of contributory factors for confined space accidents, natural language processing (Wang and Zhao, 2022), and, inherent safety and cognitive engineering as well as operator training (Srinivasan *et al*., 2019).

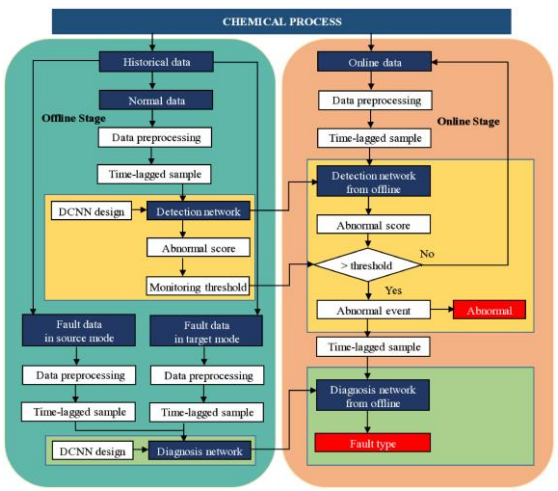


Figure 3: Framework for fault detection and diagnosis (Wu and Zhao, 2020).

3.3. Health and pharma sector

In the pharma sector, a principal challenge is to develop model (including data) based computer aided systems for synthesis, design, monitoring, control, as in chemical and petrochemical industries. Nevertheless, systems approach coupled with PSE methods and associated tools have made important contributions during the last decade, for example, in conversion from batch to continuous manufacturing (Ierapetritou *et al*., 2016), development of process analytical technologies, and data-driven approaches for active pharmaceutical ingredients (API) syntheses and design of powder-/bio- processes (*e.g*., Kim *et al*., 2021). Advances in development of health-care products (Fuentes-Garí et al., 2015) and their manufacturing processes such as monoclonal antibody drugs and stem cells have been reported (*e.g*., Hayashi *et al*. 2021). In sustainability and healthcare, COVID-19 has revealed the need for a systematic approach for vaccine/medicine development and supply chain. The urgent supply was mandated on top of maintaining the existing treatment capability. Besides COVID-19, R&D of innovative therapies, *e.g*., regenerative medicine, is ongoing. These therapies tend to be expensive while the economic aspect is becoming critical in many countries. Analyses and discussions towards sustainable healthcare society (*e.g*., Sugiyama *et al*., 2021) can be further expanded as a topic of the PSE community. Figure 4 highlights the challenges and opportunities for adoption of PSE methods and tools to prepare for future pandemics. Four potential problems to solve are highlighted together with the issues and needs that need to be addressed.

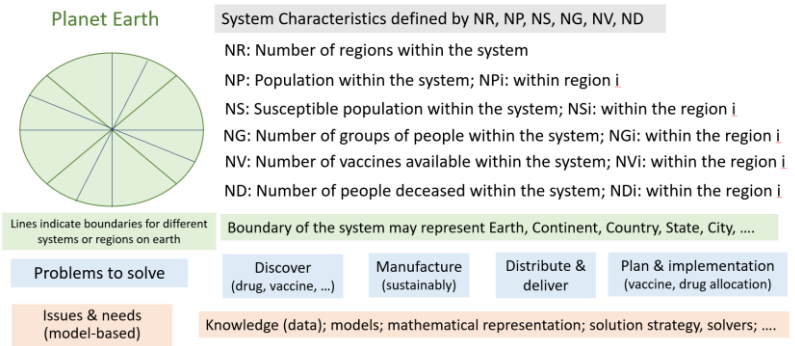


Figure 4: Visual plan to prepare for the next pandemic through PSE methods and tools

3.4. Chemicals based products and their substitution

In our changed world, we are living with chemicals that are in our food products, clothes, furniture, appliances, toys, cosmetics, medicines and many more. Society, for its existence anywhere on earth, needs to use a variety of products and/or means that are directly or indirectly connected to chemicals. Currently, more than one million chemicals can be found on planet earth and thousands of new chemicals-based products are entering the global market every year. Over 95 percent of all manufactured goods rely on some form of industrial chemical process (ICCA, 2019). As the number of chemicals grows rapidly, understanding their implications on human health and environment is increasingly becoming an issue. An important and urgent challenge is not only to identify the chemicals, which may have harmful effects, but also to substitute or control their use (Syeda *et al*., 2022). As demand for safer alternatives in products is increasing, regulatory authorities, such as EU REACH (EU, 2021), US EPA (EPA, 2021) and Occupational Safety and Health Administration (OSHA, 2021) have taken up substitution of chemicals

harmful to human health and environment as one of the central elements of their policies. As the pioneers of techniques such as computer-aided molecular design (CAMD) for chemical product design, the PSE community is well placed to assist and actively collaborate with the authorities to tackle this urgent problem. CAMD and different variations of it (Zhang *et al*., 2020, Adjiman *et al*., 2021), can easily be adopted for analysis of chemicals-based products and substitution of chemicals if hazardous chemicals are identified in the product. The chemical substitution problem and the possible solution steps are highlighted in Figure 5.

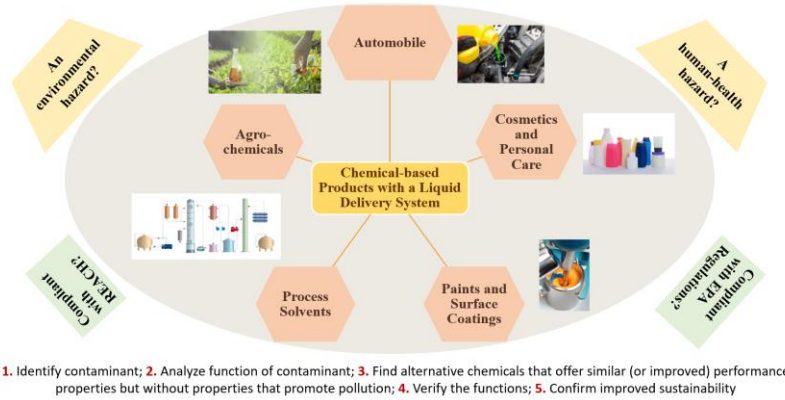


Figure 5: Chemical substitution problem and suggested solution steps.

4. Perspectives (Opportunities)

A "systems thinking" or "systems integration" approach is required, where PSE provides the glue (architecture for consistent, efficient, and smooth data transfer from one tool to another) for integration of energy supply, water management, control of greenhouse gas emissions, process safety and economics and many other major issues that impact society and earth. PSE methods and associated tools can be adopted for sustainable and secure access to food, water and energy, leading to achievement of sustainable development goals, to develop and evaluate new technologies for carbon capture, utilization and sequestration (CCUS), and to close circular production systems with near-zero or minimum waste. Simultaneously, computer-aided molecular design (CAMD) techniques could be adopted for pollution control of water caused by plastics, including disposable personal protective equipment (PPE), *etc*.; pollution of air caused by GHGs, *etc*.; and pollution of land caused by disposed chemicals-based products.

In order to tackle the challenges of our changed world, the opportunity exists to not only adopt but also to develop new methods and tools as and when necessary. As models, modelling and data are at the core of all systems-based problem solution approaches, more effort is needed to understand systems that are outside our domain knowledge. It is important to use correct and consistent models and associated decision support tools for analysing the involved complex phenomena (*e.g*., powder processing, biological reactions, cell behaviour, solid solubility, toxicity, *etc*.) in such a way that the decisions related to process and products (*e.g*., production scale and mode, design space determination, process-operation specification, *etc*.) can be made.

Recognizing that resources in earth are not uniformly distributed, nevertheless, the concept of integrated networks could be applied for desalination to obtain fresh water using solar and/or wind energy sources where these can be harnessed. In regions where

there is fresh water in abundance, technologies already available to combine hydrogen obtained through electrolysis of water with captured CO_2 to produce methanol (as an intermediate chemical from which other high-value chemicals could be synthesized), with N_2 from air to produce ammonia (as a hydrogen carrier), and, methanol and ammonia to produce urea (as a fertilizer). The analysis of competing interests as well as the presentation of feasible/optimal solutions under uncertainty (*i.e.*, societal design) would be another important contribution from PSE.

Stable and sustainable coupling of chemical industry and power generation sectors by exploiting operational flexibility, optimal integration of design, operation, and control of power-intensive chemical plants that already exist and also should be newly developed. Implementation of new technologies, such as intensified equipment and/or hybrid combination of distillation and membrane, for targeted reduction of energy consumption as a short-term solution need to be promoted through bench-mark problem solutions.

Lessons learnt from the pandemic with respect to the challenges posed to the supply chain and development of novel schemes and policies that can potentially reduce the impact of current and future pandemics should be considered as a global supply chain system. With the expertise and experience of the PSE community in the supply chain sector, an optimal vaccine allocation system that is region specific to immunize the population at the fastest rate could be developed and distributed to the appropriate agencies. Recognition of healthcare as an element of the energy-water-environment-food-health nexus will help re-designing the manufacturing processes and beyond. PSE can contribute to help prepare for the next pandemic with, for example, vaccine allocation software (to be made available globally to all countries) to reach herd immunity at the shortest time and with the minimum loss of life.

5. Conclusions

Process systems engineering as a multi-disciplinary field of research has many opportunities to tackle some of the greatest challenges faced by today's society. This opportunity is provided by the rich literature and many ongoing current and future activities to provide integrated solutions within water-energy-food-waste-health nexus. However, more efforts are needed to understand and develop models, tools and solutions strategies to address those major challenges. This can result in developing new technologies and process systems by means of predictive models. These predictive models should be able to address short-term necessities while laying the foundations for long-term solutions over a time horizon to help alleviate the current and future challenges to address different global objectives, such as the UN sustainable development goals. However, all the efforts would be wasted if the demands (due to increased growth and promising economy) cannot be limited; circular economy with zero waste and ability to sustain changes will be impossible if resources disappear because of increased demand. Thereby, a systems thinking approach is essential to not lose the opportunities within rather narrow windows to address global issues, such as global warming and the COVID-19 pandemic.

References

C. S. Adjiman, N. V. Sahinidis, D. G. Vlachos, B. Bakshi, C. T. Maravelias, C. Georgakis, 2021, Process systems engineering perspective on the design of materials and molecules, *Industrial & Engineering Chemistry Research*, 60(12), 5194-5206.

T. Al-Ansari, A. Korre, Z. Nie, N. Shah, 2015, Development of a life cycle assessment tool for the assessment of food production systems within the energy, water and food nexus, *Sustainable Production & Consumption*, 2, 52-66.

S. S. Avraamidou, G. Baratsas, Y. Tian, E. N. Pistikopoulos, 2020, Circular economy - a challenge and an opportunity for process systems engineering, *Computers & Chemical Engineering*, 133, 106629.

B. R. Bakshi, 2019, Toward Sustainable Chemical Engineering: The Role of Process Systems Engineering, *Annual Review of Chemical and Biomolecular Engineering*, 10, 265-288.

A. Bhosekar, M. Ierapetritou, 2018, Advances in surrogate based modeling, feasibility analysis, and optimization: A review, *Computers & Chemical Engineering*, 108, 250-267.

B. Burnak, N.A. Diangelakis, E.N. Pistikopoulos (2019), Towards the grand unification of process design, scheduling, and control—utopia or reality?, *Processes*, 7, 461.

J. Burre, D. Bongartz, L. Brée, K. Roh, A. Mitsos, 2020. Power-to-X: Between Electricity Storage, e-Production, and Demand Side Management, *Chemie Ing. Tech*. 92, 74–84. (https://doi.org/10.1002/cite.201900102).

D. Chaffart, L. Ricardez-Sandoval, 2018, Optimization and control of a thin film growth process: A hybrid first principles/artificial neural network based multiscale modelling approach, *Computers and Chemical Engineering*, 119, 465-479.

E. H. Chimowitz, T. F. Anderson, S. M. Macchietto, L. F. Stutzman, 1983, Local models for representing phase equilibriums in multicomponent, nonideal vapor-liquid and liquid-liquid systems. 1. Thermodynamic approximation functions, *Industrial & Engineering Chemistry Process Design and Development*, 22, 217-225.

S. Eini, H. Shahhosseini, N. Delgarm, M. Lee, A. Bahadori, 2016. Multi-objective optimization of a cascade refrigeration system: exergetic, economic, environmental, and inherent safety analysis, *Applied Thermal Engineering*, 22, 168-186.

EPA, 2021, (http://www.epa.gov/chemical-research/program-assisting-replacementindustrial-solvents-paris-iii).

Energy, 2021, (https://www.energy.gov/eere/bioenergy/biomass-resources).

EU, 2021, European Chemicals Agency (https://echa.europa.eu/).

Y. V. Fan, P. Jiang, R. R. Tan, K. B. Aviso, F. You, X. Zhao, C. T. Lee, J. J. Klemeš, 2022, Forecasting plastic waste generation and interventions for environmental hazard mitigation, *Journal of Hazardous Materials*, 424, 127330.

G. Filippini, M. A. Al-Obaidi, F. Manenti, I. M. Mujtaba, 2019, Design and economic evaluation of solar-powered hybrid multi effect and reverse osmosis system for seawater desalination, *Desalination*, 465, 114-125

M. Fuentes-Garí, E. Velliou, R. Misener, E. Pefani, M. Rende, N. Panoskaltsis, 2015, A systematic framework for the design, simulation and optimization of personalized healthcare: making and healing blood, *Computers and Chemical Engineering*, 81, 80-93.

R. Gani, J. Bałdyga, B. Biscans, E. Brunazzi, J. C. Charpentier, E. Drioli, et al., 2020, A multi-layered view of chemical and biochemical engineering, *Chemical Engineering Research and Design*, 155, 133-145. (https://doi.org/10.1016/j.cherd.2020.01.008)

D. J. Garcia, F. You, 2015, Supply chain design and optimization: Challenges and opportunities, *Computers & Chemical Engineering*, 81, 153-170.

Global Footprint Network, 2016, (https://data.footprintnetwork.org/?_ga=2.81569751.1924591506.1641417907-308979742.1641417907#/)

L. F. Greenlee, D. F. Lawler, B. D. Freeman, B. Marrot, P. Moulin, 2009, Reverse osmosis desalination: Water sources, technology, and today's challenges, *Water Research*, 43, 2317–2348.

I. E. Grossmann, I. Harjunkoski, 2019, Process systems engineering: academic and industrial perspectives, *Computers & Chemical Engineering*, 126, 474-484.

L. Guerras, D. Sengupta, M. Martín, M. El-Halwagi, (2021) Multi-layer approach for product portfolio optimization: Waste to added value products, *ACS Sustainable Chemical Engineering*, 9, 18, 6410–6426.

G. Guillén-Gosálbez, F. You, A. Galán-Martín, C. Pozo, I. E. Grossmann, 2019, Process systems engineering thinking and tools applied to sustainability problems: current landscape and future opportunities, *Current Opinion in Chemical Engineering*, 26, 170-179.
K. Hangos, I. T. Kameron, 2001, Process modelling and model analysis, Academic Press, USA.
Y. Hayashi, M. Kino-oka, H. Sugiyama, 2022. Hybrid-model-based design of fill-freeze-thaw processes for human induced pluripotent stem cells considering productivity and quality, *Computers & Chemical Engineering*, 156, 107566.
ICCA, 2019, (https://icca-chem.org/wp-content/uploads/2020/10/Catalyzing-Growth-and-Addressing-Our-Worlds-Sustainability-Challenges-Report.pdf).
IEA, 2016, (https://www.eia.gov/outlooks/ieo/pdf/transportation.pdf).
IEA, 2021, (https://www.iea.org/reports/world-energy-balances-overview).
M. Ierapetritou, F. Muzzio, G.V. Reklaitis, 2016, Perspectives on the continuous manufacturing of powder-based pharmaceutical processes, *AIChE Journal*, 62 (6), 1846-1862.
IPCC, 2021 (https://www.ipcc.ch/assessment-report/ar6/).
IRENA, 2014, (https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2014/IRENA_REmap_2030_Biomass_paper_2014.pdf).
J. Kim, H. Yonekura, T. Watanabe, S. Yoshikawa, H. Nakanishi, S. Badr, H. Sugiyama, 2021. Model-based comparison of batch and flow syntheses of an active pharmaceutical ingredient using heterogeneous hydrogenation, *Computers and Chemical Engineering*, 156, 107541.
J. Kronqvist, D. E. Bernal, A. Lundell, I. E. Grossmann, 2019, A review and comparison of solvers for convex MINLP, *Optimization Engineering*, 20(2), 397–455.
J. H. Lee, J. Shin, M. J. Realff, 2018, Machine Learning: Overview of the Recent Progresses and Implications for the Process Systems Engineering Field, *Computers & Chemical Engineering*, 114, 111-121.
Y. Li, J. Wei, Z. Yuan, B. Chen, R. Gani, 2021, Sustainable Synthesis of Integrated Process, Water Treatment, Energy Supply, and CCUS Networks Under Uncertainty, *Computers & Chemical Engineering*, 157, 107636.
C. T. Maravelias, C. Sung, 2009, Integration of production planning and scheduling: Overview, challenges and opportunities, *Computers & Chemical Engineering*, 33, 1919-1930.
M. Martin, R. Gani, I. M. Mujtaba, 2022, Sustainable process synthesis, design and analysis: Challenges and opportunities, *Sustainable Production & Consumption*, (https://doi.org/10.1016/j.spc.2022.01.002)
A. Mitsos, N. Asprion, C.A. Floudas, M. Bortz, M. Baldea, D. Bonvin, A. Caspari, P. Schäfer, 2018. Challenges in process optimization for new feedstocks and energy sources, *Computers & Chemical Engineering*, 113, 209–221.
I. M. Mujtaba, R. Srinivasan, N. Elbashir, 2018, The water-food-energy nexus: Processes, Technologies and Challenges, CRC Press, USA.
N. Nolasco, V. S. Vassiliadis, W. Kähm, S. D. Adloor, R. Al Ismaili, R. Conejeros, T. Espaas, N. Gangadharan, V. Mappas, F. Scott, Q. Zhang, 2021, Optimal control in chemical engineering: Past, present and future, *Computers and Chemical Engineering*, 155, 107528.
OSHA, 2021, (http://www.osha.gov/dsg/safer_chemicals/why_transition.html).
M. M. Papathanasiou, C. Kontoravdi, 2020, Engineering challenges in therapeutic protein product and process design, *Current Opinion in Chemical Engineering*, 27, 81-88.
E. N. Pistikopoulos, A. Barbosa-Povoa, J. H. Lee, R. Misener, A. Mitsos, G. V. Reklaitis, V. Venkatasubramanian, F. You, R. Gani, 2021, Process Systems Engineering–The Generation Next?, *Computers & Chemical Engineering*, 147, 107252.
M. Rafiei, L.A. Ricardez-Sandoval, 2020, New frontiers, challenges, and opportunities in integration of design and control for enterprise-wide sustainability, *Computers & Chemical Engineering*, 132, 106610.
D. Rangel-Martinez, K. D. P. Nigam, L. A. Ricardez-Sandoval, 2021, Machine learning on sustainable energy: A review and outlook on renewable energy systems, catalysis, smart grid and energy storage, *Chemical Engineering Research and Design*, 174, 414-441.

K. Roh, H. Lim, W. Chung, J. Oh, H. Yoo, A. S. Al-Hunaidy, H. Imran, J. H. Lee, 2018, Sustainability analysis of CO2 capture and utilization processes using a computer-aided tool, *Journal CO2 Utilization*, 26, 60–69.
A. Sanchez, L.M. Gil, M. Martín, 2019, Sustainable DMC production from CO2 and renewable ammonia and methanol, *Journal of CO2 Utilization*. 33, 521-531.
R. W. H. Sargent, 2004, 25 years of progress in process systems engineering, *Computers & Chemical Engineering*, 28, 437–439.
J. Schilling, D. Tillmanns, M. Lampe, M. Hopp, J. Gross, A. Bardow, 2017, From molecules to dollars: integrating molecular design into thermo-economic process design using consistent thermodynamic modeling, *Molecular Systems Design & Engineering*, 2 (3), 301-320.
M. Skiborowski, A. Harwardt, W. Marquardt, 2014, Conceptual design of distillation based hybrid separation processes, *Annual Review of Chemical and Biomolecular Engineering*, 4(1), 45.
P. C. Slorach, H. K. Jeswani, R. Cuéllar-Franca, A. Azapagic, 2020, Environmental sustainability in the food-energy-water-health nexus: A new methodology and an application to food waste in a circular economy, *Waste Management*, 113, 359–368.
R. Srinivasan, B. Srinivasan, M. U. Iqbal, A. Nemet, Z. Kravanja, 2019, Recent developments towards enhancing process safety: Inherent safety and cognitive engineering, *Computers and Chemical Engineering*, 128, 364–383.
G. Stephanopoulos, G. V. Reklaitis, 2011, Process systems engineering: From Solvay to modern bio- and nanotechnology.A history of development, successes and prospects for the future, *Chemical Engineering Science*, 66, 4272–4306.
H. Sugiyama, N. Harada, E. Amasawa, M. Hirao, N. Yahagi, 2021, A prototype method for selecting reduction countermeasure against unused medicine considering patient characteristics, *J. Chemical Engineering of Japan*. 54, 152–161.
S. R. Syeda, E. A. Khan, O. Padungwatanaroj, N. Kuprasertwong, A. K. Tula, 2022, A perspective on hazardous chemical substitution in consumer products, *Current Opinion in Chemical Engineering*, 36, 100748.
A. K. Tula, D. K. Babi, J. Bottlaender, M. R. Eden, R. Gani, 2017, A computer-aided software-tool for sustainable process synthesis-intensification, *Computers & Chemical Engineering*, 105, 74–95.
A. K. Tula, M. R. Eden, R. Gani, 2019, Hybrid method and associated tools for synthesis of sustainable process flowsheets, *Computers & Chemical Engineering*, 131, 106572
UN (2021) Sustainable development Goals. (https://www.un.org/sustainabledevelopment/).
V. Venkatasubramanian, 2019, The promise of artificial intelligence in chemical engineering: Is it here, finally?, *AIChE Journal*, 65 (2), 466-478.
R. Vooradi, M. -O. Bertran, R. Frauzem, S. B. Anne, R. Gani, 2018, Sustainable chemical processing and energy-carbon dioxide management: Review of challenges and opportunities, *Chemical Engineering Research and Design*, 131, 440-464.
B. Wang, J. Zhao, 2022. Automatic Frequency Estimation of Contributory Factors for Confined Space Accidents, *Process Safety and Environmental Protection*, 157, 193-207.
WBA, 2020 (http://www.worldbioenergy.org/uploads/201210%20WBA%20GBS%202020.pdf).
WHO, 2019 (https://www.who.int/news-room/fact-sheets/detail/drinking-water).
Worldometers (https://www.worldometers.info/coronavirus/).
H. Wu, J. Zhao, 2020, Fault detection and diagnosis based on transfer learning for multimode chemical Processes, *Computers & Chemical Engineering*, 135, 106731.
Z. J. Yu, L. T. Biegler, 2019, Advanced-step multistage nonlinear model predictive control: Robustness and stability, *Journal of Process Control*, 84, 192–206.
L. Zhang, H. Mao, Q. Liu, R. Gani, 2020, Chemical product design–recent advances and perspectives, *Current Opinion in Chemical Engineering*, 27, 22-34.
S. Zhang, C. Zheng, X. Chen, 2021a, SyPSE: A Symbolic Computation Toolbox for Process Systems Engineering Part I-Architecture and Algorithm Development, *Industrial and Engineering Chemistry Research*, 60 (45), 16304–16316.
S. Zhang, F. Zhao, C. Zheng, L. Zhu, X. Chen, X., 2021b, SyPSE: A Symbolic Computation Toolbox for Process Systems Engineering Part II-Design for PSE Applications, *Industrial and Engineering Chemistry Research*, 60 (45), 16317–16329.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50003-8

PSE Tools and Challenges in the Development of Advanced Pharmaceutical Manufacturing

Yingjie Chen[a], Pooja Bhalode[b], Yang Ou[b], Marianthi Ierapetritou[a,*]

[a]*Department of Chemical and Biomolecular Engineering, 150 Academy St, Newark, DE 19716, USA*
[b]*Department of Chemical and Biochemical Engineering, Rutgers University, 98 Brett Rd, Piscataway, NJ 08854, USA*
[*]*Corresponding author – mgi@udel.edu*

Abstract

Following the Industry 4.0 revolution, pharmaceutical industry is progressing towards embracing its principles for smart manufacturing. Industry 4.0 encourages the application of a robust, integrated data framework to connect physical components to virtual environment. It enables an accurate representation of the physical parts in digitized space, leading to the realization of Digital Twins (DTs). In this work, our effort on developing process systems engineering (PSE) tools towards the development of a DT for advanced pharmaceutical manufacturing are presented. These tools are demonstrated through applications in the areas of solid-based drug manufacturing and biologics production.

Keywords: Industry 4.0; Digital twin; System analysis; Pharmaceutical manufacturing; Biologics production.

1. Introduction

Driven by the Industry 4.0 revolution and the vision to develop agile, robust, and flexible manufacturing process to produce high quality drugs, the pharmaceutical industry is adopting this digitalization move (O'Connor et al., 2016, Chen et al., 2020). Efficient process monitoring, prediction, and analysis are realized using process analytical technologies (PAT), data collection and processing, Internet of Things (IoT), and big data analytics. The framework allows for the establishment of a virtual representation of the physical process with information communications, resulting in a DT capable to enhance process robustness and facilitate process design and operations (Chen et al., 2020). For DTs, maintaining precise virtual representations of processes and conducting detailed analyses are two crucial tasks. These tasks are challenging for pharmaceutical manufacturing as multi-scale information, ranging from powder and cell properties to bulk flow of materials, needs to be integrated with complex reaction networks and transport phenomena. These components lead to complex model development and high computational costs, limiting the implementation of DT in advanced pharmaceutical manufacturing.

To address these challenges, we propose the use of PSE tools focusing on modeling and analysis approaches. From a modeling perspective, mechanistic models for particle and cell level modeling, surrogate and hybrid modeling for model reduction, adaptive modeling for model updates, and flowsheet models for process integration are developed. For process analyses, efficient tools in sensitivity and feasibility analysis, techno-economic analysis (TEA), life cycle assessment (LCA), and optimization are applied. The

development and application of these methods will be illustrated for solid-based drug manufacturing and biological production.

2. Application in solid-based drug and biologics manufacturing

The in-silico design, analysis, and optimization methods provide a basis for digital manufacturing, which support resolving the bottleneck of the pharmaceutical industry in improving productivity and quality. Mechanistic models enable the incorporation of process details and material properties, providing a comprehensive digital replication of the unit. These models include discrete element modelling (DEM) to simulate dynamic powder flow in solid-based drug manufacturing (Bhalode and Ierapetritou 2020), and computational fluid dynamics (CFD) with kinetic models to capture cellular activities for biologics manufacturing (Yang and Ierapetritou, 2021). However, these models can be computationally intensive. To address such challenge, surrogate and hybrid models that combine data and process knowledge in different scales are utilized (Bhalode and Ierapetritou, 2021, Chen and Ierapetritou, 2020, Metta and Ierapetritou, 2019). Dynamic algorithms and adaptive strategies based on moving windows are used to capture time-variant process behaviours, supporting the development of DTs (Bhalode et al., 2022). Models in solid-based drug manufacturing and monoclonal antibody (mAb) production will be presented as case studies (Ding and Ierapetritou, 2021, Yang and Ierapetritou, 2021, Chopda et al., 2021).

Along with unit operation models, flowsheet models are constructed with appropriate information transfer, which facilitate early-stage design, evaluation, and decision making. To improve process understanding, PSE tools such as regression and variance-based sensitivity analysis, feasibility analysis with adaptive sampling, and deterministic optimization are performed for identification of critical process parameters, design space, and optimal operating conditions, respectively. Cases in direct compaction of solid-based drugs (Wang et al., 2017a, Metta et al., 2020, Bhalode et al., 2020) and continuous chromatography of biologics (Ding and Ierapetritou, 2021) will be demonstrated.

TEA tools are integrated with flowsheet models to analyse the cost and energy effectiveness of the process and to identify the benefits of continuous operations over traditional batch or semi-batch operation alternatives. Applications in both wet granulation for the production of solid-based drugs (Sampat et al., 2022), and mAb production will be discussed (Yang et al., 2019). To assess the sustainability potential of advanced pharmaceutical manufacturing processes, LCA tools (Luo and Ierapetritou, 2020) are utilized to obtain important environmental indicators including the global warming potential of the product. With information on sustainability and process economics, multi-objective optimization is performed to strive for a balance between the two and guide process design and operations. To reduce sampling cost, maintain process feasibility, and find accurate Pareto solutions, a two-stage optimization framework based on Wang et al. (2017b) is proposed. The feasibility stage identifies the feasible regions with promising values for all objectives, followed by the optimization stage to find the Pareto within the feasible regions. The surrogate-based feasibility-driven multi-objective optimization algorithm will be shown for wet granulation route of solid-based drug manufacturing.

Acknowledgement

The authors acknowledge funding from U.S. Food and Drug Administration through grants DHHS-FDA-U01FD006487, R01FD006588, and FDABAA-20-00123.

References

P. Bhalode, M. Ierapetrittou, 2020, Discrete element modeling for continuous powder feeding operation: Calibration and system analysis, Int J Pharm, 585, 119427.

P. Bhalode, M. Ierapetritou, 2021, Hybrid multi-zonal compartment modeling for continuous powder blending processes, Int J Pharm, 602, 120643.

P. Bhalode, Y. Chen, M. Ierapetritou, 2022, Hybrid modelling strategies for continuous pharmaceutical manufacturing within digital twin framework, Proceedings of the 14th International Symposium on Process Systems Engineering, Paper 306.

P. Bhalode, N. Metta, Y. Chen, M. Ierapetritou, 2020, Efficient Data-based Methodology for Model Enhancement and Flowsheet Analyses for Continuous Pharmaceutical Manufacturing, Comput Aided Chem Eng, 48, 127-132.

Y. Chen, M. Ierapetritou, 2020, A framework of hybrid model development with identification of plant-model mismatch, AIChE Jounal, 66, 10, e16996.

Y. Chen, O. Yang, C. Sampat, P. Bhalode, R. Ramachandran, M. Ierapetritou, 2020, Digital Twins in Pharmaceutical and Biopharmaceutical Manufacturing: A Literature Review, Process, 8, 9.

V. Chopda, A. Gyorgypal, O. Yang, R. Singh, R. Ramachandran, H. Zhang, G. Tsilomelekis, S. Chundawat, M. Ierapetritou, 2021, Recent advances in integrated process analytical techniques, modeling, and control strategies to enable continuous biomanufacturing of monoclonal antibodies, J Chem Technol Biotechnol.

C. Ding, M. Ierapetritou, 2021, A novel framework of surrogate-based feasibility analysis for establishing design space of twin-column continuous chromatography, Int J Pharm, 609, 121161.

Y. Luo, M. Ierapetritou, 2020, Comparison between Different Hybrid Life Cycle Assessment Methodologies: A Review and Case Study of Biomass-based p-Xylene Production, Ind Eng Chem Res, 59, 52, 22313-22329.

N. Metta, R. Ramachandran, M. Ierapetritou, 2019, A Computationally Efficient Surrogate-Based Reduction of a Multiscale Comill Process Model, J Pharm Innov, 15, 424-444.

N. Metta, R. Ramachandran, M. Ierapetritou, 2020, A novel adaptive sampling based methodology for feasible region identification of compute intensive models using artificial neural network, AIChE J, 67, 2, e17095.

T. O'Connor, L. Yu, S. Lee, 2016, Emerging technology: A key enabler for modernizing pharmaceutical manufacturing and advancing product quality, Int J Pharm, 509, 1-2, 492-498.

C. Sampat, L. Kotamarthy, P. Bhalode, Y. Chen, A. Dan, S. Parvani, Z. Dholakia, B. Glasser, M. Ierapetritou, R. Ramachandran, Enabling energy-efficient manufacturing of pharmaceutical products by implementing a smart manufacturing platform, J Adv Manuf Process.

Z. Wang, M.S. Escotet-Espinoza, M. Ierapetritou, 2017a, Process analysis and optimization of continuous pharmaceutical manufacturing using flowsheet models, Comput Chem Eng, 107, 77-91.

Z. Wang, M.S. Escotet-Espinoza, R. Singh, M. Ierapetritou, 2017b, Surrogate-based Optimization for Pharmaceutical Manufacturing Processes, Comput Aided Chem Eng, 40, 2797-2802.

O. Yang, M. Ierapetritou, 2021, mAb Production Modeling and Design Space Evaluation Including Glycosylation Process, Processes, 9, 2, 324.

O. Yang, M. Qadan, M. Ierapetritou, 2019, Economic Analysis of Batch and Continuous Biopharmaceutical Antibody Production: A Review, J Pharm Innov, 14, 1-19.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50004-X

Experience and Perspectives on our Journey towards Deep Decarbonization

Iftekhar A Karimi* and Shamsuzzaman Farooq*

Department of Chemical & Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117585
cheiak@nus.edu.sg and chesf@nus.edu.sg

Abstract

The enormity of mankind's decarbonization challenge precludes a simple monolithic solution. Its unprecedented scale and complexity affect every nation without exceptions. Ideally, our mission should be to fully replace the fossil fuels with zero-carbon renewable alternatives, but the path to that end is long and challenging. In this talk, we discuss our experience and perspectives on some decarbonization pathways with concrete examples (efficiency improvements, novel processes for CO2 utilization, non-polluting energy sources, …), where successful translational outcomes can be accelerated by continuous guidance from process systems engineering tools and techniques.

Keywords: carbon emissions, decarbonization, hydrogen, carbon utilization, carbon capture, supply chain.

1. Introduction

Economic development and per capita energy consumption of a nation are strongly correlated. Rising world population with a desire for better living has increased energy demand exponentially, which is mostly met with easy-to-use hydrocarbon fuels with little concern for a backlash from abusing the nature and despite warnings flagged by many scientific studies. The consequence is an unprecedented rise in the atmospheric CO2 levels. In the meantime, global warming is already showing early signs of catastrophic consequences all around the world. From Kyoto Protocol in 1997 through Paris Agreement in 2015 to Glasgow COP26 in 2021, it took 23 years for the world to fully wake up to the urgency of the situation and take decarbonization seriously. Fortunately, the scientific community was fully committed well ahead of the political consensus on this matter, and has compiled a significant body of work on decarbonization. These contributions, spanning a broad spectrum of issues, can be classified into three themes: reduce, recycle/reuse, and replace carbon; which are widely known as 3Rs. In most countries, the journey towards decarbonization has involved the 3Rs in the same sequence.

Decarbonizing existing systems require addressing inherently large-scale problems. Consequently, our collective 3R-driven decarbonization journey has witnessed a resurgence of the need to apply the classical Chemical and Process Systems Engineering (PSE) concepts and methods. Parallelly, low carbon technology innovations are required for capturing carbon emissions and producing power, chemicals, and fuels. We believe that their industrial success requires mission oriented translational work, and constant guidance from PSE right from the beginning can benefit them greatly.

In this talk, we present our perspectives on the 3Rs for our journey from high-carbon present to no-carbon future with examples from our work at the National University of Singapore that illustrate the central role of PSE in this urgent and critical mission.

2. Reduce

Reducing CO2 emissions from the power and industry sectors by increasing energy efficiency is a low hanging fruit, hence the obvious first target for many countries and research. Our work has yielded several observations.

Much literature has used energy or exergy as a KPI (Key Performance Index) for many problems. In our opinion, annualized total cost (TAC) should be the preferred KPI, as minimum-TAC solutions can be significantly different from minimum-energy/exergy solutions (Rao et al., 2016; Rao & Karimi, 2018). Furthermore, no doubt a sound theoretical concept, exergy is inadequate for economic industrial decisions in practice.

Rigorous simulation-based optimization has been quite useful for us. Our work (Dutta et al., 2018; Rao & Karimi, 2017) required detailed rigorous simulations of industrial facilities with realistic thermodynamic properties. It proved easier and more accurate to use them directly versus deriving analytical correlations for the conventional algebraic optimization. Hamedi et al. (2020) have identified several pitfalls of the latter.

Saleem et al. (2018) developed a computational fluid dynamics (CFD) model for an industry-scale LNG storage tank and showed that surface evaporation predominates, and nucleate boiling is unlikely. Furthermore, the static pressure delays internal circulation and complete mixing. Sundaram & Karimi (2021) were able to predicts pressure transients using a simpler model and their results matched both the CFD model real tank data. LNG recirculation is a major cause of BOG losses (hence power use in reliquefaction) from these tanks. Using a validated dynamic simulator, we (Karimi et al, 2019) proposed a modified recirculation scheme, a new operations schedule for BOG compressors, and a lower recondenser pressure to reduce power use by nearly 40%.

Heat integration has hugely benefitted the process industry. We addressed several key areas. First, Nair et al. (2019) developed a novel stageless superstructure that offers complete flexibility in network configuration and is seamless for both grassroots and retrofit synthesis. Second, Nair et al. (2016) showed that inter-plant (versus intra-plant) heat integration can offer substantial energy savings, but at the cost of several significant complexity, safety, control, logistics, ownership, and collaboration challenges. Capital costs become much more crucial. Third, integrating work along with heat (Hamdi et al., 2020) offers additional carbon reduction. Fourth, Rao & Karimi (2017) and Rao et al. (2019) addressed the flowsheet and operation optimizations of processes with multi-stream exchangers. Fifth, Christopher et al. (2017) used vapor recompression and self-heat recuperation to decrease energy use by 45% for propane/propylene separation.

LNG regasification terminals worldwide waste LNG's cold energy. Dutta et al. (2018a) showed that a well-configured organic Rankine cycle (ORC) can recover this energy and produce 0.5−12.9 kW/t-LNG of power with an NPV of $2.45-6.87 million at an operating regasification terminal. Furthermore, Dutta et al. (2018b) proposed an integrated regasification-production process to produce valuable heavy hydrocarbons (ethane, LPG) from rich/medium LNG by exploiting the cold energy. In fact, such a process can generate 7-10% profit for the terminal.

We have implemented process data-based pump and insulation health monitoring at the same terminal. While estimating the remaining useful life (RUL) to schedule timely preventive maintenance, the user interface also tracks increasing CO2 emission from deteriorating health. Thus, it is possible to account for environmental impact (carbon tax) to schedule maintenance.

Our above discussion illustrates how PSE methodologies for design, simulation, integration, and optimization can reduce carbon emissions in real industrial processes.

3. Recycle / Reuse

While renewable energy sources are a priority, hydrocarbons will remain in use during the transition to a zero-carbon economy. Hence, carbon capture and concentration (CCC) from various stationary emission sources for sequestration and utilization (CSU) will be important and unavoidable during the transition. CO2 recycle/ utilization can be achieved in two ways. The conventional way is to capture it from the existing processes and produce useful chemicals via green hydrogen. The alternative is to radically change the way we produce power and chemicals today by integrating their productions to achieve zero net CO2 emissions.

Our studies (Khurana & Farooq, 2017, 2019) on CCC have led to the following perspectives. First, the scale and footprint of a CCC plant, irrespective of the technology, are far larger than the largest chemical plant known today. Second, adsorption-based CCC processes (VSA, TSA, etc) do not show any significant cost advantage compared to the established industrial amine processes. Third, the cost reduction from designing both adsorbent and process simultaneously (process inversion) by considering adsorbent attributes also as optimization variables was minimal. This implies that process rather than material innovation is more promising. The idea of process inversion can help other R&DD efforts such as catalysts for CO2 utilization (CU).

In order to gain a broader understanding of the prospects and challenges of CO2 reuse, we (Dutta et al., 2017) conceptualized a zero-emissions scheme (Figure 1) to integrate both power and chemicals production. All CO2 produced was recycled internally and reused to produce useful fuels and chemicals. All reaction and separation operations along with heat and power integration were also built within the scheme. Using the scheme as a surrogate to achieve zero net carbon emissions made us quickly realize that an abundant supply of green renewable hydrogen is a prerequisite. If the scheme was used to meet the current global demands of all chemicals, CO2 avoidance would be at most 59% of 2013 emissions. This implies that CU must produce fuels to increase CO2 avoidance. Furthermore, CU products must be cost-competitive and profitable.

Our recent study (Vo et al., 2021) on the feasibility of producing transportation fuels, methanol and 1-propanol from CU shows that only 1-propanol is competitive at the present levels of carbon tax, CO2 cost, and green hydrogen price. Our proposed scheme for producing 1-propanol from CO2 utilization is a three-step process. Even though this process is economically profitable and emits less CO2 than the current commercial process, a novel catalyst allowing a single step process may seem more attractive intuitively. However, our technoeconomic study based on preliminary data from a single-step catalyst shows contrary results. The first problem is low selectivity towards propanol versus non-alcoholic by-products. The second problem is the high H2 burden arising from the oxygen molecules introduced with each mole of CO2. This is an excellent illustration of how a complete process study at an early stage of catalyst development can guide

further materials research. In view of the current high green H2 cost, its higher consumption in CO2 utilization compared to the conventional routes deserves further study.

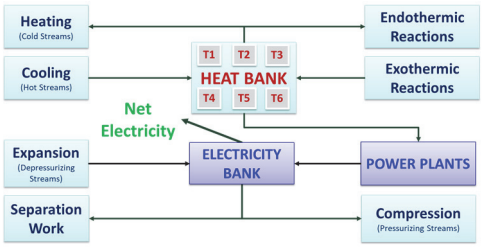


Figure 1: Scheme for the integrated production of power and chemicals with zero net carbon emissions.

4. Replace

The grand vision for deep decarbonisation foresees the emergence of an energy landscape based on global hydrogen supply chains (HSCs, Figure 2). While much effort is underway on technologies (e.g. water electrolysis) to produce green hydrogen via renewable electricity, it will take time before such green H2 is cost competitive with the blue hydrogen produced from NG with CCS. Other material and equipment hurdles along with the lack of infrastructure will keep both NG and CCSU critically relevant for the foreseeable future. In other words, HSCs will be inseparable from NG/LNG and CO2 supply chains. Several countries (e.g. EU, USA, Korea, Australia, Japan) have developed roadmaps for how H2 could develop in specific regions. Since some countries (e.g. Singapore) with high energy needs have few renewable energy resources, export/import of H2 via global transport routes will be a key aspect of HSCs with low density of H2 as a significant hurdle. Hence, countries are asking three main questions: what are the technoeconomic and environmental costs of producing, importing, and transporting H2?

Fully funded by and in collaboration with ExxonMobil USA, we are working on a hydrogen roadmap for ASEAN (Association of South East Asian Nations) under the umbrella of Singapore Energy Center. We conducted a holistic study of various options for producing and transporting hydrogen from both technoeconomic and environmental perspectives, and are analysing various scenarios for the planning of HSCs in ASEAN from now to 2050. For this, we (Hong et al., 2021) have developed a multiperiod mixed-integer nonlinear programming model for geographically distributed SC capacity planning, and a tool called HEART (Hydrogen Economy Assessment & Resource Tool). HEART enables the long-term design and planning of H2 production and import, and computes the cost of producing and transporting H2. In the near future, we will also be collaborating with Chiyoda and Mitsubishi Japan to plan and optimize the local distribution of H2 in Singapore, when MCH (Methyl CycloHexane) is used as a carrier for importing H2.

Using data from various H2 reports, we developed simplified analytical correlations for the cost of producing H2 from NG, coal, biomass, and water electrolysis at various locations in ASEAN. We (Hong et al., 2021) then analysed four options for transporting hydrogen, namely methyl cyclohexane (MCH), liquid hydrogen (LH2), compressed hydrogen (CH2), and liquid ammonia (LNH3) and computed the landed cost of H2 as a function of various technoeconomic and geographic parameters. Our study showed that HSC costs are comparable to production costs, hence both require careful attention for

H2 economy. H2 produced via steam reforming with CCS and supplied via pipelines is the cheapest option for distances within 2000 km. The next best option is the direct use of NH3 instead of H2 for power generation. MCH and liquid NH3 showed comparable landed costs for H2, and using green H2 is essential for achieving >80% carbon avoidance. Liquid H2 has the highest energy penalty. Contrary to the misgiving that direct NH3 burning will increase NOx emissions, our recent comparative study (Saleem et al., 2022) of literature data on burning hydrogen, ammonia, and natural gas shows that NOx generation from H2 combustion is more than that from an equivalent amount of NH3.

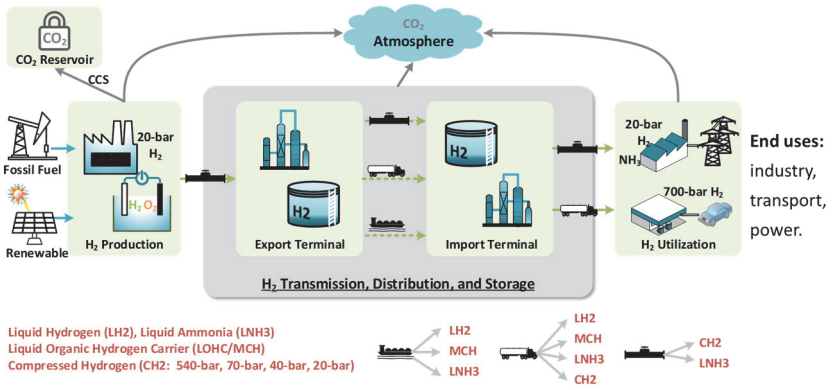


Figure 2: Schematic of hydrogen supply chain and its key elements. (Hong et al., 2021)

5. Conclusions

Various branches of chemical engineering and related sciences have sufficiently progressed where we can have high confidence in simulation-based prediction of a process performance obtained via PSE tools and techniques. In this paper, we have demonstrated with examples drawn from our work how that can play important roles in all three Rs to arrest our existential threat from global warming.

We strongly advocate a paradigm shift where PSE would be the guiding vector for materials research from the beginning in order to provide timely feedback necessary to ensure that what looks exciting at a small scale meets the requirements of a successful commercial translation. This new paradigm has the potential to maximize resource utilization and shorten development times.

Acknowledgements

This work was funded by grants from National University of Singapore, Singapore Energy Center, ExxonMobil Research & Engineering, and Energy Market Authority of Singapore and National Research Foundation of Singapore under their Energy Innovation Research Program.

References

C. Christopher, A. Dutta, S. Farooq, I. Karimi, 2017, Process synthesis and optimization of propylene/propane separation using vapor recompression and self-heating recuperation, Ind. Eng. Chem. Res., 56, 49, 14557-14564.

A. Dutta, S. Farooq, I. Karimi, S. Khan, 2017, Assessing the potential of CO2 utilization with an integrated framework for producing power and chemicals. J of CO2 Utilization, 19, 49-57.

A. Dutta, I. Karimi, S. Farooq, 2018a, Heating Value Reduction of LNG (Liquefied Natural Gas) by Recovering Heavy Hydrocarbons: Technoeconomic Analyses Using Simulation-Based Optimization, Ind. Eng. Chem. Res., 57, 17, 5924-5932.

A. Dutta, I. Karimi, S. Farooq, 2018b, Economic feasibility of power generation by recovering cold energy during LNG (Liquefied Natural Gas) regasification. ACS Sustainable Chemistry & Engineering, 6, 8, 10687-10695.

H. Hamedi, I. Karimi, T Gundersen, 2020, Simulation-based approach for integrating work within heat exchange networks for sub-ambient processes. Energy Conv & Management, 203, 112276.

X. Hong, V. Thaore, I. Karimi, S. Farooq, X. Wang, A. Usadi, B. Chapman, R. Johnson, 2021, Techno-enviro-economic analyses of hydrogen supply chains with an ASEAN case study, Int J of Hydrogen Energy, in press.

I. Karimi, S. Farooq, L. Samavedham, 2019, Final Report for grant NRF2014EWT-EIRP003-008.

M. Khurana, S. Farooq, 2017, Integrated Adsorbent-Process Optimization for Carbon Capture and Concentration Using Vacuum Swing Adsorption Cycles, AIChE Journal 63 (7), 2987–2995.

M. Khurana, S. Farooq, 2017, Integrated Adsorbent-Process Optimization for Carbon Capture and Concentration Using Vacuum Swing Adsorption Cycles, AIChE Journal 63 (7), 2987–2995.

A. Saleem, S. Farooq, I. Karimi, R. Banerjee, 2018, A CFD simulation study of boiling mechanism and BOG generation in a full-scale LNG storage tank, Comp. Chem. Eng., 115, 112-120.

S. Nair, Y Guo, U Mukherjee, I. Karimi and A Elkamel, 2016, Shared and Practical Approach to Conserve Utilities in Eco-Industrial Parks. Comp. Chem. Eng., 93, 221-233.

S. Nair, I. Karimi, 2019, Unified Heat Exchanger Network Synthesis via a Stageless Superstructure". Ind. Eng. Chem. Res., 58, 15, 5984-6001.

H. Rao, I. Karimi, 2017, A Superstructure-based Model for Multistream Heat Exchanger Design within Flowsheet Optimization, AIChE J, 63, 9, 3764-3777.

H. Rao, I Karimi, 2018, Optimal design of boil-off gas reliquefaction process in LNG regasification terminals, Comp Chem Eng, 117, 171-190.

H. Rao, S. Nair, I. Karimi, 2019, Operational Optimization of Processes with Multistream Heat Exchangers". Ind. Eng. Chem. Res., 58, 15, 5838-5850.

H. Rao, K Wong and I Karimi, 2016, Minimizing Power Consumption Related to BOG Reliquefaction in an LNG Regasification Terminal. Ind. Eng. Chem. Res., 55, 27, 7431-7445.

A. Saleem, I. Karimi, S. Farooq, 2022, Estimating NOx emissions of useful two-fuel blends from literature data, Fuels, in press.

A. Sundaram, I. Karimi, 2021, State Transients in storage systems for energy fluids, Comp. Chem. Eng., 144, 107128.

C. Vo, C. Mondelli, H. Hamedi, J. Pérez-Ramírez, S. Farooq, I. Karimi, 2021, Sustainability Assessment of Thermocatalytic Conversion of CO2 to Transportation Fuels, Methanol, and 1-Propanol, ACS Sustainable Chem & Eng, 9, 10591−10600.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50005-1

Surrogate Modeling and Surrogate-Based Optimization with Stochastic Simulations

Samira Mohammadi[‡], Bianca Williams[‡], Selen Cremaschi*

Department of Chemical Engineering, Auburn University, Auburn, AL, United States
**selen-cremaschi@auburn.edu*
[‡]*Authors contributed equally*

Abstract

In recent years, high-fidelity simulation models have become widespread to study, design, and optimize engineering systems. However, the complexity of the models often requires computational power beyond what is feasible. One common way to reduce computational cost is to use surrogate/meta models, simplified approximations of more complex, higher-order models, and map input data to output data. Many techniques have been developed for surrogate modeling; however, there remains a need for a systematic method for selecting suitable techniques. In addition, surrogate models built using traditional methods do not accurately represent the outputs of high-fidelity stochastic simulations, e.g., simulations with uncertain parameters. This work describes a new framework that combines PRESTO (Predictive REcommendations of Surrogate models To Optimize), a surrogate model selection tool, with a novel technique, PARIN (PARameter as INput-variable), for building accurate surrogate models of stochastic simulations. We applied the framework to two stochastic test functions with one uncertain parameter. The results reveal that the framework yielded lower normalized root mean square errors than stochastic kriging in predicting the mean and standard deviation of the test function outputs.

Keywords: surrogate model, stochastic simulation, surface approximation, process design/optimization

1. Introduction

In recent years, high-fidelity simulation models have become widespread to study, design, and optimize engineering systems (e.g., (Burnak et al., 2019; Al et al., 2020; Marvi-Mashhadi et al., 2020; Wang et al., 2020)). However, in many cases, the complexity of the models requires computational power beyond what is available for applications like sensitivity analysis or optimization studies. One common way to reduce computational costs is to use surrogate/meta models. Surrogate models, also known as response surfaces, black-box models, metamodels, or emulators, are simplified approximations of more complex, higher-order models and map input data to output data (Jiang et al., 2020).

With all the surrogate modeling techniques currently available, there is a need for a systematic procedure for selecting the appropriate technique. Recent works (Cui et al., 2016; Garud et al., 2018; Jia et al., 2020) have made progress in generalizing the selection of a surrogate model to approximate a design space by using meta-learning approaches avoiding expensive trial-and-error methods. However, selecting surrogate models for surrogate-based optimization remains an open challenge. Furthermore, surrogate models built using traditional techniques do not accurately represent the outputs of high-fidelity

stochastic simulations, e.g., simulations with uncertain parameters (Staum, 2009). High-fidelity simulations may contain different sources of uncertainty, including uncertain inputs, parameters, and model form (Ankenman et al., 2008). The existing machine learning (ML) methods are mainly built for deterministic problems, and they usually fail in representing the stochastic simulation outputs properly (Staum, 2009). The current ML techniques can be used to model outputs of stochastic simulation only when the source of uncertainty is the inputs. A surrogate model is built using an appropriate ML technique with the input/output data generated using the deterministic simulation. Then, the input uncertainty may be propagated to the ML model outputs using uncertainty propagation methods (Kim, 2016).

Three current approaches to model the stochastic simulations with uncertain parameter(s) are 1) fixing the value of the uncertain parameters (Hüllen et al., 2019), 2) using a subset of realizations of the uncertain parameters (Hüllen et al., 2019), and 3) stochastic kriging (Ankenman et al., 2008). Using the first method leads to a deterministic output and the loss of uncertainty. The second method requires training several surrogate models to estimate the output and its uncertainty. Applying stochastic kriging limits the ML technique to kriging; however, it has been shown that the best ML technique for building a surrogate model depends on the data characteristics, which is a function of the underlying phenomena the model represents (Williams and Cremaschi, 2021).

In this paper, we propose a new framework to address the shortcomings of the previous methods and build surrogate models of simulations with uncertain parameter(s). The approach combines our recently developed surrogate model recommendation tool (PRESTO - Predictive REcommendations of Surrogate models To Optimize) with a novel approach, PARIN (PARameter as INput-variable), for building accurate surrogate models of stochastic simulations with uncertain parameters (Section 2). Computational studies use two test functions with different dimensions to evaluate the new framework (Section 3). The training data set is generated using Sobol sampling methods, and then the output for each given input point is calculated. The resultant training data set is fed to the PRESTO to obtain a set of recommended ML techniques for the data set. The recommended models are trained for each of the test functions. The performance of the new framework is compared to stochastic kriging (Section 3). Normalized root mean square error is used as the metric to compare the accuracy of the mean and standard deviation estimations of the test function outputs (Section 3). The comparisons reveal that the mean and standard deviation estimates obtained by the new framework are closer to their true values than the ones obtained using stochastic kriging. These results are presented and discussed in Section 4, followed by conclusions in Section 5.

2. Framework for Training Surrogate Models for Stochastic Simulations

The framework (Figure 1) starts by collecting input/output data, i.e., the training data set, from the stochastic simulation using a space-filling design. PARIN, PARameter as INput-variable, converts the stochastic simulation into a deterministic one by extracting its uncertain parameters and considering these parameters as additional inputs to the simulation (Section 2.2). Therefore, the input data set also includes the uncertain parameters. PRESTO (Section 2.1) selects the best surrogate modeling technique given the training data set, and a surrogate model is trained using the selected technique. The parameter uncertainty is incorporated into the surrogate model outputs utilizing an appropriate uncertainty propagation method (Section 2.2).

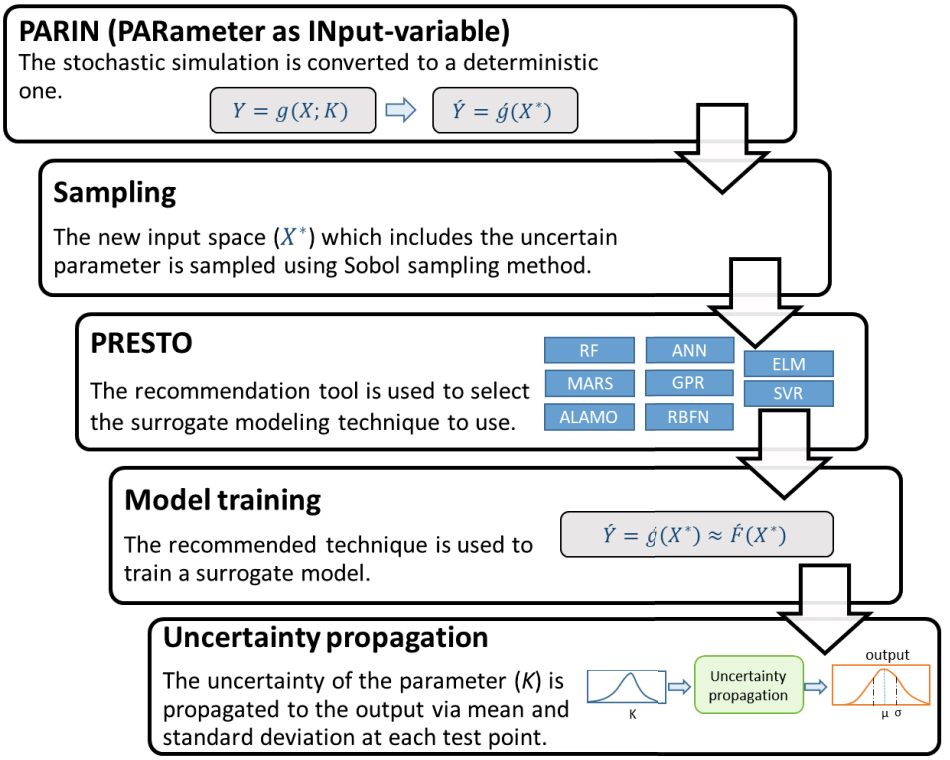


Figure 1 – Workflow of the new framework

2.1. PRESTO - Predictive REcommendations of Surrogate models To approximate and Optimize

PRESTO is a random forest-based framework for selecting appropriate surrogate modeling techniques (Williams et al., 2021). Given a dataset of input-output values, it provides a list of surrogate models predicted to give the most accurate surface approximation of the underlying model of the given input-output pairs. In order to collect the data required to construct PRESTO, data sets were generated using test functions from the optimization test suite of the Virtual Library of Simulation experiments (Surjanovic and Bingham, 2013) at various input dimensions and sample sizes. Surrogate models were trained for each generated data set using a set of candidate surrogate modeling techniques. The calculated performance of each model was used to determine if a surrogate model was appropriate ("recommended") or not ("not recommended") for modeling a particular data set. This information was used to train random forest classifiers using data set characteristics as inputs and the assigned recommendation class as outputs (Williams et al., 2021).

PRESTO recommends models based on the prediction of their performance for adjusted-R^2. The formula for calculating adjusted-R^2 ($\hat{R}^2$) is shown in Eq. (1).

$$\hat{R}^2 = 1 - (1 - R^2)\left[\frac{n-1}{n-(k+1)}\right] \quad (1)$$

In Eq. (1), R^2 is the R-squared regression coefficient, n is the number of data points in the training set, and k is the number of model parameters (or hyperparameters). The adjusted-R^2 takes into account both the surrogate model accuracy and complexity. Taking complexity into account is important in preventing overfitting of the model as overfit models do not generalize well to data outside of the training set. R^2 values typically fall between zero and one, with an R^2 of one indicating an exact fit to the original, more complex model data. However, with the adjustment for model size, adjusted-R^2 values can become negative.

PRESTO calculates characteristics, or attributes, of the underlying model using only the available input-output information. These attributes include calculations related to the location and distribution of the data points and estimations of the gradients of the surface. Based on the values of the data set attributes, PRESTO classifies each of a set of candidate surrogate model forms as being either "recommended" or "not recommended" for that data set (Williams et al., 2021).

2.2. PARIN - PARameter as INput-variable: A novel approach for surrogate modeling of stochastic simulations

PARIN builds surrogate models of stochastic simulations by considering the uncertain parameter(s) as additional inputs to the system. Suppose we are given a stochastic simulation, $Y = g(X; K)$, where X is the input vector, K is the uncertain parameter, and Y is the stochastic output (Figure 2). PARIN converts the stochastic simulation to a deterministic one, $Y' = \acute{g}(X^*)$, where the vector X^* now also includes the uncertain parameter K (Figure 2), and the output of the deterministic model is denoted Y'. A surrogate model, $\widehat{Y}' = F'(X^*)$, is trained to predict the deterministic simulation output, Y', with data generated from the deterministic simulation, $Y' = \acute{g}(X^*)$. Different surrogate modeling techniques can be employed to train $F'(X^*)$. We select the surrogate modeling technique using PRESTO (Section 2.1). Then, the stochastic output of the original simulation model is estimated by propagating the uncertainty of the parameter K to the surrogate model output. Here, again, a number of uncertainty propagation methods can be employed.

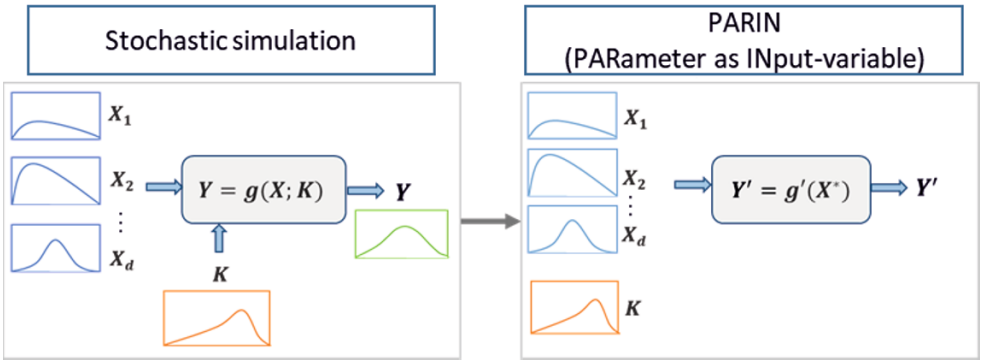


Figure 2 – PARIN (PARameter as INput-variable) - A novel approach for building surrogate models of stochastic simulations

3. Computational Experiments

We evaluate the performance of the framework via computational experiments using two test functions, Griewank and Rastrigin functions given in Eqs. (2) and (3), from the

Virtual Library of Simulation Experiments optimization test suite (Surjanovic and Bingham, 2013). Each test function includes one uncertain parameter and one, two, or four inputs. A normal distribution is assumed for the uncertain parameter (p) of each function. Variable d in both functions denotes the input dimension, which can be changed. The dimension is increased from one to four geometrically to investigate the impact of the input dimension on the performance of the proposed framework.

$$f(x) = \sum_{i=1}^{d} \frac{x_i^2}{p} - \prod_{i=1}^{d} cos\left(\frac{x_i}{\sqrt{i}}\right) + 1 \tag{2}$$

where,

$$p \sim Normal(4000, 400)$$

$$f(x) = pd + \sum_{i=1}^{d} \left[x_i^2 - p\left(cos(2\pi x_i)\right)\right] \tag{3}$$

where,

$$p \sim Normal(10, 1)$$

The training data set for each test function included the inputs, the uncertain parameter, and the corresponding function values for 1000 evaluations generated using Sobol sampling (Sobol', 1967). We used Sobol sampling because it has been shown to yield robust results in comparative studies of uncertainty propagation methods (Mohammadi and Cremaschi, 2019). PRESTO is used to select the best surrogate modeling technique for each training data set. A surrogate model is trained using the selected technique. During training, the hyperparameters of the models are optimized using 5-fold cross-validation (Wong, 2015).

The performance is evaluated using 10,000 test points sampled using Halton method (Halton, 1960). The parameter uncertainty is propagated to the output of each of these test points using a simulation-based method using Halton sampling. With $m = 1000$ points sampled from the uncertain parameter distribution, the mean (μ_l) and standard deviation (σ_l) of the l^{th} test point is calculated using Eq. (4) and Eq. (5), respectively.

$$\mu_l = E[F'(X_l^*)] \approx \frac{1}{m}\sum_{j=1}^{m} h(x_j) \tag{4}$$

$$\sigma_l^2 = (E[F'(X_l^*)^2] - E[F'(X_l^*)]) \approx \frac{1}{m}\sum_{j=1}^{m} h\left(x_j\right)^2 - \mu^2 \tag{5}$$

3.1. Metric for evaluating framework's performance

The metric used to evaluate the performance of the new framework is the normalized Root Mean Squared Error (nRMSE), shown in Eq. (6). The metric is calculated using 10,000 test points generated by the Halton sampling method (Halton, 1960). The framework's ability to estimate the mean and standard deviation of the stochastic simulation output is assessed using nRMSE for each test function.

$$nRMSE = \left(\sqrt{\frac{\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2}{N}} \Big/ Y_{max} - Y_{min} \right) \quad (6)$$

In Eq. (6), Y_i and $\hat{Y}_i$ are the true and predicted values of the desired moments (mean/standard deviation), respectively, for i^{th}test point. N is the total number of test points. Y_{max} and Y_{min} are the maximum and minimum values of the true moment values within the test set, respectively.

3.2. Stochastic kriging

Among three existing approaches for building surrogate models of stochastic simulations, fixing the value of the uncertain parameter does not provide information about the output uncertainty. The second approach, building multiple surrogate models, each built using a value from a select subset of parameter values, requires training multiple models, which introduces additional uncertainty, and may become computationally expensive as the number of surrogate models increases. Hence, here, the accuracy of the proposed framework is only compared to that of stochastic kriging, described briefly below. The performance metric, nRMSE, is also used to assess the stochastic kriging model's ability to estimate the mean and standard deviation of the stochastic simulation output.

Stochastic kriging (SK) (Ankenman et al., 2008), which extends classic kriging (Rasmussen and Williams, 2005), was initially proposed to account for the intrinsic variance in addition to the extrinsic one. The prediction of an SK model, $\hat{\mathbf{y}}$, can be represented by Eq. (7), where $\sum_M$ is the covariance matrix across all sample points, $\sum_M (x, .)$ is the covariance vector consisting of the covariance of the point x and other points, and $\sum_\varepsilon$ is the diagonal covariance matrix of the intrinsic uncertainty. β_0 is the unknown parameter estimated by the maximum likelihood.

$$\hat{\boldsymbol{y}} = \boldsymbol{\beta_0} + \sum_{\boldsymbol{M}} (\boldsymbol{x}, .)^{\boldsymbol{T}} [\sum_{\boldsymbol{M}} + \sum_{\boldsymbol{\varepsilon}}]^{-\mathbf{1}} (\bar{\boldsymbol{y}} - \boldsymbol{\beta_0} \mathbf{1}_{\boldsymbol{k}}) \quad (7)$$

The training data set for SK included 1000 input/output pairs for each test function at each input dimension. The set is constructed by evaluating the test functions for 100 input values at ten samples from the uncertain parameter distribution. Both input values and the samples from the distribution were determined using Sobol sampling.

4. Results and Discussion

4.1. Selected Surrogate Modeling Techniques by PRESTO

The candidate surrogate model techniques considered in this study included single hidden layer artificial neural network (ANN) models (Haykin, 2009), extreme learning machines (ELM) (Huang et al., 2006), Gaussian process regression (GPR) (Rasmussen and Nickisch, 2010), multivariate adaptive regression splines (MARS) (Friedman, 1991), random forests (Breiman, 2001), and support vector regression (Smola and Scholkopf, 2004). The models selected for the test functions at each studied input dimension are listed in Table 1.

Figures 3 and 4 show the average adjusted-R^2 value for the trained surrogate models that PRESTO recommended compared to the average adjusted-R^2 value of the models that PRESTO did not recommend for the Griewank and Rastrigin functions. The models trained using the recommended techniques for both functions have higher adjusted R^2

values and thus better predictive capability than the not recommended ones. The difference is more pronounced for higher dimensions and the Rastrigin function than the Griewank function.

Table 1 – PRESTO selected models for Griewank and Rastrigin functions

	Recommended Surrogate Models		
	1D	2D	4D
Griewank	ANN ELM **GPR** MARS	ANN ELM **GPR** MARS	**GPR** MARS
Rastrigin	GPR **MARS**	GPR **MARS**	**MARS**

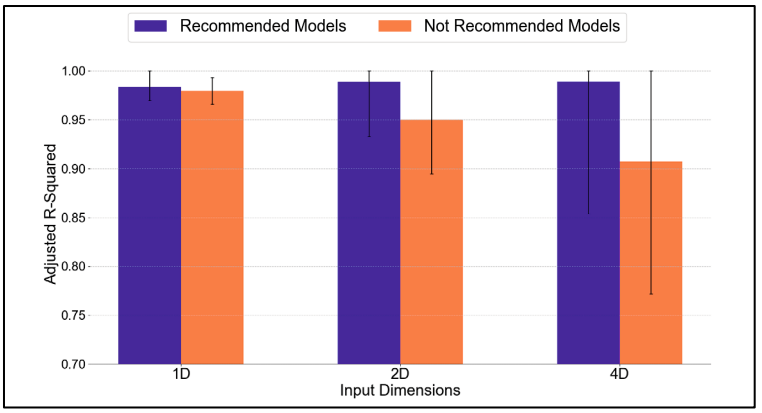


Figure 3 – Average Adjusted R^2 values for Griewank Function. Error bars represent +/- one standard deviation.

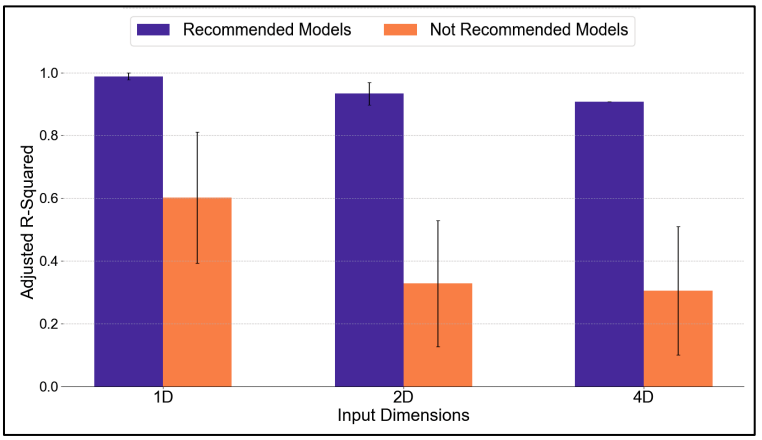


Figure 4 – Average Adjusted R^2 values for Rastrigin Function. Error bars represent +/- one standard deviation.

4.2. Surrogate modeling of stochastic simulations using PARIN

We selected GPR and MARS techniques (both recommended by PRESTO, bold in Table 1) for training surrogate models for deterministic Griewank and Rastrigin functions at each input dimension. The deterministic functions are obtained by including the uncertain parameter as an additional input for each function. We also trained SK models for the test functions. The mean and standard deviation of the test points are estimated using the GPR and MARS surrogate models with Halton-sampling-based uncertainty propagation (Section 3) and the SK models (Section 3.1). Then, the nRMSEs of the mean and standard deviation are calculated via Eq. (6). The results are summarized in Figures 5 and 6.

Figure 5 includes bar plots of the nRMSE obtained using the new framework and SK models for estimating the mean of the two test function outputs. For the Grienwank function, the nRMSE obtained by the new framework for estimating the mean is lower than the nRMSE obtained by SK models for all dimensions (Figure 5). However, the nRMSE yielded for the mean estimates by the SK models for one and two inputs are lower than the nRMSE yielded by the new framework for the Rastrigin function though the difference is relatively small (Figure 5). As shown in Figure 5, the trend is reversed for the Rastrigin function with four inputs with the new framework yielding a significantly lower nRMSE value for estimating the mean of the output.

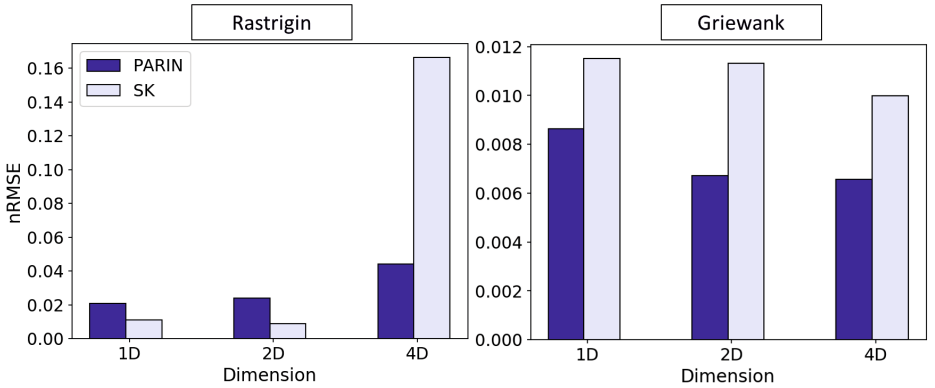


Figure 5 - Bar plots of nRMSE for predicting mean of the two test function outputs using the new framework and stochastic kriging (SK)

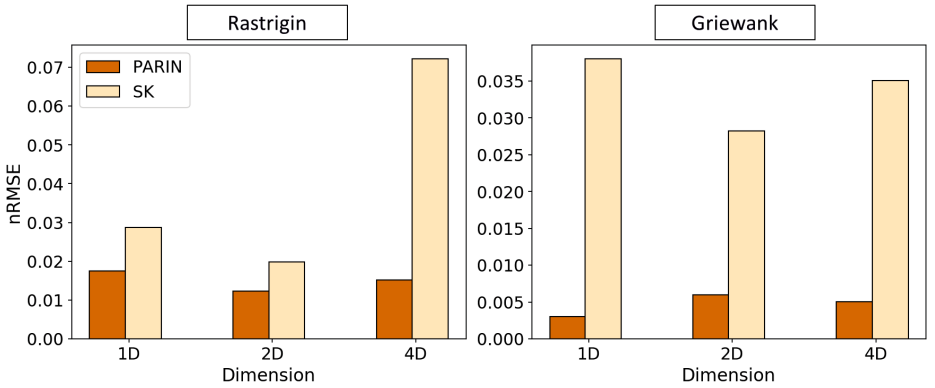


Figure 6 - Bar plots of nRMSE for predicting standard deviation of the two test function outputs using the new framework and stochastic kriging (SK)

Figure 6 presents similarly-formatted bar plots of the nRMSE for estimating the standard deviation of the test function outputs. As shown in Figure 6, for both functions and input dimensions, the nRMSEs of the standard deviation estimates obtained by the new method are lower than those obtained by the SK models. Figure 6 reveals that the difference is larger for the Grienwank function and at higher dimensions for both functions.

Overall, these results suggest that the new framework generates closer mean and standard deviation estimates of the output to the true values (i.e., the stochastic simulation outputs) compared to the same obtained by SK. Furthermore, it is better than SK in capturing the uncertainty of the output due to parameter uncertainty.

5. Conclusions

High-fidelity simulations are complicated and expensive to run. Surrogate models are used to represent these simulations with cheaper to evaluate functions. However, most of the existing surrogate modeling techniques cannot model the stochastic simulation outputs with high accuracy, and the current methods do not capture the uncertainty of the output. This study introduces a new framework to build surrogate models of stochastic simulations where the source of stochasticity is the uncertain model parameters. The framework converts the stochastic simulation to a deterministic one by incorporating uncertain parameters as model inputs (PARIN - PARameter as INput-variable) and uses PRESTO, Predictive REcommendations of Surrogate models To Optimize, to select the best modeling technique for training surrogate models. Comparing the new framework to stochastic kriging, the most popular method to train surrogate models for stochastic simulations, revealed that the framework yielded output mean and standard deviation estimates closer to the true values than those obtained by SK.

References

Al, R., Behera, C.R., Gernaey, K.V., and Sin, G. (2020). Stochastic simulation-based superstructure optimization framework for process synthesis and design under uncertainty. Computers & Chemical Engineering 143. doi: ARTN 10711810.1016/j.compchemeng.2020.107118.

Ankenman, B., Nelson, B.L., Staum, J., 2008. Stochastic kriging for simulation metamodeling. Proc. - Winter Simul. Conf. 362–370. https://doi.org/10.1109/WSC.2008.4736089

Breiman, L. (2001). Random forests. Machine Learning 45(1), 5-32. doi: Doi 10.1023/A:1010933404324.

Burnak, B., Diangelakis, N.A., Katz, J., and Pistikopoulos, E.N. (2019). Integrated process design, scheduling, and control using multiparametric programming. Computers & Chemical Engineering 125, 164-184. doi: 10.1016/j.compchemeng.2019.03.004.

Cui, C., Hu, M.Q., Weir, J.D., and Wu, T. (2016). A recommendation system for meta-modeling: A meta-learning based approach. Expert Systems with Applications 46, 33-44. doi: 10.1016/j.eswa.2015.10.021.

Friedman, J.H. (1991). Multivariate Adaptive Regression Splines - Rejoinder. Annals of Statistics 19(1), 123-141. doi: DOI 10.1214/aos/1176347973.

Garud, S.S., Karimi, I.A., and Kraft, M. (2018). LEAPS2: Learning based Evolutionary Assistive Paradigm for Surrogate Selection. Computers & Chemical Engineering 119, 352-370. doi: 10.1016/j.compchemeng.2018.09.008.

Halton, J.H., 1960. On the efficiency of certain quasi-random sequences of points in evaluating

multi-dimensional integrals. Numer. Math. 2, 84–90. https://doi.org/10.1007/BF01386213

Haykin, S. (2009). Neural Networks and Learning Machines. Upper Saddle River, New Jersey: Pearson Education, Inc.

Huang, G.B., Zhu, Q.Y., and Siew, C.K. (2006). Extreme learning machine: Theory and applications. Neurocomputing 70(1-3), 489-501. doi: 10.1016/j.neucom.2005.12.126.

Hüllen, G., Zhai, J., Kim, S.H., Sinha, A., Realff, M.J., Boukouvala, F., 2019. Managing Uncertainty in Data-Driven Simulation-Based Optimization. Comput. Chem. Eng. 106519. https://doi.org/10.1016/j.compchemeng.2019.106519

Jia, L.Y., Alizadeh, R., Hao, J., Wang, G.X., Allen, J.K., and Mistree, F. (2020). A rule-based method for automated surrogate model selection. Advanced Engineering Informatics 45. doi: ARTN 10112310.1016/j.aei.2020.101123.

Jiang, P., Zhou, Q., and Shao, X. (2020). Surrogate Model-Based Engineering Design and Optimization. Surrogate Model-Based Engineering Design and Optimization, 1-240. doi: 10.1007/978-981-15-0731-1.

Marvi-Mashhadi, M., Lopes, C.S., and LLorca, J. (2020). High fidelity simulation of the mechanical behavior of closed-cell polyurethane foams. Journal of the Mechanics and Physics of Solids 135. doi: ARTN 10381410.1016/j.jmps.2019.103814.

Mohammadi, S., Cremaschi, S., 2019. Efficiency of Uncertainty Propagation Methods for Estimating Output Moments, in: Muñoz, S.G., Laird, C.D., Realff, M.J. (Eds.), Proceedings of the 9th International Conference on Foundations of Computer-Aided Process Design, Computer Aided Chemical Engineering. Elsevier, pp. 487–492. https://doi.org/https://doi.org/10.1016/B978-0-12-818597-1.50078-3

Rasmussen, C.E., and Nickisch, H. (2010). Gaussian Processes for Machine Learning (GPML) Toolbox. Journal of Machine Learning Research 11, 3011-3015.

Smola, A.J., and Scholkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing 14(3), 199-222. doi: Doi 10.1023/B:Stco.0000035301.49549.88.

Sobol', I.M., 1967. On the distribution of points in a cube and the approximate evaluation of integrals. Zhurnal Vychislitel'noi Mat. i Mat. Fiz. 7, 784–802.

Staum, J., 2009. Better Simulation Metamodeling: The why, what, and how of Stochastic Kriging 119–133.

Surjanovic, S., and Bingham, D. (2013). Virtual Library of Simulation Experiments [Online]. Simon Fraser University. Available: http://www.sfu.ca/~ssurjano. [Accessed 2018].

Wang, L.N., Chen, X.Y., Kang, S.K., Deng, X.W., and Jin, R. (2020). Meta-modeling of high-fidelity FEA simulation for efficient product and process design in additive manufacturing. Additive Manufacturing 35. doi: ARTN 10121110.1016/j.addma.2020.101211.

Williams, B., and Cremaschi, S. (2021). Selection of surrogate modeling techniques for surface approximation and surrogate-based optimization. Chemical Engineering Research & Design 170, 76-89. doi: 10.1016/j.cherd.2021.03.028.

Williams, B., Otashu, J., Leyland, S., Eden, M., and Cremaschi, S. (2021). PRESTO: Predictive REcommendation of Surrogate models To approximate and Optimize. Chemical Engineering Science. doi: https://doi.org/10.1016/j.ces.2021.117360.

Wong, T.T., 2015. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. Pattern Recognit. 48, 2839–2846 https://doi.org/10.1016/j.patcog.2015.03.009

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50006-3

Application of PSE into social changes: biomass-based production, recycling systems, and regional systems design and assessment

Yasunori Kikuchi[a,b,c*]

[a] *Institute for Future Initiatives, The University of Tokyo, Tokyo 113-8654, Japan*
[b] *Presidential Endowed Chair for "Platinum Society", The University of Tokyo, Tokyo 113-8656, Japan*
[c] *Department of Chemical System Engineering, The University of Tokyo, Tokyo 113-8656, Japan*
ykikuchi@ifi.u-tokyo.ac.jp

Abstract

Towards the goal of zero fossil-based greenhouse gas emissions, a trend is growing to change the raw materials for energy and materials to those derived from renewable sources. When considering the introduction of any technology, the basics of PSE, i.e., mathematical modelling and simulation of changes to understand the impact on mass and heat balances, are essential for appropriate technology and system assessments including life cycle assessment (LCA). In this study, the role of PSE is discussed through case studies in the assessment of several technologies and systems under consideration, such as cellulose nanofibers reinforced plastics (CNFRP), recycling of lithium-ion batteries (LIB), and regional material and energy systems design in Tanegashima. Although the technology options for those issues are under development, the performances of systems applying them are necessitated for current decision making. The data for LCA, however, is not sufficiently collected due to their low technology readiness levels. Prospective LCA for such emerging technologies is employed in the filling of data gaps and interpretation of assessment results with uncertainties. PSE can be applied into such assessments and have an important role of design of systems.

Keywords: life cycle assessment, technology readiness level, sociotechnical analysis, socioeconomical analysis, technoeconomic analysis.

1. Introduction

Towards the goal of zero fossil-based greenhouse gas emissions, a trend is growing to change the raw materials for energy and materials to those derived from renewable sources. When considering the introduction of any technology, the basics of PSE, i.e., mathematical modelling and simulation of changes to understand the impact on mass and heat balances, are essential for appropriate technology and system assessments including life cycle assessment (LCA).

In this study, the role of PSE is discussed through case studies in the assessment of several technologies and systems under consideration, such as cellulose nanofibers reinforced plastics (CNFRP), recycling of lithium-ion batteries (LIB), and regional material and energy systems design in Tanegashima. The related previous literatures are briefly reviewed. The applicability of PSE basics is discussed considering the requirements for the technology and systems design and assessments towards social changes. Although the

technology options for those issues are under development, the performances of systems applying them are necessitated for current decision making. The data for LCA, however, is not sufficiently collected due to their low technology readiness levels (TRLs).

2. Reviews on application of PSE

2.1. CNFRP production from lignocellulosic biomass (Kanematsu et al., 2021)

Cellulose nanofibers (CNF) can be produced from plant-derived renewable resources and have advantage of mechanical properties in lightness and strength when it was applied as the filler of the composites. Acetylated cellulose nanofiber-reinforced plastics (AcCNF-RP) have been developed as substitutes for conventional structural materials (Eichhorn et al., 2010). CAPE tools enabled simulation-based life cycle inventory analysis to reveal the environmental and economic performance of AcCNF-RP considering the future scale-up of production processes. CAPE tools have huge potentials for systems design and assessment adopting emerging technologies, which are necessitated towards carbon neutral society. Especially in chemical production, biomass-derived production can become one of the production routes with sustainable feedstocks. Not only conversion routes, but also the acquisitions of feedstocks from agriculture or forestry are now under development and construction. Before their huge installation, CAPE tools should be combined with prospective LCA to visualize the performances of such low TRL emerging technologies.

2.2. LiB recycling systems (Kikuchi et al., 2021)

When designing the target recycling systems, best mixture of physical segregation and chemical treatment should be pursued considering the specific characteristics of respective components. The recycling of cathode particles and aluminum (Al) foil from positive electrode sheet (PE sheet) dismantled from spent LiBs was experimentally demonstrated by applying a high-voltage pulsed discharge (Tokoro et al., 2021). This separation of LIB components by pulsed discharge was examined by means of prospective LCA (Kikuchi et al., 2021). The indicators selected were life cycle greenhouse gas (LC-GHG) emissions and life cycle resource consumption potential (LC-RCP). CAPE tools can become methods applicable for acquiring data for prospective assessments. Prospective LCA should be applied into the technology assessment that employs modelling tools which focus on potential environmental impacts arising from various technologies even still at the R&D stage, i.e., low technology readiness level. With CAPE tools, the inventory data for prospective LCA can be connected with the design methods for optimizing the throughputs of unit operations, analyzing the upscaled process systems, and conducting the quantification of environmental loads with plausible process systems design.

2.3. Regional systems design in Tanegashima (Kikuchi et al., 2020)

Well-coordinated, multifaceted actions, including a shift from imported fossil to locally available renewable resources and empowering of rural areas are vital in tackling the social challenges such as resource security, sustainable food production, and forest management. Co-learning approach to practice the multifaceted actions with a case study on Tanegashima, an isolated Japanese island, was applied to move the society towards sustainability. In these actions, thorough understandings in the feasible technologies, the locally available resources and the socioeconomic aspects of the local community should be shared among the stakeholders to acquire the momentum for a change. In addition to

the technoeconomic analysis, several other analyses were conducted to reveal the concerns of respective stakeholders, share the understandings on the possibilities of technology options, and their socioeconomic implications on local sustainability. Tools such as the life-cycle assessment, input–output analysis, and choice experiments based on questionnaire surveys on the residents' preferences are used for the analyses. The stakeholders were provided with the results. These opportunities gradually converted the concerns of the local stakeholders on their future regional energy systems into expectations and yielded constructive alternatives in technology implementation that can use the locally available resources. PSE basics were employed in the simulation and visualization of the possible future visions achieved by feasible technologies and available resources.

3. Application of PSE into social changes

3.1. Arguments for social changes

3.1.1. Design and assessment considering TRL

Novel technologies, including processes, systems, and ways of thinking, are expected to play a critical role in transforming regional societies to become revitalized and sustainable. However, technology development has the "valley of death" in the transfer to society, as is often seen for various energy technologies (e.g., Weyant 2011). Although many types of subsidies are designed to bridge the valley by accelerating technology development based on the TRL (e.g., Debois et al. 2015), the public may perceive that an insufficiently mature technology, or the new installation of existing technology even if it has previous implementation examples in other regions, could have unpredicted consequences associated with its implementation in their regional societies, resulting in the creation of a difficult obstacle to overcome for innovative change in social systems (Weyant 2011).

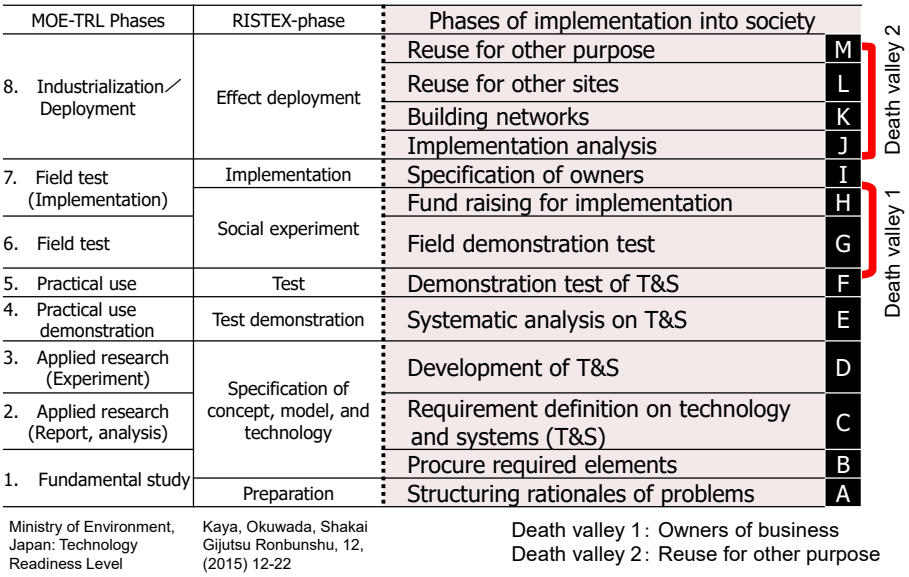


Figure 1 Phases of social implementation considering the TRLs adopted in government subsidized projects towards decarbonization in Japan (Ministry of Environment Japan, 2014) and surveyed on the projects funded by RISTEX (Research Institute of Science and Technology for Society, Japan Science and Technology Agency) (Kaya and Okuwada, 2015).

Figure 1 shows the phases of social implementation considering various types of TRLs. Especially in energy-related technologies that mitigate fossil resource consumption, decentralized and multiple-generation technologies are often seen as promising, but the barriers in progressing to demonstration tests are often too high, hindering the implementation of such technologies. Although technology road-mapping has become a method to address such obstacles by making the effects of technology implementation qualitatively or quantitatively transparent, technology road-mapping has limited roles in practical technology implementation. Appropriate technology and systems design and assessments could support the progressing phases of social implementation.

3.1.2. Prospective LCA for emerging technology

Conventional LCA does not take into account changes in technology level, because it refers to information on the current technology level and specifically estimates the environmental impacts of each process related to the provision of products and services. The significance of conducting a strategic LCA of emerging technologies for the 30-year time horizon up to the target year of 2050 arose regarding the issues on the climate change. Emerging technologies, as defined by Rotolo et al. (2015), are characterized as "innovative", "rapid growth", "consistent", "significant impact" and "uncertain", which makes technology assessment difficult due to lack of existing data and knowledge. Four main issues were identified as needing to be addressed in conducting prospective LCAs of emerging technologies (Thonemann et al., 2020; Moni et al., 2020). (1) comparability of technologies; (2) availability and quality of data; (3) scale-up challenges; and (4) uncertainty of assessment results. Process modeling and simulation are effective in estimating the missing process inventories in industrial scale production, because these technologies are under development in lab or pilot scale.

3.1.3. Technoeconomic, socioeconomic, and sociotechnical analyses

Elements of technology assessment that have been proposed for implementation include the shift in social systems such as the relationships between the socio-, econo-, and techno- spheres through transformation in aviation systems (Kikuchi et al., 2020b). Economic aspects of technology implementation have been examined in technoeconomic (TE) analyses to clarify the relationships between the characteristics of technologies and various economic indicators, such as direct and indirect costs, fixed capital investment, and product price. Socioeconomic (SE) analysis has also become an essential method for analyzing the impacts of technology implementation on SE systems. The benefits should be analyzed within a sociotechnical (ST) approach to ensure that society benefits from the technology implementation.

3.1.4. Social changes with process systems design and assessments

Geels and Schot (2007) argue that transitions occur through interactions among niche innovations, sociotechnical regimes, and the sociotechnical landscape. The seeds of niche innovations were generated by university researchers, e.g., AcCNF-RP for structural materials, a high-voltage pulsed discharge as physical separation methods for products, and energy systems applying regionally available renewables. The windows of opportunity for such seeds of niche innovation are created by the destabilization of regimes such as the policy/regulation, market, infrastructure, industrial network and ecosystem as the specific conditions for technologies and systems. The landscape, such as the public movements towards carbon neutral society, may have placed pressure on the regimes. To grow the seeds of niche innovation, niche actors should be involved and motivated by technology assessments by university researchers, triggering adjustments in existing systems (Geels et al. 2017).

The social embeddedness of emerging technology options should be addressed through the holistic application of scientific technology assessments into co-learning. The main questions are whether systematic technology assessments could contribute to the bridging of the valley of death between research development and actual implementation, how the settings of assessment, i.e., boundary, indicators, and raw data, could be defined through co-learning for mitigating concerns of stakeholders, and how the assessment results could become informative for the stakeholders to understand the necessity of the implementation of the technology options. The TRLs of potential technologies may be lab-scale demonstration, where the process inventory data required for LCA was not sufficiently obtained from the experimental demonstration considering their upscaling. For such technology, modeling and simulation can be employed to fill the gap of foreground data (Tsoy et al., 2020), which can take into account the future potential of the technology and aim to predict the environmental impacts on the technology under development (Arvidsson et al., 2018; Moni et al., 2020; Thonemann et al., 2020).

4. Conclusions

Prospective assessments for novel technology options are employed in the filling of data gaps and interpretation of assessment results with uncertainties. PSE can be applied into such assessments and have an important role of design of systems. The basics of PSE, i.e., mathematical modelling and simulation of changes to understand the impact on mass and heat balances, are essential for appropriate technology and system assessments. The obtained information applying PSE can become the essential information for the social changes which involve various stakeholders. The elaborated interpretation for those who are not experts in PSE is needed to accurately convey the quantitative and qualitative essences clarified by PSE.

Acknowledgement

This work was supported by MEXT/JSPS KAKENHI Grant Number JP21H03660, JST-Mirai Program Grant Number JPMJMI19C7, and JST COI-NEXT JPMJPF2003. Activities of the Presidential Endowed Chair for “Platinum Society” at the University of Tokyo are supported by the KAITEKI Institute Incorporated, Mitsui Fudosan Corporation, Shin-Etsu Chemical Co., ORIX Corporation, Sekisui House, Ltd., the East Japan Railway Company, and Toyota Tsusho Corporation.

References

R. Arvidsson, A.M. Tillman, B.A. Sandén, M. Janssen, A. Nordelöf, D. Kushnir, S. Molander. 2018, Environmental assessment of emerging technologies: recommendations for prospective LCA. J Ind Ecol, 22(6), 1286-1294.

S. Debois, T. Hildebrandt, M. Marquard, T. Slaats, 2015. Bridging the valley of death: a success story on Danish funding schemes paving a path from technology readiness level 1 to 9. Software Engineering Research and Industrial Practice (SER&IP), 2015 IEEE/ACM 2nd International Workshop 54–57.

S.J. Eichhorn, A. Dufresne, M. Aranguren, N.E. Marcovich, J.R. Capadona, S.J. Rowan, C. Weder, W. Thielemans, M. Roman, S. Renneckar, W. Gindl, S. Veigel, J. Keckes, H. Yano, K. Abe, M. Nogi, A.N.. Nakagaito, A. Mangalam, J. Simonsen, A.S. Benight, A. Bismarck, L.A.. Berglund, T. Peijs. 2010. Review: Current International Research into Cellulose Nanofibres and Nanocomposites. J. Mater. Sci., 45, 1−33.

F.W. Geels, J. Schot, 2007. Typology of sociotechnical transition pathways. Res. Policy, 36, 399–417.

F.W. Geels, B.K. Sovacool, S. Sorrell, 2017. Sociotechnical transitions for deep decarbonization. Science, 357, 1242–1244.

Y. Kanematsu, Y. Kikuchi, H. Yano. 2021. Life Cycle Greenhouse Gas Emissions of Acetylated Cellulose Nanofiber-reinforced Polylactic Acid Based on Scale-up from Lab-scale Experiments, ACS Sustainable Chem. Eng., 9(31), 10444-10452

A. Kaya, K. Okuwada, 2015. Investigating the courses of implementation by describing resarch perfomance, Shakai Gijutsu Ronbunshu, 12, 12-22.

Y. Kikuchi, M. Nakai, Y. Kanematsu, K. Oosawa, T. Okubo, Y. Oshita, Y. Fukushima, 2020a. Application of technology assessments into co-learning for regional transformation: A case study of biomass energy systems in Tanegashima, Sustain. Sci., 15, 1473-1494.

Y. Kikuchi, A. Heiho, Y. Dou, I. Suwa, I.C. Chen, Y. Fukushima, C. Tokoto, 2020b. Defining Requirements on Technology Systems Assessment from Life Cycle Perspectives: Cases on Recycling of Photovoltaic and Secondary Battery, Int. J. Autom. Technol., 14(6), 890-908.

Y. Kikuchi, I. Suwa, A. Heiho, Y. Dou, S. Lim, T. Namihira, K. Mochidzuki, T. Koita, C. Tokoro, 2021, Separation of cathode particles and aluminum current foil in lithium-ion battery by high-voltage pulsed discharge Part II: Prospective life cycle assessment based on experimental data, Waste Manage., 132, 86-95.

Ministry of Environment Japan, 2014. User manual of TRL setting tool, https://www.env.go.jp/earth/ondanka/biz_local/26_01/trl_manual.pdf

S.M. Moni, R. Mahmud, K. High, M. Carbajales-Dale. 2020. Life cycle assessment of emerging technologies: A review. J Ind Ecol, 24, 52‑63.

D. Rotolo, D. Hicks, B.R. Martin. 2015. What is an emerging technology? Res. Policy, 44(10), 1827-1843.

N. Thonemann, A. Schulte, D. Maga. 2020. How to Conduct Prospective Life Cycle Assessment for Emerging Technologies? A Systematic Review and Methodological Guidance. Sustainability, 12(3), 1192.

C. Tokoro, S. Lim, K. Teruya, M. Kondo, K. Mochizuki, T. Namihira, Y. Kikuchi, 2021, Separation of cathode particles and aluminum current foil in Lithium-ion battery by high-voltage pulsed discharge part I: Experimental investigation, Waste Management, 125, 58-66.

N. Tsoy, B. Steubing, C. van der Giesen, J. Guinée. 2020. Upscaling methods used in ex ante life cycle assessment of emerging technologies: a review. Int. J. Life Cycle Assess. 25, 1680-1692.

J.P. Weyant, 2011. Accelerating the development and diffusion of new energy technologies: beyond the “valley of death”. Energ. Econ. 33, 674-682.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50007-5

Q-MPC: Integration of Reinforcement Learning and Model Predictive Control for Safe Learning

Tae Hoon Oh[a], Jong Min Lee[a*]

[a]*School of Chemical and Biological Engineering, Institute of Chemical Processes, Seoul National University, Seoul, 08826, Republic of Korea*
jongmin@snu.ac.kr

Abstract

Model-free reinforcement learning (RL) learns an optimal control policy by using the process data only. However, simple application of model-free RL to a practical process has a high risk of failure because the available amount of data and the number of trial runs are limited. Moreover, it is likely that state constraints are violated during the learning period. In this work, we propose Q-MPC framework, an integrated algorithm of RL and model predictive control (MPC) for safe learning. The Q-MPC learns the action-value function in an off-policy fashion and solves a model-based optimal control problem where the trained action-value function is assigned as the terminal cost. Because the Q-MPC utilizes a model, the state constraints can be respected during the learning period. For simulation study, Q-MPC, MPC, and double deep Q-network (DDQN) were applied with varying prediction horizons. The results show the advantages of Q-MPC that outperforms MPC by reducing the model-plant mismatch and shows much fewer constraint violations than DDQN.

Keywords: Reinforcement Learning; Model Predictive Control; Optimal Control; Safe Learning

1. Introduction

As the digitalization of manufacturing processes progresses, an unprecedented amount of operational data are measured and stored. Accordingly, there is a growing interest in developing data-based methods that can improve the existing process performance. Model-free reinforcement learning (RL) is a data-based optimal control method that aims to learn an optimal control policy in the absence of a process model. Model-free RL can be applied to any discrete-time system as long as the system has the Markov property. Therefore, optimal control policies for a wide range of complex systems characterized by nonlinearity, discrete events, and stochasticity can be obtained if one can secure a sufficient amount of data. In addition, the online computation of trained control policy is much less than that of the model-based control such as model predictive control (MPC). In line with these advantages, several studies conduct apply model-free RL methods to chemical processes, such as simulated moving bed (Oh et al. 2021), microfluid (Dressler et al. 2018), polymerization (Ma et al. 2019), polishing, and photo-product bioprocesses (Petsagkourakis et al. 2020).

However, using model-free RL to obtain an optimal control policy of the manufacturing process may pose several practical challenges. First, the amount of data required to learn an optimal control policy may not be practical to obtain even for a digitalized process. Also, the data should be generated with a certain degree of explorations that may do harm

to the process performance. Finally, the model-free RL does not have any model for the state transition, which cannot guarantee the satisfaction of state constraints. More specifically, model-free RL cannot explicitly ensure the state constraints, but it can consider the constraints implicitly by modifying either the optimality criterion such as adding penalty terms or the exploration procedure with the guidance of a risk metric (Garcia and Fernandez 2015). Therefore, the model-free RL can learn the state constraint only after it experiences the constraint violation. This is a major drawback as the state constraints are typically imposed for safety reasons.

In most cases, a model built on *a priori* knowledge of system dynamics is available. This model may not be precise, but it can provide the information of correlation between the state and input and can be used to ensure the safety constraints with a sufficient margin. Therefore, using data to improve the existing model-based control can be a more practical approach instead of completely ruling out the model like model-free RL. We propose an algorithm that integrates RL and MPC, referred to as Q-MPC. Q-MPC improves the performance of existing MPC by incorporating the advantage of data-based learning of RL. We first formulate a double deep Q-network (DDQN) optimization problem on the continuous action space, which uses gradient-based numerical optimization. This method is an off-policy algorithm where only the critic is approximated by a deep neural network. Then, the actor, originally represented as the optimizer of the trained action-value function, is extended to an open-loop model-based optimal control problem. This model-based optimal control problem predicts the states and costs up to the prediction horizon with a model and assigns the action-value function as a terminal cost. Therefore, the Q-MPC is a generalization method of MPC as the actor implements the control input by solving the optimization problem in a receding horizon fashion. Furthermore, Q-MPC becomes equivalent to DDQN in continuous action space by setting the prediction horizon length to 0. The Q-MPC can explicitly impose the state constraints and explicitly schedule the exploration. In addition, it can improve the control policy with a much less amount of data than the model-free RL methods. For the simulation study, MPC, Q-MPC, and DDQN are applied, where the length of the prediction horizon is scheduled. The simulation results show that Q-MPC outperforms MPC by learning and can guarantee the satisfaction of state constraints even during the learning period. DDQN also outperforms MPC after sufficient learning, but it violates the state constraints much more frequently than Q-MPC during the learning period.

2. Q-model predictive control

Suppose that the dynamic model and constraints are represented as $\dot{x} = f(x,u)$ and $g(x,u) \leq 0$. In addition, let the path-wise and terminal cost of the system be $L(x,u)$ and $\phi(x,u)$, respectively. Even though the system dynamics is given as continuous time, the control input is assumed to be implemented on the system in a discrete-time manner with zero-order hold. The time interval between control inputs is denoted as Δt and is fixed as a constant value. Let the system be terminated at a finite time step N_T. Then, the total cost is given by

$$J(x_{0:N_T}, u_{0:N_T}) = \phi(x_{N_T}, u_{N_T}) + \sum_{k=0}^{N_T-1} L(x_k, u_k). \tag{1}$$

Suppose that the system is controlled by a control policy $u(\cdot)$. The value function, $V^l_{u(\cdot)}(x)$ associated with this control policy $u(\cdot)$ is defined as

$$V^l_{u(\cdot)}(x_0) = E\left[\phi\big(x_{N_T}, u(x_{N_T})\big) + \sum_{k=l}^{N_T-1} L(x_k, u(x_k)) \mid x_l = x_0\right], \tag{2}$$

where the integer $l \in [0, \dots, N_T]$ denotes the time step. Similarly, the action-value function (Q-function) is defined as

$$Q^l_{u(\cdot)}(x_0, u_0) = E\left[\phi\big(x_{N_T}, u(x_{N_T})\big) + \sum_{k=l}^{N_T-1} L(x_k, u(x_k)) \mid x_l = x_0, u_l = u_0\right]. \tag{3}$$

The optimal control policy $u^*(\cdot)$ is defined as a control policy that gives the minimal return, $\mathrm{V}^{\mathrm{l}}_{u^*(\cdot)}(\mathrm{x}) \leq \mathrm{V}^{\mathrm{l}}_{u(\cdot)}(\mathrm{x})$, for all feasible x and l. Suppose that an optimal control policy exists, then the value function with $l = 0$ is equal to the optimal value. In addition, once the optimal action-value function is given, then the optimal control policy can be obtained without the information of system dynamics by

$$u^*(x) \in \underset{u}{\operatorname{argmin}}\, Q^l_{u^*(\cdot)}(x, u). \tag{4}$$

Therefore, learning the action-value function implies learning an optimal control policy, and the Q-learning based RL aims to approximate the optimal action-value function without any knowledge of state dynamics.

The Q-learning based RL is classified as an off-policy algorithm, that is the action-value function can be learned from any data obtained from the system. Therefore, all the data obtained from the system can be stored in a single data set, and learning can proceed with the data randomly sampled from this set (Van Hasselt et al, 2016). This random sampling helps to break the correlations in the measured sequence and smooth over changes in the data distribution. The Bellman equation states that the optimal action-value function satisfies the following recursive equation (Sutton and Barto 2018)

$$Q^l_{u^*(\cdot)}(x_l, u_l) = E\left[L(x_l, u_l) + \min_u Q^{l+1}_{u^*(\cdot)}(x_{l+1}, u)\right]. \tag{5}$$

The Bellman equation (5) breaks the sequence of costs into a single time step by using the principle of optimality. Then, the input choice made from the behaviour policy for the next time step is replaced by the target policy which enables to update the actor in an off-policy fashion.

To prevent the selection of the under-estimated value (over-estimated for maximization) in (5), two function approximators can be used to approximate the action-value function. This algorithm is called Double Deep Q-Network (DDQN). The target deep neural networks is denoted as $Q_{\theta_t}(x, u)$ and the online deep neural networks is denoted as $Q_{\theta_{on}}(x, u)$, respectively. The target network is only utilized to evaluate the minimal value of action-value function in (5) to learn the online network. In this case, the squared error for a single list tuple $D = [x_l, u_l, L(x_l, u_l), x_{l+1}]$ is given as

$$Error(\theta_{on}, \theta_t, D) = E\left[Q^l_{\theta_{on}}(x_l, u_l) - L(x_l, u_l) - \min_u Q^{l+1}_{\theta_t}(x_{l+1}, u)\right]. \tag{6}$$

Let $B = \{D_1, D_2, \dots, D_n\}$ be the batch data set which is the set of several data lists randomly sampled from the data set. Then, the online network is updated by applying the one-step gradient descent with the appropriate learning rate α as

$$\theta_{on} \leftarrow \theta_{on} + \alpha \nabla_{\theta_{on}} \frac{1}{|B|} \sum\nolimits_{D \in B} Error(\theta_{on}, \theta_t, D). \tag{7}$$

The target network can be updated by

$$\theta_t \leftarrow (1 - \tau)\theta_t + \tau\theta_{on}. \tag{8}$$

where $\tau \in [0, 1]$ is the update rate. Then, DDQN selects the control input u_l associated with the state x_l by solving the following simple optimization problem:

$$u_l = \underset{u}{\operatorname{argmin}}\, Q^l_{\theta_{on}}(x_l, u). \tag{9}$$

The input of DDQN is totally determined by the learned action-value function Q_{on}. This implies that the control performance of DDQN can be degraded, and the constraints can be violated with inaccurate action-value function. This is a common but crucial problem for all model-free RL that the constraints can be learned only after they are violated.

Instead of using (4) to calculate control input, the Q-MPC solves the following open-loop optimal control problem:

$$\min_u Q_{\theta_{on}}\left(x_{l+N_p}, u_{l+N_p}\right) + \sum_{k=l}^{l+N_p-1} L(x_k, u_k)$$

$$\text{subject to} \tag{10}$$

$$x_l \text{ is given, } x_{k+1} = f(x_k, u_k) \text{ and } g(x_k, u_k) \le 0$$

where N_p denotes the prediction horizon and the continuous system is converted into its discrete-time counterpart. If the prediction reaches the terminal time, then the exact terminal cost ϕ is assigned to (10) instead of action-value function. Note that, solving the optimization problem (10) gives the open-loop control trajectory $u_l, \dots, u_{l+N_p}$ but only u_l is implemented to the system.

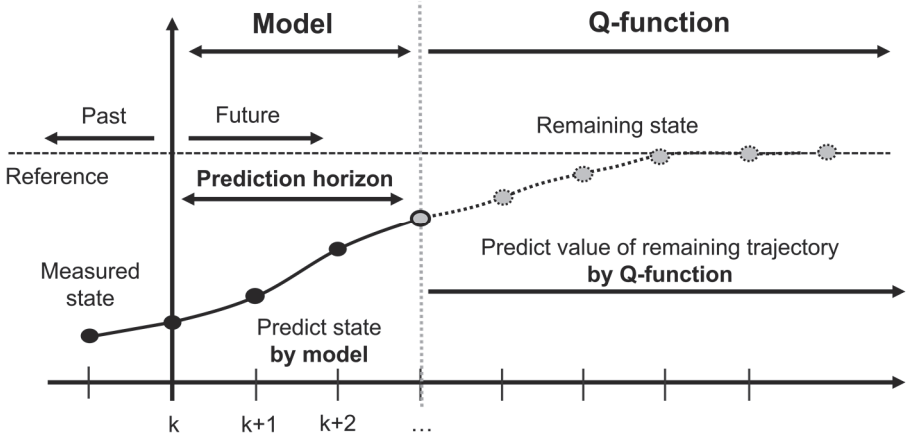


Figure 1: The scheme of Q-MPC.

The scheme of Q-MPC is presented in Figure 1. Compared with nominal MPC, the terminal cost of Q-MPC is replaced by the trained action-value function. Because the action-value function Q learns the value from the data, Q-MPC can adapt the system dynamics. Compared with DDQN, the actor of Q-MPC is also presented as an optimization problem formulated as a mathematical program such as the quadratic program (10). However, the model is used to predict the state transition and associated cost for N_p time step. The use of the model helps to satisfy the constraints and dramatically reduces the required amount of data to improve the control policy. The receding horizon control makes it easy to satisfy the constraints even with a short prediction horizon.

The prediction horizon of Q-MPC determines how long model is involved in prediction. For example, if the length of prediction horizon is 0, then the Q-MPC is equivalent to DDQN in the continuous domain. In this case, the model is completely excluded in determining control policy. On the other hand, if the prediction reaches the terminal time, the action-value function is excluded in calculating control inputs, and Q-MPC becomes equivalent to nominal MPC. Therefore, the performance of Q-MPC is directly affected by the length of prediction horizon which is another tuning parameter. We suggest setting the prediction horizon much smaller than the whole batch operation because even for the short prediction horizon length, the input is highly affected by the model. In addition, we suggest setting relatively long prediction horizon at the early stage of learning where the accuracy of action-value function is low.

3. Simulation studies

A simple photo-production system having 3 states ($x_1, x_2, and\ x_3$) and 2 inputs ($u_1\ and\ u_2$) is considered where the system dynamics are given as (Petsagkourakis et al. 2020)

$$\frac{dx_1}{dt} = 1, \tag{11}$$

$$\frac{dx_2}{dt} = -(u_1 + 0.5u_1^2)x_2 + \frac{0.5u_2}{(x_2 + x_3 + 0.1)}, \tag{12}$$

$$\frac{dx_3}{dt} = u_1 x_2 - 0.2u_2 x_2, \tag{13}$$

where the first state denotes the time. The time interval is selected as 0.05 and system is assumed to be terminated at 1. Therefore, the system contains 20 horizons. Note that because the system terminates in finite steps, the time should be included in the state.

The lower and upper bounds for all states are 0 and 1, respectively, and the bounds for inputs are 0 and 5, respectively. In addition, the second state has additional lower bound presented as

$$x_2 \geq 0.45. \tag{14}$$

The path-wise cost and terminal cost is given as

$$L(x,u) = 0.01\left(\frac{u_1}{25}\right)^2, \text{and} \tag{15}$$

$$\phi(x,u) = 3(1 - x_3), \tag{16}$$

and the penalty $\max(0,\ 0.45 - x_2)$ for violating constraint (*) is added to the cost.

The surrogate model that Q-MPC used is given as

$$\frac{dx_1}{dt} = 1, \tag{17}$$

$$\frac{dx_2}{dt} = -(u_1 + 0.55u_1^2)x_2 + \frac{0.5u_2}{(x_2 + x_3 + 0.5)}, \tag{18}$$

$$\frac{dx_3}{dt} = u_1 x_2 - 0.1u_2 x_2. \tag{19}$$

The Q-MPC solves the following optimization problem

$$\min_u V\left(x_{N_p}\right) + \sum_{k=0}^{N_p-1} L(x_k, u_k)$$

subject to

(4), (5), (6), and $x_2 \geq 0.45$ for k = 0, ..., N_p. (20)

The action-value function is approximated by deep neural networks that is consisted of three layers. The number of nodes for each layer are 16, 4, and 1, respectively. The

following smooth activation function is utilized to optimize the deep neural network by IPOPT method

$$A(x) = \log(1 + x^2). \tag{21}$$

The online network was trained for every episode, and the target network was updated for every 5 episodes. The learning and updating rates were selected as 0.02 and 0.01, respectively. The batch size |B| was selected as 64. Each method was applied to the system for 1,000 episodes, but the first 10 episodes were simulated with random control inputs for comparison. The prediction horizon of Q-MPC was set to 20 for episodes 10 to 30, 5 for episodes 30 to 200, and 1 for episodes 500 to 1,000.

The simulation results are presented in Figures 2 and 3. Figure 2 shows the optimal total cost and total cost obtained by Q-MPC, MPC, and DDQN. Because the model in MPC does not change throughout the simulations, the total cost of MPC is kept constant. The Q-MPC cannot outperform MPC with a prediction horizon of 5, as the mismatch between the model and action-value function can worsen the performance. In addition, the total cost increases around 200 episodes because the prediction horizon of Q-MPC is changed from 5 to 3. Then, the total cost gradually decreases to that of MPC by learning. The first episode that Q-MPC outperforms MPC is the 297^{th} episode. DDQN also successfully learns the system dynamics and eventually outperforms MPC, but the first episode that outperforms MPC is the 718^{th} episode. Figure 2 clearly shows that Q-MPC improves the control policy much faster than DDQN. Figure 3 shows the number of constraint violations. Because the number of horizons for a single episode is 20, the maximum possible number of violations is 20. Q-MPC never violates the constraint throughout learning, whereas DDQN violates the constraint even if the learning is nearly finished. Note that the violations in the early stage are made by randomly implemented input and not by Q-MPC. Figure 3 shows the advantages of Q-MPC that it can safely learn the system dynamics and improve the existing control policy.

4. Conclusions

We proposed a novel Q-MPC algorithm to learn the system safely. Q-MPC is a generalization method of both DDQN and MPC, where each method can be recovered by adjusting the length of the prediction horizon. The simulation results showed that Q-MPC could improve the control policy by satisfying the state constraint. In addition, Q-MPC requires much less amount of data to improve the control policy than DDQN.

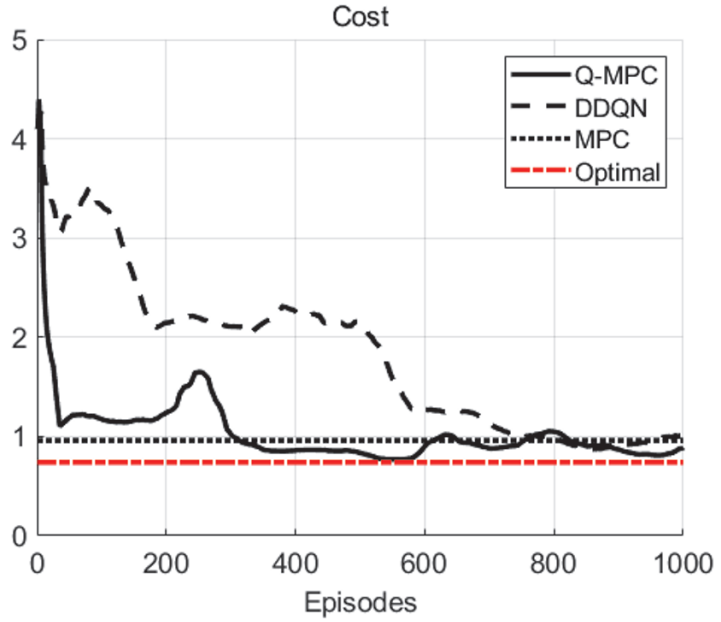


Figure 2: The moving averaged value of total cost with 50 samples.

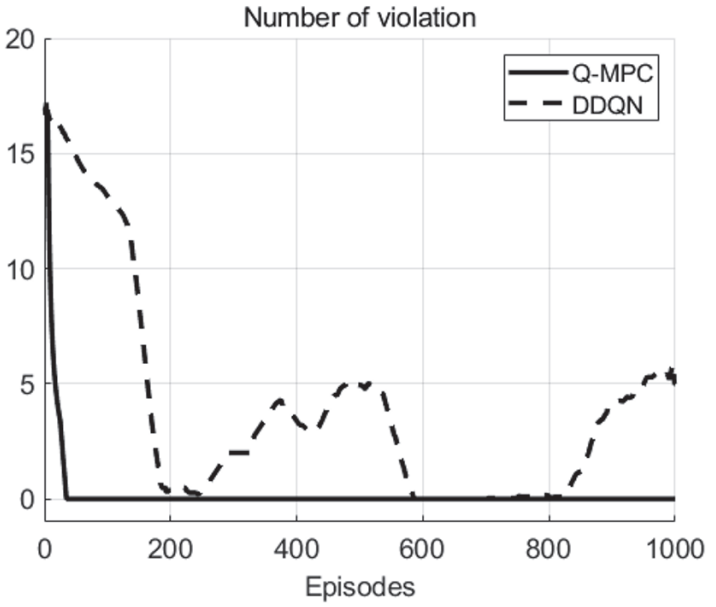


Figure 3: The number of constraint violations.

References

T. H. Oh, J. W. Kim, S. H. Son, H. Kim, K. Lee, and J. M. Lee, 2021, Automatic control of simulated moving bed process with deep Q-network, Journal of Chromatorgraphy A, 1647, 462073.

O. J. Dressler, P. D. Howes, J. Choo, and A. J. deMello, 2018, Reinforcement learning for dynamic microfulidic control, ACS omega, 3, 8, 10084 – 10091.

Y. Ma, W. Zhu, M. G. Michael, and J. Romagnoli, 2019, Continuous control of a polymerization system with deep reinforcement learning, Journal of Process Control, 75, 40 – 47.

P. Petsagkourakis, I. O. Sandoval, E. Bradford, D. Zhang, and E. A. del Rio-Chanona, 2020, Reinforcement learning for batch bioprocess optimization, Computers & Chemical Engineering, 133, 106649.

J. Garcia, and F. Fernandez, 2015, A comprehensive survey on safe reinforcement learning, Journal of Machine Learning Research, 16, 1, 1437 – 1480.

H. Van Hasselt, A. Guez, and D. Silver, 2016, Deep reinforcement learning with double q-learning, Proceedings of the AAAI conference on artificial intelligence, 30, 1.

R. S. Sutton, and A. G. Barto, 2018, Reinforcement learning: An introduction, MIT press.

This page intentionally left blank

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50008-7

Presentation abstract: Optimization formulations for machine learning surrogates

Francesco Ceccon[a*], Jordan Jalving[b*], Joshua Haddad[b], Alexander Thebelt[a], Calvin Tsay[a], Carl D Laird[c+], Ruth Misener[a+]

[a]*Department of Computing, Imperial College London, 180 Queen's Gate, SW7 2AZ, UK*
[b]*Center for Computing Research, Sandia National Laboratories, Albuquerque, NM 87123, USA*
[c]*Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA*
[*]*Contributed equally,* [+]*Contributed equally*
claird@andrew.cmu.edu, r.misener@imperial.ac.uk

Abstract

In many process systems engineering applications, we seek to integrate surrogate models, e.g. already-trained neural network and gradient-boosted tree models, into larger decision-making problems. This presentation explores different ways to automatically take the machine learning surrogate model and produce an optimization formulation. Our goal is to automate the entire workflow of decision-making with surrogate models from input data to optimization formulation. This presentation discusses our progress towards this goal, gives examples of previous successes, and elicits a conversation with colleagues about the path forward.

Keywords: neural networks, gradient-boosted trees, Pyomo, optimization formulations

1. Main Text

The optimization and machine learning toolkit (https://github.com/cog-imperial/OMLT, OMLT 1.0) is an open source software package enabling optimization over high-level representations of neural networks (NNs) and gradient-boosted trees (GBTs). Optimizing over trained surrogate models allows integration of NNs or GBTs into larger decision-making problems. Computer science applications include maximizing a neural acquisition function (Volpp et al., 2019) or verifying neural networks (Botoeva et al., 2020). Engineering applications of grey-box optimization (Boukouvala et al., 2016) hybridize mechanistic, model-based optimization with surrogate models learned from data. OMLT 1.0 supports GBTs through an ONNX (https://github.com/onnx/onnx) interface and NNs through both ONNX and Keras interfaces. OMLT transforms these pre-trained machine learning models into the algebraic modeling language Pyomo (Bynum et al., 2021) to encode optimization formulations.

OMLT is a general tool incorporating both NNs and GBTs, many input models via ONNX interoperability, both fully-dense and convolutional layers, several activation functions, and various optimization formulations. The literature often presents these different optimization formulations as competitors, e.g. our partition-based formulation competes with the big-M formulation for ReLU NNs (Kronqvist et al., 2021; Tsay et al., 2021). In

OMLT, competing optimization formulations become alternatives: users can switch between the formulations and find the best for a specific application.

References

Elena Botoeva, Panagiotis Kouvaros, Jan Kronqvist, Alessio Lomuscio, and Ruth Misener. Efficient verification of ReLU-based neural networks via dependency analysis. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, pages 3291–3299, 2020.

Fani Boukouvala, Ruth Misener, and Christodoulos A. Floudas. Global optimization advances in Mixed-Integer Nonlinear Programming, MINLP, and Constrained Derivative- Free Optimization, CDFO. European Journal of Operational Research, 252(3):701 – 727, 2016.

Michael L Bynum, Gabriel A Hackebeil, William E Hart, Carl D Laird, Bethany L Nicholson, John D Siirola, Jean-Paul Watson, and David L Woodruff. Pyomo—Optimization Modeling in Python, volume 67. Springer Nature, 2021.

Jan Kronqvist, Ruth Misener, and Calvin Tsay. Between steps: Intermediate relaxations between big-M and convex hull formulations. In International Conference on Integration of Constraint Programming, Artificial Intelligence, and Operations Research, pages 299– 314. Springer, 2021.

Artur M Schweidtmann and Alexander Mitsos. Deterministic global optimization with artificial neural networks embedded. Journal of Optimization Theory and Applications, 180(3):925–948, 2019.

Calvin Tsay, Jan Kronqvist, Alexander Thebelt, and Ruth Misener. Partition-based formulations for mixed-integer optimization of trained ReLU neural networks. In Advances in Neural Information Processing Systems, 2021.

Michael Volpp, Lukas P Fröhlich, Kirsten Fischer, Andreas Doerr, Stefan Falkner, Frank Hutter, and Christian Daniel. Meta-learning acquisition functions for transfer learning in Bayesian optimization. In International Conference on Learning Representations, 2019.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50009-9

Pharma PSE: a multiscale approach for reimagining pharmaceutical manufacturing

Hirokazu Sugiyama

Department of Chemical system Engineering, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
sugiyama@chemsys.t.u-tokyo.ac.jp

Abstract

Confronted with the global challenges including COVID-19, pharmaceutical manufacturing needs to simultaneously achieve long-term efficiency and short-term resilience. Process systems engineering (PSE) can provide scientific basis here, and in fact, PSE researchers have made significant contributions to pharma in the last decade. The author, after having worked for a global pharmaceutical company, initiated research on pharmaceutical process systems engineering: Pharma PSE. The research tackles different challenges in small molecules, biopharmaceuticals, and regenerative medicine, at the scales of molecules/cells, processes, and the society. This paper first introduces the viewpoint of Pharma PSE, followed by showcasing a research example that involved a range of computer-aided analyses at different scales. The multiscale approach of Pharma PSE can provide a new horizon to "reimagine" pharmaceutical manufacturing processes and beyond, towards establishment of a sustainable healthcare society.

Keywords: Pharmaceuticals, Regenerative medicine, Process modelling, Process design, Sustainable healthcare society.

1. Introduction

The relevance of pharmaceuticals is more apparent than ever before. The Sustainable Development Goals (SDGs; United Nations Development Programme, 2021) defined the achievement of the "access to safe, effective, quality and affordable essential medicines and vaccines for all" as a part of Goal No. 3. Long-term efficiency is critical for manufacturing while the development pipeline of new drugs needs to be enhanced further. Another mandate for manufacturing is to cope with pandemics, especially COVID-19, by dealing with the short-term surges in demand and disruptions in the supply chain. The pharmaceutical industry needs to establish a system where long-term efficiency and short-term resiliency are achieved at the same time.

In the last decade, the community of process systems engineering (PSE) have been introducing and practicing systems approaches in the design, operation, and control of pharmaceutical production processes. The previous studies have covered various topics in the manufacture of active pharmaceutical ingredients (APIs) as well as dosage forms (e.g., tablets and injectables). Continuous manufacturing and process analytical technologies (PATs) have been intensively researched (e.g., Badr and Sugiyama, 2020; Bhalode et al., 2021; Diab et al., 2021, Ghijs et al., 2021; Hong et al., 2021). Furthermore, advanced model-based approaches for quality assurance (e.g., Ochoa et al., 2021) and the subjects related to personalized healthcare (e.g., Içten et al., 2015; Wang et al., 2018; Papathanasiou et al., 2020) have been investigated.

Table 1. Characteristics and research opportunities in the pharmaceutical domain

Small molecules	Biopharmaceuticals	Regenerative medicine
• Large quantity in supply • Most conventional form of medicine (e.g., tables) • Intensive research performed on "flow chemistry" & "continuous manufacturing" • Rigorous modeling needed for the G/L/S interfaces in flow chemistry • Difficulty in modeling heterogeneity in powder processing	• Market rapidly growing • Monoclonal antibodies (mAbs) and vaccines attracting recent attentions • Intensive research performed on host cells, media, manufacturing technologies and equipment, and measurement devices • Challenges in modeling biological behavior (e.g., heterogeneity, dynamics, and impurities)	• Future therapy based on stem cells • Clinical trials intensively performed • Need to establish manufacturing processes as well as supply chain • High cost of R&D and manufacturing raising social attentions • Models required for describing cell behavior, process performance, supply chain, and cost-effectiveness

In a world of ever-increasing demand for advanced pharmaceuticals, there is a need for increasing efficiency, flexibility, and production capacity. At the same time, there are innovations that span all categories of pharmaceutical products. Table 1 shows the characteristics of different product categories, and the associated modeling challenges. Innovations here include the introduction of novel therapeutics, materials, and shifts in production scales especially for personalized medicine. Such innovations create a need and an opportunity to "reimagine" pharmaceutical manufacturing to better accommodate the changes and developments in the industry and in society.

The author, after having worked for a global pharmaceutical company, launched a research group on pharmaceutical process systems engineering: Pharma PSE. The research tackles challenges in small molecules, biopharmaceuticals, and regenerative medicine from a multiscale viewpoint. The research aims to expand PSE into a critical domain in society by incorporating the systems approach into the development of new products and processes. This paper first introduces the multiscale viewpoint of Pharma PSE, followed by a case study on regenerative medicine. This paper serves as the basis for the keynote lecture at PSE2021+ with more materials to be added.

2. Multiscale research viewpoint

Figure 1 describes the multiscale viewpoint of Pharma PSE. At the molecule/cell level, elements of a manufacturing process are investigated such as the choice of host cells, nutrition media, or protective agents are considered. At the process level, alternatives regarding manufacturing technologies, equipment specification, and operation strategy in manufacturing processes are investigated. Higher level assessments are conducted up to the level of the healthcare society.

Figure 1 The multiscale viewpoint in Pharma PSE research.

Recognizing the entire system as in Figure 1 would facilitate bottom-up analyses that allow for the comprehensive assessment of impacts of lower-level modifications on higher-level targets. For example, the performance of novel host cells (cell level) could be assessed regarding lead time (process level), and supply agility (society level). Top-down analyses could also be enabled for determining promising alternatives at lower levels given higher levels goals. In conducting Pharma PSE research, such "zoom-in and zoom-out" is supported by the appropriate consideration/use of modeling strategy (first-principle, data-driven, or hybrid), simulation methods, objective function(s), design and operational alternatives, and physical- and cyber-space information.

3. Research example on cryopreservation of hiPS cells

This paper introduces design of cryopreservation processes for hiPS cells as an example.

3.1. Molecule/cell level: computational screening of cryoprotective agents

As a study at the molecule/cell level, this work (Hayashi et al., 2021a) investigated cryoprotective agents (CPAs) that are used in the cryopreservation of cells including hiPS cells. A computational screening was performed for candidate compounds using quantum chemistry and molecular dynamics (MD) simulations. The motivation was to search for an alternative CPA to dimethyl sulfoxide (DMSO), which is currently widely used but is known to be toxic to cells. Figure 2 shows the overview of the work. For forty compounds, the solvation free energy and partition coefficient, and the root mean square deviation (RMSD) of a phospholipid bilayer which composes a cell membrane, were calculated by quantum chemistry simulation and by MD simulation, respectively. These three indicators were used to assess osmoregulatory ability, affinity with a cell membrane, and ability to stretch a cell membrane, respectively. The quantum chemistry simulation revealed that trimethylglycine, formamide, urea, thiourea, diethylene glycol, and dulcitol were better than DMSO, regarding either or both of the indicators considered. Further analysis with the MD simulation suggested formamide, thiourea, and urea as promising candidates within the simulated conditions.

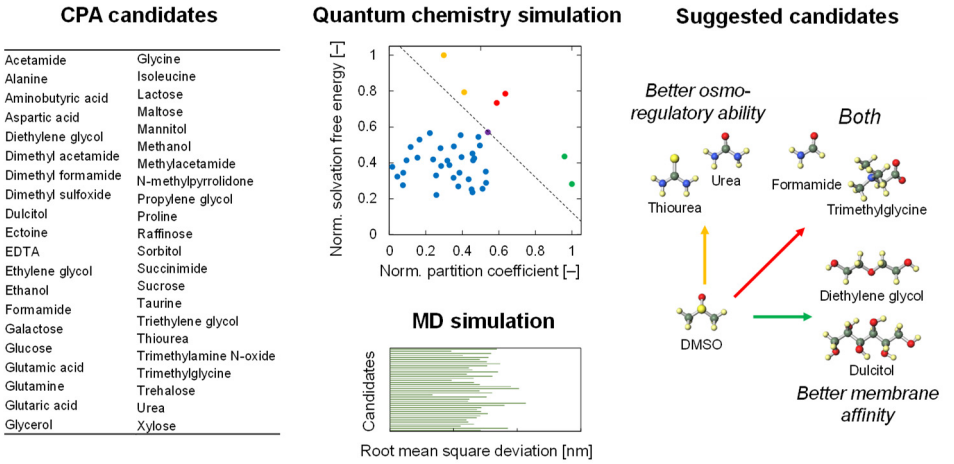


Figure 2 Computational screening of CPA candidates using quantum chemistry and MD simulations (Hayashi et al., 2021a)

3.2. Process level: model-based assessment of temperature profiles in slow freezing

As a study at the process level, this work (Hayashi et al., 2021b) presented a model-based assessment of temperature profiles in slow freezing for hiPS cells. Figure 3 shows the summary. The basis here was our previously developed single-cell model (Hayashi et al., 2020) that consists of heat transfer, mass transfer, and crystallization models. The three models can quantify temperature distribution in a vial, cell volume change through transmembrane water transport, and intracellular ice formation during freezing, respectively. These first-principle models was then extended to cover the cell survival rate through data-driven modeling. Experiments using hiPS cells provided the necessary parameter values of the multivariate regression model. The newly developed hybrid single-cell model can, given a temperature profile of freezing, estimate the cell survival rate and required freezing time as the quality and productivity indicators, respectively. As a case study, the model was used to assess ca. 16,000 temperature profiles. The simulation results suggested that fast-slow-fast (i.e., non-linear) cooling in the dehydration, nucleation-promoting, and further cooling zones, respectively, as a promising profile.

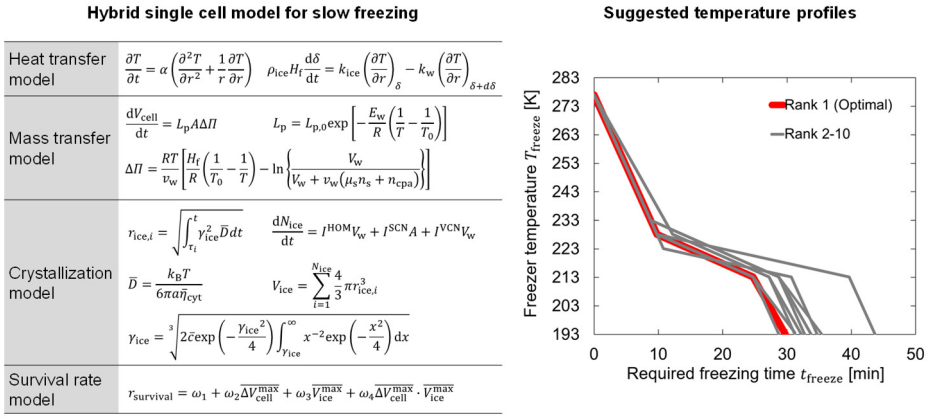


Figure 3 Model-based assessment of freezing temperature profiles (Hayashi et al., 2021b)

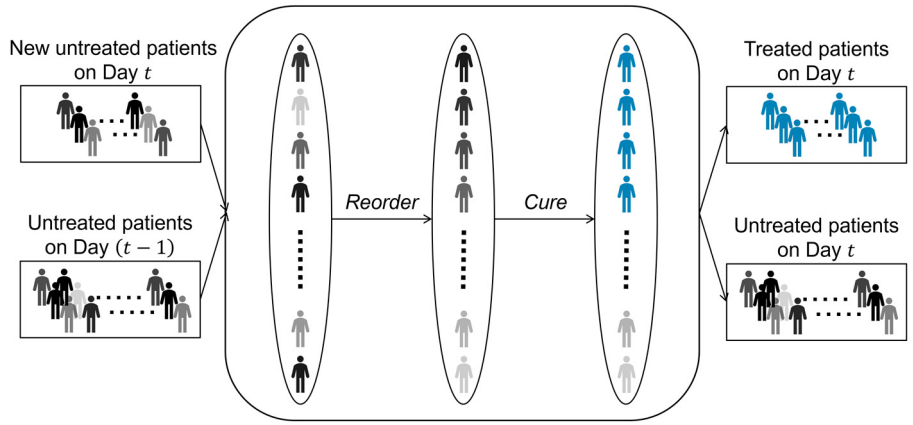


Figure 4 Agent-based model for analyzing cost-effectiveness in the manufacture of allogeneic hiPS cells in Japan (Hayashi et al., 2022)

3.3. Society level: cost-effectiveness analysis in the manufacture of allogeneic hiPS cells in Japan

Finally as a study at the society level, this work (Hayashi et al., 2022) proposed an agent-based model (ABM) for cost-effectiveness analyses in the manufacture of allogeneic hiPS cells in Japan. The ABM (see Figure 4) was set up for estimating the disability-adjusted life years (DALYs) of each patient and the total required cost for manufacturing allogeneic hiPS cells. The DALYs and the total required cost for manufacturing were used as the indicators of effectiveness and cost, respectively. Cryopreservation was considered as a part of the manufacturing processes. The developed ABM can calculate these two indicators, given the disease, the annual number of treated patients, and the treatment mode. The developed model was then applied to analyze therapy for two diseases using allogeneic hiPS cells, which are currently undergoing clinical studies in Japan. A case study suggested that the treatment mode (i.e., treating patients from the youngest to the oldest) would affect the cost-effectiveness significantly.

4. Lessons learnt towards future research

The interconnections between each of the above studies across different scales are visible by taking the multiscale view shown in Figure 1. For example, the choice of CPA can influence the extracellular condition during freezing, which can affect the process duration, and thus the supply performance. Taking a multiscale approach requires the use of flexible modelling strategies, e.g., in this case the use of models spanning quantum chemistry and agent-based modelling with different degrees of detail. The approach also requires taking the position of multiple stakeholders, such as varying the research viewpoint form experimental investigator (for CPA) to process practitioner (for freezing unit), and further to manufacturers and the government (for cell supply). Similar characteristics could be observed in other workpieces in small molecule (e.g., Matsunami et al., 2020) and biopharmaceuticals (e.g., Badr et al., 2021). Recognition of these characteristics can provide new research opportunities such as multiscale scenario analyses and optimization, investigation of appropriate model resolution considering cross-scale interconnections, and multiobjective decision-making. The multiscale approach can provide a richer insight for “reimagining” the manufacturing processes and the associated systems of pharmaceuticals.

5. Conclusions

This paper presented the research viewpoint of Pharma PSE, which aims to cover topics in different drug categories at various scales in one frame. As an example, a study on cryopreservation of hiPS cells was showcased, which involved various simulation-based analyses. The observation indicated that the multiscale approach of Pharma PSE can provide a new horizon to “reimagine” the manufacturing processes and beyond, towards establishment of a sustainable healthcare society.

Acknowledgement

The author would like to thank Mr. Yusuke Hayashi, Dr. Sara Badr and Dr. Isuru Udugama at Sugiyama Lab in The University of Tokyo for the support in preparing the paper, and all lab members of for the hard work and innovation. All research partners and the funding supports (currently: Japan Society for the Promotion of Science (JSPS) [grant No. 21H01699 and 20K21102]; Japan Agency for Medical Research and Development (AMED) [grant No. JP21ae0121015AMED]) are sincerely appreciated.

References

S. Badr, H. Sugiyama, 2020. A PSE perspective for the efficient production of monoclonal antibodies: Integration of process, cell, and product design aspects. Curr. Opin. Chem. Eng. 27, 121–128.

S. Badr, K. Okamura, N. Takahashi, V. Ubbenjans, H. Shirahata, H. Sugiyama. 2021. Integrated design of biopharmaceutical manufacturing processes: Operation modes and process configurations for monoclonal antibody production. Comput. Chem. Eng. 153, 107422.

P. Bhalode, H. Tian, S. Gupta, S. M. Razavi, A. Roman-Ospino, S. Talebian, R. Singh, J.V. Scicolone, F.J. Muzzio, M. Ierapetritou. 2021. Using residence time distribution in pharmaceutical solid dose manufacturing–A critical review. Int. J. Pharm. 610, 121248.

S. Diab, M. Raiyat, D.I. Gerogiorgis, 2021. Flow synthesis kinetics for lomustine, an anti-cancer active pharmaceutical ingredient. React. Chem. Eng. 6, 1819–1828

M. Ghijs, B. Vanbillemont, N. Nicolaï, T. DeBeer, I. Nopens. 2021. Two-dimensional moisture content and size measurement of pharmaceutical granules after fluid bed drying using near-infrared chemical imaging. Int. J. Pharm. 595, 120069

Y. Hayashi, I. Horiguchi, M. Kino-oka, H. Sugiyama, 2020. Slow freezing process design for human induced pluripotent stem cells by modeling intracontainer variation. Comput. Chem. Eng. 132, 106597.

Y. Hayashi, Y. Nakajima, H. Sugiyama, 2021a. Computational screening of cryoprotective agents for regenerative medical products using quantum chemistry and molecular dynamics simulations. Cryobiol. 100, 101–109.

Y. Hayashi, I. Horiguchi, M. Kino-oka, H. Sugiyama, 2021b. Model-based assessment of temperature profiles in slow freezing for human induced pluripotent stem cells. Comput. Chem. Eng. 144, 107150.

Y. Hayashi, K. Oishi, H. Sugiyama, 2021. An agent-based model for cost-effectiveness analysis in the manufacture of allogeneic human induced pluripotent cells in Japan. Comput. Aided Chem. Eng. (Proceedings of PSE2021+), Paper ID 143, in submission.

M.S. Hong, A.E. Lu, R.W. Ou, J.M. Wolfrum, S.L. Springs, A.J. Sinskey, R.D. Braatz, 2021. Model-based control for column-based continuous viral inactivation of biopharmaceuticals. Biotechnol. Bioeng. 118, 3215–3224.

E. Içten, A. Giridhar, L.S. Taylor, Z.K. Nagy, G.V. Reklaitis. 2015. Dropwise additive manufacturing of pharmaceutical products for melt-based dosage forms. J. Pharm. Sci. 104, 1641–1649.

K. Matsunami, T. Nagato, K. Hasegawa, H. Sugiyama. 2020. Determining key parameters of continuous wet granulation for tablet quality and productivity: A case in ethenzamide. Int. J. Pharm. 579, 119160.

M.P. Ochoa, S. García-Muñoz, S. Stamatis, I.E. Grossmann. 2021. Novel flexibility index formulations for the selection of the operating range within a design space. Comput. Chem. Eng. 149, 107284.

M.M. Papathanasiou, Ch. Stamatis, M. Lakelin, S. Farid, N. Titchener-Hooker, N. Shah. 2020. Autologous CAR T-cell therapies supply chain: challenges and opportunities? Cancer Gene Ther. 27, 799–809.

United Nations Development Programme (UNDP), Sustainable Development Goals, https://www.undp.org/sustainable-development-goals#good-health (accessed January 04, 2022)

X. Wang, Q. Kong, M.M. Papathanasiou, N. Shah, 2018. Precision healthcare supply chain design through multi-objective stochastic programming. Comput. Aided Chem. Eng. 44, 2137–2142.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50010-5

Artificial Intelligence and Process Systems Engineering

Raghunthan Rengaswamy[a,b*]

[a]*Department of Chemical Engineering, Indian Institute of Technology, Madras*
[b]*Robert Bosch Center for Data Science and AI*
raghur@iitm.ac.in

Abstract

Process systems engineering is a thriving field within chemical engineering. PSE deals with several design and operational tasks that allow process systems to work efficiently and safely. There is a large intersection between PSE tools and Artificial Intelligence (AI) algorithms, recognized for decades now. With unprecedented availability of various forms of data and significant improvement in computational prowess, AI techniques have started to address large and meaningful engineering problems.
In this talk, we will explore the relevance and importance of AI techniques in the next generation process systems engineering applications. Various aspects of PSE and the impact of AI cross-cutting these aspects will be described as outlined below. The focus of this talk will be on the most recent developments and industrial applications that the author has been involved in.

PSE as an area has implications in process modelling, process design, process optimization and process operations. Computer-aided tools are at the centre of all modelling activities. With the advent of AI, automated model building tools are being increasingly researched. Assembling first principles models in a purely data-driven manner is a promising area. Of course, process design is a key aspect of PSE. Design is an inverse problem, where a set of requirements are provided and designs that can satisfy the requirements are desired. As a result, any data-driven modelling tool can also be used in design if there are many exemplar designs that are available for training. As the result, the strength of AI in modelling can be leveraged for this inverse modelling problem. Natural evolution inspired techniques such as genetic algorithms also continue to play an important part in addressing complicated inverse design problems. Recently, reinforcement learning has also been used in solving design problems.

The use of AI techniques in optimization is another exciting area of research. Many core AI algorithms themselves use optimization techniques in their development; use of learning approaches in optimization is an interesting synergy between the two fields. Convex representations using neural networks that allow convex optimization approaches to be used in optimization is an emerging area of research. Other convenient representations from an optimization viewpoint are likely to pursued. An example of such a representation is the difference of convex representation.

The biggest impact of AI in PSE is in the area of process operations. With the ability of systems to collect data at an unprecedented level and the possibility of collecting variegated datatypes, AI algorithms can now be comprehensively explored for various process operations tasks. In process monitoring and operator training, natural language

processing ideas have a large role to play. Further, data from different types of sensors such as vision, noise and so on, over and above the standard sensor data, is likely to revolutionize the way process monitoring and fault detection and diagnosis tasks are performed. This is particularly powerful when data from different plants are centralized allowing for the possibility of transfer learning to occur.

Standard data rectification and gross error detection techniques that used to rely on process models are now being addressed by purely data driven approaches. This brings in several important questions that need to be satisfactorily addressed by the machine learning techniques. Interestingly, sensor placement for data reconciliation, fault detection and diagnosis algorithms that used to rely on process models are also being reimagined as data driven problems.

Work on the use of neural networks and knowledge-based systems in control has been around for more than three decades. However, with renewed interest in AI, these approaches are being explored again with better architectures and larger computational power. Reinforcement learning is a natural approach to address several learning-based control problems. There has been a flurry of activity in this area, and one would expect this area to progress quite rapidly. There are several challenges related to inclusion of constraints, robustness and so on that need to be addressed comprehensively.

Looking forward, two important streams of work can be identified. One of those is the hybridization of existing knowledge with the data driven AI systems. This will be a very profitable area of research and will bring in systems that are explainable, robust and more deployable in engineering problems. Another avenue that will assume significance is moving towards purely unsupervised learning. Many successful applications use supervised and/or semi-supervised learning approaches. However, in the future, several concepts for unsupervised learning will be explored. This, we believe, will lead to truly intelligent process systems that are safe, efficient and robust to inherent variations that cannot be controlled.

Keywords: PSE, AI, ML.

1. Christoph Thon, Benedikt Finke, Arno Kwade and Carsten Schilde

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50011-7

Reinventing the Chemicals/Materials Company: Transitioning to a Sustainable Circular Enterprise

George Stephanopoulos[a,c,*], Bhavik R. Bakshi[a,b], and George Basile[a]

[a]*The Global KAITEKI Center, Arizona State University, Tempe, AZ 85257, USA*
[b]*Department of Chemical Engineering, The Ohio State University, Columbus, OH 43210, USA*
[c]*Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
geosteph(@mit.edu

Abstract

Decarbonization of the Chemicals/Materials Industry (CMI) is feasible, even though technical and economic hurdles exist. However, given the prevailing constraints (economics, green electricity, available biomass), and weak decoupling of GDP from resource utilization, it is clear that the industry cannot achieve the Paris Agreement targets without transforming itself to a Sustainable Circular Enterprise. Such transformation will have deep and broadly-based ramifications on the economy, the structure of CMI and the character of CMI-companies, which need to reinvent themselves. In this paper we will offer data and arguments to substantiate the above statements, and will outline the questions that need to be answered by academic research.

Keywords: Climate change; Circular economy; Sustainability; Process/product redesign; Energy, Environmental systems.

1. The Chemicals and Materials Industry (CMI)

Industrial activities create all the physical products (e.g., cars, agricultural equipment, fertilizers, building construction materials, transportation vehicles, electronic devices, textiles, household items, food, health and security related products, etc.), whose use delivers the services that satisfy specific human needs. In this paper we consider industrial activities over the whole supply chain, from extraction of primary materials (e.g., ores, coal, petroleum, natural gas) or recycling of waste materials, through chemicals-materials-products manufacturing, to the services these products offer and the demand that such services satisfy. For the purposes of this paper, the Chemicals and Materials Industry (CMI) includes the classical chemical/petrochemical industry (organic, inorganic), the cement industry, the iron and steel industry, the non-ferrous materials industries (aluminium, magnesium, copper, and others), and a variety of industrial activities producing smaller amounts of a broad array of chemicals and materials.

Figure 1 shows the complete supply chain of what we consider as the CMI's position in the economy. The supply chain is composed of the following components: (a) *Raw Materials*: Earth stock of ores, minerals, petroleum, coal, natural gas, raw biomass. (b) *Extractive Industry*: Extracts the Raw Materials from the earth stock of raw materials and generates the Feedstocks. For example, natural deposits of petroleum contain gases and solids, which are separated, before the petroleum satisfies the specs to be a feed in a petroleum refinery. Similar extractive processing is required for the preparation of mined

coal, minerals, and natural gas. (c) *Feedstocks*: The form and state of Raw Materials, which satisfy the required specs for feeds to the Processing Industry, in order to produce the Materials that the Manufacturing Industry needs. (d) *Processing Industry*: The set of activities that converts the Feedstocks to Materials. Examples include the conversion of coal, petroleum, or natural gas to fuels and chemicals; iron ore to various grades of iron and steel; conversion of clay, marl, lime, sand, into cement; raw biomass into grades of lignin, sugars, proteins. (e) *Materials*: All chemicals and materials generated by the Processing Industry, which are used for the manufacturing of the various Products the market needs. Examples include: all polymers; various grades of iron and steel; various grades of cement; various types of pulp and paper; etc. (f) *Products*: Buildings, roads, general infrastructure, automobiles, airplanes, electronic devices, pharmaceuticals, household items, etc. (g) *Services*: The satisfaction of the specific need that a product satisfies, such as: housing, clothing, transportation, food, therapeutics, entertainment, security, etc.

In all activities of the supply chain in Figure 1, scrap/wastes are generated and are processed by the "Waste Industry". The corresponding wastes can be reused, repaired, remanufactured, recycled, discarded (landfilled), or destroyed (incinerated).

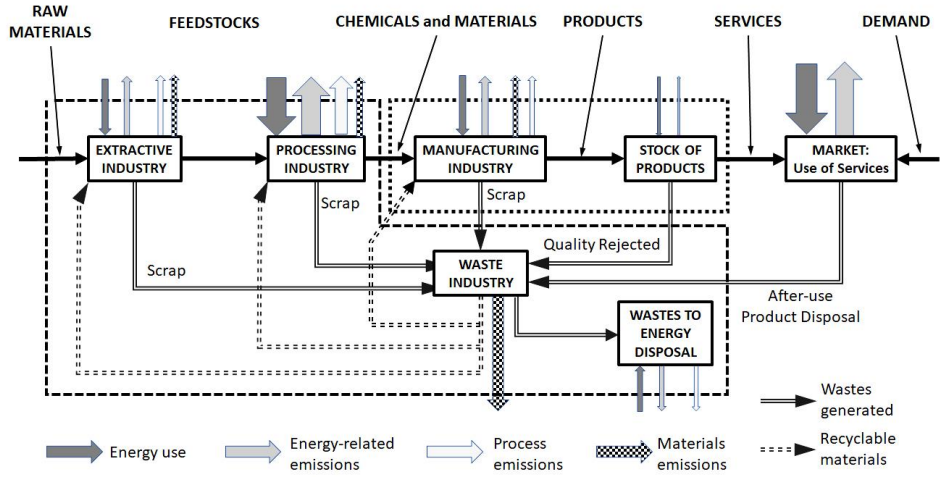


Figure 1. Complete supply chain of the *Chemicals and Materials Industry* (enclosed by the dashed and dotted envelops) from *Raw Materials* to *Services* that satisfy human needs.

2. GHG and Materials Emissions from CMI

Nearly 3/4 of Green House Gas (GHG) emissions (i.e. 73.2%) come from energy use, and 1/3 of it (24.2%) is attributed to industrial use of energy. Adding 5.2% of process-related GHG emissions (primarily from chemicals and cement), we take 30% of total emissions stemming from CMI. When compared with the emissions from transportation (16%), buildings (17.5%, heating, cooling, lighting), and agriculture, forestry and land use (18.4%), we realize that industry is the largest contributor of GHG emissions. The largest contributors of industrial GHG emissions are: iron and steel (24%), cement (19%), chemicals (18%), aluminium (6%), pulp and paper (3%). For the chemicals industry, the largest contributors are; ammonia, olefins (ethylene, propylene), methanol, and aromatics (benzene, toluene, xylenes).

As indicated above, from a strict accounting point of view (i.e. see dashed envelop in Figure 1), the CMI accounts for 30% of global GHG emissions; CO_2, CH_4, N_2O, and F-gases. However, within the scope of CMI's complete supply chain of production and consumption activities, as shown in Figure 1 (i.e. append the activities in the dotted envelop), the GHG emissions corresponding to "chemicals/materials handling and use" are much higher, and account for about 70% of the total GHG emissions. In the extreme case, these emissions include all emissions, except passenger mobility and energy use for residential purposes (space, water heating, and lighting). For example, GHG emissions related to freight transport are not "energy related", because they serve material needs, i.e. move consumer goods around.

In addition, CMI produces large amounts of materials emissions with possible adverse effects on the environment; e.g. 0.5 Gt/yr of plastics with ~40% going to landfills (where carbon is sequestered for hundreds of years) and ~ 20% leaking to the environment with disastrous health effects.

During the period 1990-2016, the global GHG emissions from industry increased by 175%, while the global GDP increased by 110%. Emissions from other sectors were far lower: transport 70%, manufacturing 50%, agriculture 20%, buildings 5%. Furthermore, we note that over the 20-year period of 1998-2018 despite the fact that the share of industry in global GDP has declined, and the annual per cent growth of manufacturing's value added has remained roughly constant, around 2%, the rate of growth of emissions from industry has far outpaced the emissions from any other sector of the world economy (https://data.worldbank.org/indicator/NV.IND.MANF.KD.ZG). The conclusion is clear and inescapable: *Industry's emissions are closely related to the rates of GDP growth*.

The relationship between GDP and GHG emissions has been the subject of many studies and is characterized by the *absolute* and *relative decoupling* between GDP and resource utilization or emissions, which are defined as follows:

$$\begin{aligned} \text{Absolute Decoupling} &= \frac{\Delta(\text{Resource Utilization})}{\Delta(\text{GDP})} < 0 \\ \text{Relative Decoupling} &= \frac{\Delta(\text{Resource Utilization})}{\Delta(\text{GDP})} < 1 \end{aligned} \tag{1}$$

Analysis of nearly 900 studies, based on empirical data on levels of emissions versus levels of consumption and production per capita, have led to important observations, which frame the scope of analysis for the transition of the Chemicals and Materials Industry to a net-zero fossil carbon industry, and can be summarized as follows (Haber, et al., 2020; Mir and Storm, 2016): (i) There is econometric evidence which supports the Carbon-Kuznets-Curve (CKC) hypothesized pattern (see Figure 2), between CO_2-eq emissions production or consumption, and GDP per capita. Such pattern would lead to absolute decoupling after the turning point. (ii) The turning point, for a production-based CKC curve, has been estimated to lie in the range of 50 to 100 $GtCO_2$-eq, which is far beyond the COP21 emissions reduction goals. (iii) Examples of consumption-based absolute decoupling are very rare. (iv) Relative decoupling is frequent for material use as well as GHG and CO_2 emissions, but not for useful exergy. From the above observations we reach two important conclusions: (i) Absolute decoupling, i.e. reduction of resource (energy, materials) utilization and GHG emissions per unit of GDP growth, cannot be achieved through observed decoupling rates. (ii) To reach the Paris Agreement goals by 2050, efforts to decouple resource utilization from GDP growth are necessary conditions

but not sufficient. Sufficiency-oriented strategies must include strict enforcement of absolute targets.

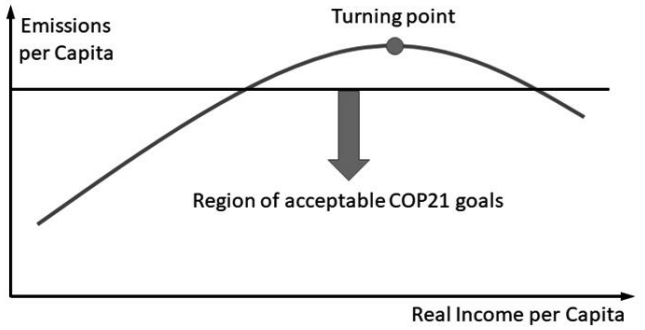


Figure 2. The Carbon-Kuznets-Curve (CKC) relationship between CO2-eq emissions and real income.

3. Towards the Circularization of CMI

The following formula offers a simple way for computing GHG emissions. It also identifies the points where major interventions would lead to emissions reductions.

$$G = \frac{G}{E} \times \frac{E}{M} \times \frac{M}{Pr} \times \frac{Pr}{S} \times S \tag{2}$$

G, GHG emissions. *G/E*, Emissions Intensity; GHG emissions per unit of energy used. *E/M*, Energy Intensity; energy used per unit of material. *M/Pr*, Materials Intensity; materials used per unit of product to create the product and maintain stock of the product. It depends both on the design of the product and on the scrap discarded during its production. *Pr/S*, Product-Service Intensity; it determines the level of service provided by a product, and depends on whether the product is consumable or durable. *S*=Total Service Demand=(Population)×(GDP/capita); the total global demand for service. Traditional programs of Continuous Improvement (KAIZEN) and process optimization can reduce the values of the first three factors, *G/E*, *E/M*, and *M/Pr*, with potential reduction of emissions by 25% - 40%. The remaining must come from the following sources: (i) Renewable energy supplies and major technological breakthroughs, such as: carbon capture sequestration and/or use. (ii) Major reductions in Materials Intensity (*M/Pr*), Product-Service Intensity (*Pr/S*), and Total Service Demand (*S*). *S* is directly related to GDP/capita and we discussed earlier. (iii) Introduction of Circular practices: reuse, repair, remanufacturing, recycle products, wastes or scraps. The conclusion is inescapable: *To meet the Paris Agreement goals we must do the following:* (1) *Change the focus from the energy sector to the chemicals/materials sector*. (2) *Enhance circularization of all supply chain activities* (*processing, manufacturing, distributions, sales, recovery, reuse, reprocessing, remanufacturing*). (3) *Reduce virgin material demand by extending the percent utilization of all material products (housing, mobility, nutrition, communications, consumables).*

A series of obstacles prevents the full and idealized circularization of the CMI. These obstacles are: (a) *Products may be too complex* to recycle, reuse, or remanufacture. The large-scale use of synthetic materials makes the closing of the cycle nearly impossible. Furthermore, the recycling of synthetic materials (e.g. polymers) invariably produces inferior materials. Redesign of products with easily assembled and disassembled material components, as well as extensive

use of biomass-based, degradable materials, could address these concerns to a large degree. (b) *How do you recycle fossil fuels?* A large part of fossil materials is used to provide energy for heating and electricity, leading to exhaust streams, whose useful energy has diminished and is unusable. Two options are open: Shift to progressively larger amounts of renewable energy, and optimize the process of capturing and sequestering or/and using CO_2. (c) *Growing human needs*. Continuous growth of material human needs leads to the extraction of continuously larger amounts of natural resources. By increasing the amount of materials recycled, we may be able to establish steady state. However, in order to achieve this, we ned to have a holistic approach to the circular economy that involves the complete life cycle of materials; something that many advocates of the circular economy fail to account for, by focusing on limited segments of the whole. (d) *Accumulation of natural resources*. A significant portion (about 30-35%) of processed natural resources remains in the economy and accumulates in the form of buildings, infrastructure, and consumer products; it is not recycled, destroyed, or disposed in a landfill. Therefore, the circular economy is not truly a steady state situation, but one which continues to be extractive economy, with increasing inventory of materials over various time horizons. (e) *Recycling and reuse are not enough*. The economy needs to adjust to the above limitations, by increasing the percent utilization of all material goods. For example, we cannot have a sustainable circular economy with cars unused more than 90% of the time, office buildings used only 60% of the time, or more than 30% of food wasted.

4. The Research Scope of Sustainable Circular CMI (S-CCMI)

So far, the prevailing discussion on circular economy has been driven by the following simple definition of the sustainable circular CMI (S-CCMI): It is an economic model, focused on designing and manufacturing products, components and chemicals/materials for reuse, remanufacturing and recycling. However, the S-CCMI must sustain economic growth, and this definition is not sufficient to delineate its actionable scope. While the specific characteristics of S-CCMI can differ for different sectors of CMI, they must be driven by a simple principle: *The resource inputs and recycled materials should maintain dynamic material balances of "wastes" at the sustainably highest allowable materials-accumulation levels.* Pure "steady state" requirements are impossible; materials accumulate in the economy continuously. For example, carbon extracted from earth is equal to the amount of carbon returned to the earth, while the amount of carbon accumulated in the system remains below the sustainably highest level. The "sustainably highest level" allowance is determined by climate change (GHG emissions) and environmental impact (materials emissions) constraints.

Design elements of the S-CCMI. The above simple principle has a broad range of implications for all activities in the "materials handling and use" network of Figure 1. At The Global KAITEKI Center (TGKC) of Arizona State University, we have undertaken an extensive research program, which addresses all questions related to the transition of CMI to S-CCMI, such as: (i) Reduction in the input and use of non-renewable material resources. (ii) Reduction in the generation of wastes and emissions (GHG and materials). (iii) Increase in inner materials use and recycle up to the sustainably highest level of materials use. (iv) Redesign of products and associated components, materials, and chemicals to facilitate recycle, and reuse. (v) Development of new technologies for the capture, sequestration and use of carbon.

Transformation of CMI companies to S-CCMI companies. Aspects of the research program evaluate the following implications on the structure and operations of the CMI companies, as they transform to S-CCMI companies: (1) Transition from centralized and vertically integrated, open-chain large-scale processing and manufacturing systems, to smaller-size,

decentralized, distributed and locally managed processing and manufacturing activities. (2) Economies of scale, which have dominated the large-scale processing in CMI will no longer be the drivers for large-scale investments. Materials, transportation, logistical services, and financial costs will drive the structure of S-CCMI. (3) Redefinition of the scope of the business activities, with the CMI companies transforming themselves from producers of virgin chemicals and materials, to producers of integrated components and products, and suppliers of technical services to support the life-cycle of their products.

The economics of transition from CMI to S-CCMI. The economic benefits of such transformation S-CCMI companies have been estimated to be very attractive: (1) Significantly higher returns on investment. (2) Reduction in volatility between supply and demand, due to effective recycling and self-regulating system, leading to resilient economic growth. (3) Stronger market position and competitive differentiation for the companies, which espouse the future of S-CCMI. This will be particularly true for the companies, which are presently strong and have strong R&D, engineering, and supply-chain market positions.

Societal and human adjustments during the transition from CMI to S-CCMI. The consumers will need to adjust their behavior away from "owning physical products" towards "using services of physical products", but the benefits are many and attractive: (i) The transition *from "owning, using and disposing" to "using and returning"* will force improvement of service quality, durability and reliability of products. (ii) This transition will foster the appearance of new emerging trends of sharing, lending, swapping etc. that will benefit the consumers. (iii) The companies of present CMI employ very few people per dollar of asset values. The S-CCMI will increase these numbers significantly and fuel more predictable and sustainable demand of products and services at higher levels. (iv) The demands of S-CCMI can only be met by human resources of significantly higher education and skills. This will be in-line with historical expectations.

5. Conclusions

The transition of the CMI to a Sustainable Circular CMI is a necessary (and possibly sufficient) condition for industry to maintain under control the growing mountains of environmental "wastes"; GHG emissions and discarded materials. The implications are broad and deep and require restructuring of CMI, reinvention of the CMI companies and realignment of human behavioural traits in the new market place. The question is not whether to transition to S-CCMI or not, but how; the scope of research at TGKC.

Acknowledgements

The authors gratefully acknowledge the support from "The Global KAITEKI Center" at Arizona State University (ASU), a university-industry partnership between ASU and The KAITEKI Institute of Mitsubishi Chemical Holdings Corporation.

References

M. Brudermüller, 2020, How to build a more climate-friendly chemical industry, World Economic Forum, Annual meeting.

H. Haber, et al., 2020, A systematic review of the evidence on decoupling of GDP, resource use and GHG emissions, part II: synthesizing the insights, Environ. Res. Lett. 15 065003

G. Mir and S. Storm, 2016, Carbon Emissions and Economic Growth: Production-based versus Consumption-based Evidence on Decoupling, Working Paper No. 41, Institute for New Economic Thinking.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50012-9

Value Chain Optimization of a Xylitol Biorefinery with Delaunay Triangulation Regression Models

Nikolaus I. Vollmer[a*], Krist V. Gernaey[a], Gürkan Sin[a]

[a]*Process and Systems Engineering Center (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søltofts Plads, Building 228A, 2800, Kgs. Lyngby, Denmark*
nikov@kt.dtu.dk

Abstract

The presented work focuses on the value chain optimization of a conceptually designed biorefinery, considering the plant capacity and other logistic and design constraints. An existing framework is used to create surrogate models, which are then used to reformulate the underlying optimization problem for performing value chain optimization. The used Delaunay triangulation regression surrogate model performs well and is a suitable candidate for value chain optimization. The results indicate an apparent effect of the economics of scale, and the market conditions mainly constrain the designed value chain.

Keywords: Biorefinery, Surrogate Modelling, Delaunay Triangulation, Mixed-Integer Linear Program, Value Chain Optimization

1. Introduction

A key approach in expediting the transition towards a bio-based economy is the conceptual design and implementation of value chains based on integrated second-generation biorefineries. Although these biorefineries have been investigated for several decades, and despite their vast potential regarding a sustainable production of fuels and chemicals, the major challenge remaining concerns their economic viability (Ubando et al., 2020). Among other factors that influence the economic viability, the capacity and location of the plant and the design of suitable feedstock and product supply chains are essential considerations to take (Gargalo et al., 2017). What is classically referred to as economies of scale can improve the economic key performance indicators (KPIs) of a plant up to a certain point, as the capital expenditures do not increase proportionally with the plant capacity. In contrast, additional necessary equipment, as well as increased operational costs for logistics, can thwart this effect.

Hence, it is crucial to conceptually design both the biorefinery process and the entire value chain in which the biorefinery will be embedded. Vollmer et al. (2021) have recently developed a framework (S3O) that allows for the conceptual design of biorefinery processes based on mechanistic modeling for all unit operations in the process. The framework utilizes flowsheet simulations and different types of surrogate models to perform a superstructure optimization to determine candidate process topologies. This procedure is applied to eliminate nonlinearities, which are inherent to all unit operation models and constitute the superstructure optimization as a mixed-integer nonlinear optimization problem (MINLP) (Vollmer et al., 2021b). The surrogate models aim either at linearizing the original model or eliminating the integer variables.

Similarly, in value chain optimization, nonlinear models are commonly linearized with piecewise linear approaches (Krämer et al., 2021).

In the scope of this work, the S3O framework is extended by using the already present Delaunay Triangulation Regression (DTR) surrogate model to perform value chain optimization based on a conceptually designed process through the framework. By flowsheet sampling with the process flowsheet with relevant input and output variables for the value chain optimization, piecewise linear DTR surrogate models are created. The value chain optimization is set up and constituted as a mixed-integer linear program (MILP), using the DTR surrogate model and solved with a suitable solver. This solution is benchmarked with a Gaussian Process Regression (GPR) surrogate model, which is equally incorporated in the S3O framework. The solution is analyzed and compared to the initial base case process design to conclude how to design an optimal value chain.

2. Methodology

2.1. Mechanistic Process Models

The basis of the work in this manuscript is a running flowsheet simulation model created in the first step of the S3O framework. Each flowsheet model consists of various unit operation models. All unit operation models are mechanistic models, consisting of mass and energy balances and a kinetic description of the respective reaction or transfer process. Based on the mass and energy balance calculations, the fixed capital investment for each unit operation is calculated by a plant capacity ratio based on a report by the National Renewable Energy Laboratory (NREL) regarding a similar biorefinery setup. Based on the fixed capital investment and other report data, capital expenditures and operational expenses (CAPEX and OPEX) and different KPIs, e.g., the net present value (NPV) of the plant, are calculated. Input parameters for the flowsheet simulation can be operational variables and the feedstock mass, and other setup parameters. Output variables can be all mass and energy flows, as well as design parameters, e.g., vessel sizes, or economic variables referring to CAPEX or OPEX, or ultimately the KPIs of the plant. All flowsheet simulations are performed through the S3O framework as it is implemented in MATLAB.

2.2. Delaunay Triangulation Regression

DTR is based on a triangulation of points as a logical extension of piecewise linear regression in any dimension (Vollmer et al., 2021b). In a two-dimensional case, a triangulation consists of triangles or 2-simplices. For any dimension n, the triangulation hence consists of n-simplices. Each n-simplex itself is constituted by n+1 vertices. Within each simplex, each point can be described as a linear affine combination of the vertices. In this manuscript, the vertices are assigned to be the sampling points of the flowsheet simulations. The DTR utilizes Delaunay triangulation, which imposes the criterion to each simplex not to contain any other vertex of another simplex within the circumcircle – or its pendant in other dimensions – of the simplex. For a more detailed description, the reader is referred to Vollmer et al. (2021b). The results show an excellent functionality of the DTR surrogate for superstructure optimization applications within the S3O framework, despite impaired validation metrics. Furthermore, DTR has been used in other research works for performing operations optimization (Obermeier et al., 2021). In this work, the DTR surrogate model is created with functions provided through the scipy library for Python.

2.3. Gaussian Process Regression

Another very popular type of surrogate model is a GPR model, a machine learning model with a broad variety of possible applications (McBride and Sundmacher, 2019). The model itself harvests its potential through a stochastic process – the eponymous Gaussian Process – which correlates the given input and output data, in this case, the flowsheet samples (Vollmer et al., 2021b). Also, using GPR surrogate models for process design applications within the S3O frameworks shows good functionality with good validation metrics (Vollmer et al., 2021a). In this work, the GPR functionalities of the Statistics & Machine Learning Toolbox in MATLAB are used.

2.4. Value Chain Optimization

The original value chain optimization in this manuscript is an MINLP of the form given in Eq. (1).

$$MINLP: \quad \begin{aligned} &\max z = f(x,y) \\ &s.t. \;\; g(x,y) \leq 0 \\ &h(x,y) = 0 \\ &x \in X, \;\; y \in [0,1] \end{aligned} \tag{1}$$

The functional relation $f(x,y)$ denotes the flowsheet simulation, including continuous input variables x and binary input variables y. The objective z is equally one of the model outputs. Inequality constraints are represented by $g(x,y)$ and equality constraints are represented by $h(x,y) = 0$. Both inequality and equality constraints can describe conditions regarding logistics, markets, plant capacity, location, and supply and demand for the biorefinery. When utilizing the DTR surrogate model, the MINLP converts into a MILP due to eliminating the nonlinearities by the piecewise linear approach. When using the GPR surrogate model, the MINLP converts into a set of nonlinear programs (NLP), with each element of the set representing one realization of the combinations of all binary variables that need to be solved separately. The MILP can be solved with the GUROBI solver, whereas each NLP is solved with the fmincon solver in MATLAB while employing a multi-start procedure to guarantee global optimality.

3. Application

3.1. Case Study

The biorefinery in this case study is a multi-product biorefinery that converts wheat straw into xylitol, succinic acid, and heat. The latter is used for process integration purposes in the downstream processing of the former two products. It consists of a biomass pretreatment unit, a unit for enzymatic hydrolysis, two evaporation units in the upstream process, two fermentation units for the production of xylitol and succinic acid, two evaporation units, four crystallization units, and one combustion unit for the lignin.

3.2. Optimization Problem

The original biorefinery is designed for an annual feedstock mass of $m = 150.000\,t$ designed for being located in Denmark, corresponding to approximately 3% of the nationally harvested amount of wheat straw (Danish Agriculture & Food Council, 2015). In order to see a significant effect of the economies of scale, the capacity of the biorefinery could be potentially increased up to $m = 600.000\,t$, as practiced in the mentioned report of the NREL (Humbird et al., 2011). However, as wheat straw is harvested and centrally collected, a higher amount of feedstock correlates with a longer

transportation distance. Per each full additional 150.000 t of feedstock, it is considered to increase the transportation distance stepwise by 100km, with a transportation price of $0.05\ \$/(t \cdot km)$. As this increases the costs for transportation, a potential option is to not install one plant with a capacity of bigger than $m = 300.000\ t$, but to install two plants with the ability to be able to process $m = 600.000\ t$ in a decentralized manner which alleviates the economic impact (Galanopoulos et al., 2020). This is expressed by a binary variable y_p as part of the optimization problem. Lastly, as fermentation processes need equipment for inoculation, a higher capacity than the original one requires an additional fermentation tank, which increases the CAPEX of the plant. The costs of the additional tank scale linearly with the feedstock mass. The capacity itself, however, is calculated through the capacity ratio. For the flowsheet simulations, the operational variables, as well as the feedstock mass and the plant number y_p are used. The effects of all three mentioned aspects are calculated in the flowsheet simulation. As output, the NPV of the plant and the mass of produced xylitol and succinic acid are calculated. The objective in the optimization is set to be the NPV of the plant, and the mass of xylitol and succinic acid are constrained to be $g_1 = m_{xyo} \in [5,33]\%$ and $g_2 = m_{xyo} \in [20,50]\%$ of the global production of the respective substance. The lower limits are imposed to assure a minimal production, whereas the upper limits are imposed to avoid decreasing market prices, which would negatively influence the plant profitability due to the comparatively small market size of both products.

3.3. Results

For illustrative purposes, the NPV as the objective function of the feedstock mass with fixed operational conditions is displayed in Figure 1.

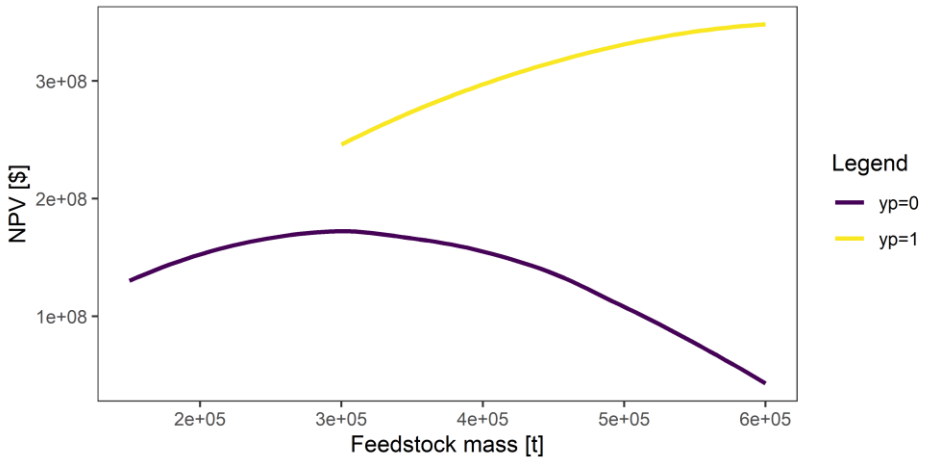


Figure 1: Plot of the objective function with the feedstock mass as a variable for $y_p = 0$ and $y_p = 1$

The effect of the economies of scale, the constraining effects of increased costs for logistics and equipment, and the reduction of the expenses through decentralized production are visible. For the value chain optimization, the process flowsheet model is sampled with N=1000 simulation samples for all five operational variables, the feedstock mass, and the binary variable y_p with two sets of the sample size sampled by Latin Hypercube sampling for each realization of y_p. The DTR and the GPR surrogate

models are fitted to the sampling data and cross-validated. The results of the cross-validation for both models are presented in Table 1.

Table 1: Results of the surrogate model validation for the NPV as output variable

	DTR		GPR	
	$y_p = 0$	$y_p = 1$	$y_p = 0$	$y_p = 1$
R^2_{train}	1	1	1	1
R^2_{test}	0.552	0.408	0.957	0.937
$RMSE_{train}$	0	0	$3.19 \cdot 10^5$	$1.04 \cdot 10^5$
$RMSE_{test}$	$3.07 \cdot 10^8$	$2.97 \cdot 10^8$	$9.48 \cdot 10^6$	$1.57 \cdot 10^5$

It becomes evident that the DTR surrogate metrics are impaired compared to the ones of the GPR. Particularly for the case of $y_p = 1$, the DTR surrogate model does fail to predict accurately, while the GPR shows improved metrics compared to the case of $y_p = 0$. The optimization problem for the value chain optimization is set up as described in section 2.4 with all variables, the objective function, and the constraints as described in section 3.2. The results are presented in Table 2.

Table 2: Results of the value chain optimization with the optimization result (left) and the validation simulation result (right) for each binary decision and surrogate model

	DTR				GPR			
	$y_p = 0$		$y_p = 1$		$y_p = 0$		$y_p = 1$	
M_{feed} $[10^3 \cdot t]$	302		293		292		292	
$\boldsymbol{NPV}$ $\boldsymbol{[10^8 \cdot \$]}$	**1.96**	**2.47**	**2.93**	**3.29**	**7.1**	**3.56**	**5.8**	**2.60**
M_{xyo} $[10^3 \cdot t]$	19.8	20.4	22.1	22.2	29.1	22.4	21.8	20.2
M_{suc} $[10^3 \cdot t]$	35.9	38.0	35.9	35.1	35.9	35.9	35.9	35.9

Primarily, the result from both the optimization with the DTR and the GPR surrogate shows that the effect of the economics of scale is visible, and the feedstock capacity is increased to a maximally feasible limit, which is imposed by the market saturation bounds. The effects of increased transportation costs and additional equipment do not influence the objective to the same degree. In other studies investigating products as bioethanol with higher market volumes, these effects become more significant (Galanopoulos et al., 2020). Furthermore, the operational conditions were not significantly influenced. Also, the results of both optimization problems with the different surrogate models agree, indicating that the DTR model is a suitable candidate for value chain optimization despite insufficient validation metrics, since it performs better in predicting the objective function and the constraints. The differences in prediction with the GPR surrogate model for the case of $y_p = 1$ and $y_p = 0$ can be explained by the different fit of each model, which is reflected by the validation metrics. The differences in prediction for the DTR surrogate model can be equally attributed to the different fit expressed by the validation metrics. This indicates a potential increase in prediction quality for both models by using larger sampling sizes. It is noted that the economic metrics are calculated with fixed price considerations for feedstock and products. Hence the emphasis is on the trend of NPV rather than its absolute value. Further studies of uncertainties that will yield a distribution of NPV are needed as the volatility in the market prices for feedstock and products, as well as the market demand uncertainty, will affect the nominal values presented here.

4. Conclusion

Based on a conceptually designed process for an integrated second-generation biorefinery, this study aimed to design an optimal value chain with the given process design by taking into account a varying feedstock mass and other logistic and design constraints for the plant. The value chain optimization was performed using two surrogate models to transform the underlying MINLP into either a MILP or a set of NLPs. The results from solving the optimization problems show results in agreement with each other. This proves the suitability of DTR surrogate models for the use in value chain optimization problems, and this despite impaired validation metrics of the surrogate model, which has been addressed before (Vollmer et al., 2021a; Vollmer et al., 2021b) With regards to the actual value chain, it becomes evident that the economics of scale positively influence the KPIs of the biorefinery. The limitations for economic resilience are found in the actual market sizes of the products rather than in the additional costs for increased CAPEX and OPEX of a larger biorefinery. These results are essential and contribute to facilitating the further design and implementation of biobased value chains to create more sustainable production patterns in the future.

Acknowledgments

The research project is part of the Fermentation-Based Biomanufacturing Initiative funded by the Novo Nordisk Foundation (Grant no. NNF17SA0031362).

References

Danish Agriculture & Food Council, 2015, Facts and Figures: Denmark – a Food and Farming Country, 1-56

C. Galanopoulos et al., 2020, An integrated methodology for the economic and environmental assessment of a biorefinery supply chain, Chemical Engineering Research and Design, 160, 199-215.

C.L. Gargalo et al., 2017, Optimal Design and Planning of Glycerol-Based Biorefinery Supply Chains under Uncertainty, Industrial & Engineering Chemistry Research, 56, 41, 11870-11893

D.Humbird et al., Process design and economics for biochemical conversion of lignocellulosic biomass to ethanol: dilute-acid pretreatment and enzymatic hydrolysis of corn stover, No. NREL/TP-5100-47764, National Renewable Energy Laboratory, Golden (CO), United States

A. Kämper et al., 2021, AutoMoG: Automated data-driven Model Generation of multi-energy systems using piecewise-linear regression, Computers & Chemical Engineering, 145, 107162

K. McBride and K. Sundmacher, 2019, Overview of surrogate modeling in chemical process engineering, Chemie Ingenieur Technik, 91, 3, 228-239

A. Obermeier et al., 2021, Generation of linear-based surrogate models from nonlinear functional relationships for use in scheduling formulation, Computers & Chemical Engineering, 146, 107203

A.T. Ubando et al., 2020, Biorefineries in circular bioeconomy: A comprehensive review. Bioresource Technology, 299, 122585

N.I. Vollmer et al., 2021a, Benchmarking of Surrogate Models for the Conceptual Process Design of Biorefineries, Computer Aided Chemical Engineering, 50, 475-480

N.I. Vollmer et al., 2021b, Synergistic optimization framework for the process synthesis and design of biorefineries, Frontiers of Chemical Science and Engineering, 1-23

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50013-0

Evaluating the Impact of Model Uncertainties in Superstructure Optimization to Reduce the Experimental Effort

Stefanie Kaiser[a*], Sebastian Engell[a]

[a] *Department of Chemical Engineering, TU Dortmund University, Emil-Figge-Str. 70, 44227 Dortmund, Germany*
stefanie2.kaiser@tu-dortmund.de

Abstract

Optimization-based process design can be an efficient tool for finding synergies between process units, but it strongly relies on accurate process models. Hence, experiments for model refinement may be necessary. We present an optimization-based methodology to enhance the process development by integrating superstructure optimization under uncertainties and optimal design of experiments. In this manner, experiments for model refinement can be focussed on the parameters which are critical for discrete design decisions. These parameters are identified by a local discrimination analysis followed by a computation of the partial dependence or the permutation feature importance. The methodology is applied to the hydroaminomethylation of 1-decene. It is shown that it reduces the number of experiments needed for the decision between alternative process structures.

Keywords: Superstructure Optimization, Process Design, Optimal Design of Experiments, Linear Discrimination Analysis.

1. Introduction

Superstructure optimization has been developed as a tool to support design decisions in process development by the optimization of discrete and continuous parameters of a superstructure that describes a range of process alternatives. The formulation of the superstructure and the solution of the resulting large mixed-integer optimization problems have been widely studied ((Chen & Grossmann, 2017), (Skiborowski et al., 2014)). However, the prerequisite of the application of the approach is the availability of models that describe the chemical and physical phenomena in the different pieces of equipment as well as the necessary investments and the costs of operation accurately. Since the models that are available for process design are usually uncertain in the early process design phases, where nonetheless often important structural decisions are taken, these uncertainties should be considered in the optimization. Steimel and Engell (2016) proposed a two-stage formulation for superstructure optimization under uncertainty where the uncertainty is modelled by discrete scenarios and the design degrees of freedom are identical for all scenarios but the operational degrees of freedom are adapted to the realization of the uncertainties, i.e. the real behavior of the plant. However, often there will be several process structures that are optimal for different scenarios, and the uncertainty should be reduced by experimental work to obtain a unique solution. In order to identify the optimal design with the smallest experimental effort, the authors proposed an integrated methodology that combines superstructure optimization under uncertainty,

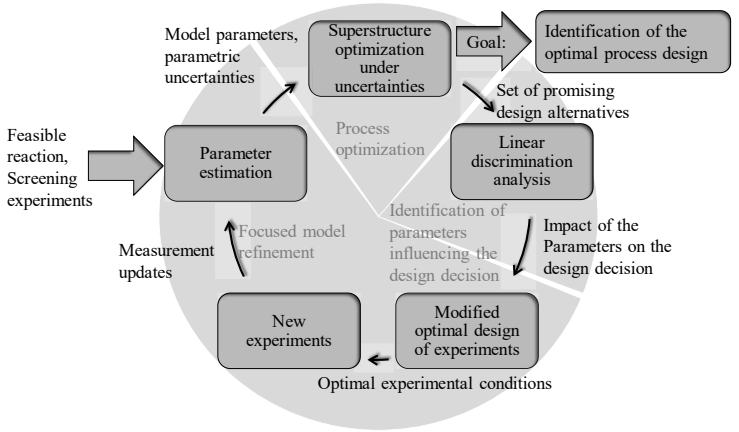


Figure 1: Schematic representation of the integrated methodology.

sensitivity analysis, and optimal design of experiments ((Kaiser & Engell, 2020), (Kaiser et al., 2021)). The parameters with the highest impact on the process cost for the most promising design are identified using a sensitivity analysis. A modified optimal design of experiment is used to plan experiments that focus on determining these parameters.

Our earlier work showed that the process design could be accelerated, but it did not yet take the impact of the uncertain parameters on the structural decisions into account. We improve the methodology by evaluating the impact of the model parameters on the discrete design decisions and planning experiments that are targeting to determine the parameters that influence the discrete design decisions the most. In the following, we will describe the methodology and apply it to the hydroaminomethylation of 1-decene.

2. Methodology

The idea is to integrate superstructure optimization under uncertainty, the identification of relevant parameters, the design of focused experiments, and parameter estimation as can be seen in Figure 1. By this, efficient experiments can be planned for model refinement that are focused on determining the parameters that have the highest impact one the discrete design decisions. The single steps are explained in the next section.

2.1. Superstructure optimization under uncertainty

For superstructure optimization under uncertainty Steimel and Engell (2016) proposed to formulate a two-stage mixed integer optimization problem.

$$\begin{aligned} &\min_{y_d, y_c, x_\omega} G(y_d, y_c) + \sum_{\omega=1}^{\Omega} \pi_\omega F_\omega(y_d, y_c, x_\omega, z_\omega) \\ \text{s.t.}\quad &h(y_d, y_c, x_\omega, z_\omega) = 0 \\ &g(y_d, y_c, x_\omega, z_\omega) \leq 0 \end{aligned} \tag{1}$$

The objective function (1) consists of two terms. The first term accounts for the cost when fixing the discrete (y_d) and continuous (y_c) design degrees of freedom, which cannot be adapted to the realization of the uncertainties. The second term is the weighted sum of the scenario dependent costs for the Ω discrete scenarios of the uncertain parameters. The operational degrees of freedom x_ω are assumed to be adapted to the realization of the uncertainty and are therefore optimized separately for each scenario.

2.2. Linear discrimination analysis

To identify the effect that the uncertainties have on the design decision, the outputs are divided into g classes K_l– here the best design decisions – depending on the values of the uncertain parameters. By a linear discrimination analysis, the features that separate the classes can be identified. A realization x_i of a feature carrier $\boldsymbol{e}$ is assigned to the class K_l for which the value of the discrimination function is maximum (Rinne, 2008):

$$d_{l*}(x_i) = \max_{1 \le l \le g} d_l(x_i) \Rightarrow x_i \in K_{l*}. \tag{2}$$

Supposing that a number or realizations (samples) x_i have been classified, arbitrary values of $\boldsymbol{e}$ can be assigned to the classes under the assumption that the features are independent, Gaussian distributed, and have an equal feature-independent variance

$$\boldsymbol{\Sigma}_l = \sigma^2 \boldsymbol{I} \; \forall \; l. \tag{3}$$

A realization with vector $\boldsymbol{x}$ is assigned to the class with the minimum Euclidean distance between the class center $\boldsymbol{\mu}_l$ and $\boldsymbol{x}$. The linear discrimination function is defined as:

$$d_l(\boldsymbol{x}) = -\frac{1}{2\sigma^2}\boldsymbol{\mu}_l'\boldsymbol{\mu}_l + \frac{1}{\sigma^2}\boldsymbol{\mu}_l'\boldsymbol{x}. \tag{4}$$

$\boldsymbol{\mu}_l$ is computed as the average of the classified samples that belong to class l .
The relationship between the features and the predicted class can be represented by the partial dependence (PD) as described by (Greenwell, 2017). It is computed as the average probability of the output belonging to the different classes when the value of this feature is varied, averaged over the values of all other features. The assignment of the class for a value of the feature is based on (4) and the averaging is done over the set of the classified realizations of the parameter vector (samples). This gives the fractions of the predictions of each design decision being the optimum one depending on the values of the individual features (parameters in our case). If $\boldsymbol{x}^s$ is the feature of interest and $\boldsymbol{X}^{\boldsymbol{c}}$ is its complement (i.e. all other features), the partial dependence of the output f at $\boldsymbol{x}_{\boldsymbol{s}}$ is defined as:

$$f^s(\boldsymbol{x}^s) = E_C[f(\boldsymbol{x}^s, \boldsymbol{X}^c)] = \int f(\boldsymbol{x}^s, \boldsymbol{X}^c) p_c(\boldsymbol{X}^c) d\boldsymbol{X}^c \tag{5}$$

The results of the PD are compared to the permutation feature importance (PFI) that also capture interaction effects (Fisher et al., 2019). For each feature, the values are exchanged with values from a different sample to generate a new feature matrix X^{perm}. The classification of the discrimination model of X^{perm} is then compared to the true optimal discrete design as indicated in the training data and the number of wrong classifications is compared for all features. If permuting one feature leads to a wrong classification, it implies that this feature as a large influence on the model output.

2.3. Optimal design of experiments

To design an experiment to determine the parameters that influence the design decision, an optimal design of experiment (ODoE) is used. In ODoE a metric of the inverse of the Fisher information matrix (FIM) is minimized (Franceschini & Macchietto, 2008).

$$\boldsymbol{FIM} = \sum_{\tau=\tau_1}^{\tau_N} \boldsymbol{Q}^{\boldsymbol{T}}(\boldsymbol{u}_\tau)\mathbf{diag}^{-\mathbf{1}}(\sigma_1^2, \dots, \sigma_{n_y}^2)\boldsymbol{Q}(\boldsymbol{u}_\tau) \tag{6}$$

with $\boldsymbol{Q}(\boldsymbol{u}_\tau)$ being the matrix of the derivatives of the model outputs with respect to the parameters of the experiment with the input $\boldsymbol{u}_{\boldsymbol{\tau}}$. As only some of the parameters will have an impact on the design decision, as identified by PD and PFI, only these are included in the ODoE. Here an A-optimal design of experiments is used which minimizes the trace of the inverse of the FIM.

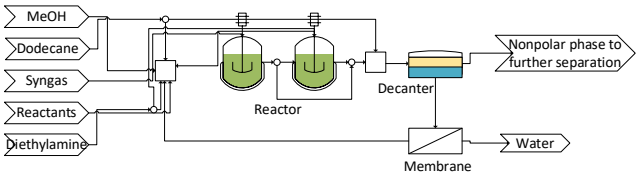


Figure 2: Superstructure of the process.

3. Case study

The presented methodology is applied to the homogeneously catalysed hydroaminomethylation of 1-decene in a thermomorphic solvent system of methanol and dodecane. In this process, long-chain amines are produced and water is formed as a by-product. Thermomorphic solvent systems are single phase at reaction temperature and separate into two phases when cooled down which enables a recycling of the expensive catalysts. The reaction can either be performed as a tandem reaction (HAM) in one reactor or in two subsequent steps, thus performing the two subsequent reactions hydroformylation (HYFO) and reductive amination (RA) in different reactors. The superstructure is depicted in Figure 2. The process consists of three steps: reaction, separation and removal of water. Kinetic models for the two reactions steps can be found in (Hentschel et al., 2015) and (Kirschtowski et al., 2021). The HAM was modelled by combining both model structures and fitting the parameters to twelve experiments. The gas solubilities as well as the phase separation are predicted using the equation of state PC-SAFT. As the iterative solution of the PC-SAFT equations is not feasible in the optimization, surrogate models were trained as proposed in (Nentwich & Engell, 2019). The membrane separation is modelled using a solution-diffusion model. The uncertainties considered in this case study are the pre-exponential factors and the activation energies of all reaction rates resulting in 31 uncertain parameters. The binary design degrees of freedom are the choices whether the tandem reaction or the subsequent reactions are used and whether the nonpolar solvent dodecane is fed before or after the reaction, and the continuous design decisions are the volumes of the reactors and the area of the membrane. As recourse variables, the temperatures in the reactors and the decanter, the partial pressures of syngas, the solvent ratio and the catalyst concentration are optimized. The cost function is the production cost per kg of product for a constant capacity of 10.000 t/a.

3.1. Application of the integrated methodology

The superstructure optimization was performed for the four structurally different process alternatives. For each alternative, the predicted costs for the best designs are shown in Figure 3 (left). Each line in the figure represents one scenario of uncertain parameters. Initially, design 1 is optimal for most but not for all scenarios. As one cannot make a design decision based on this result, a further model refinement is necessary. As a next step, we analyzed which parameters influence the design decision. Therefore, a linear discrimination analysis was performed. Designs 2 and 4 are not optimal for any scenario and therefore, only designs 1 and 3 are considered. The influence of the parameters on the class allocation was analysed via PD and PFI. The partial dependence plots are shown in Figure 4, where the scores of all 31 parameters are presented. One can see that

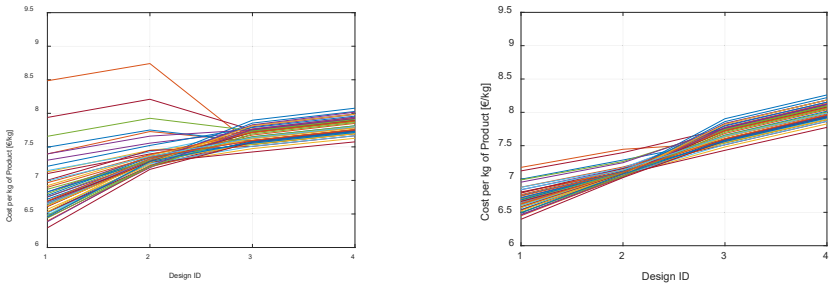


Figure 3: Predicted costs for the four best structurally different designs for 50 scenarios of the uncertain parameters after the initial superstructure optimization (left) and after 7 iterations. The design IDs indicate the structurally different designs. Design 1 and design 2 correspond to the tandem reaction with a dodecane feed after and before the reaction respectively and design 3 and 4 to the two subsequent reactions with a dodecane feed after and before the reaction.

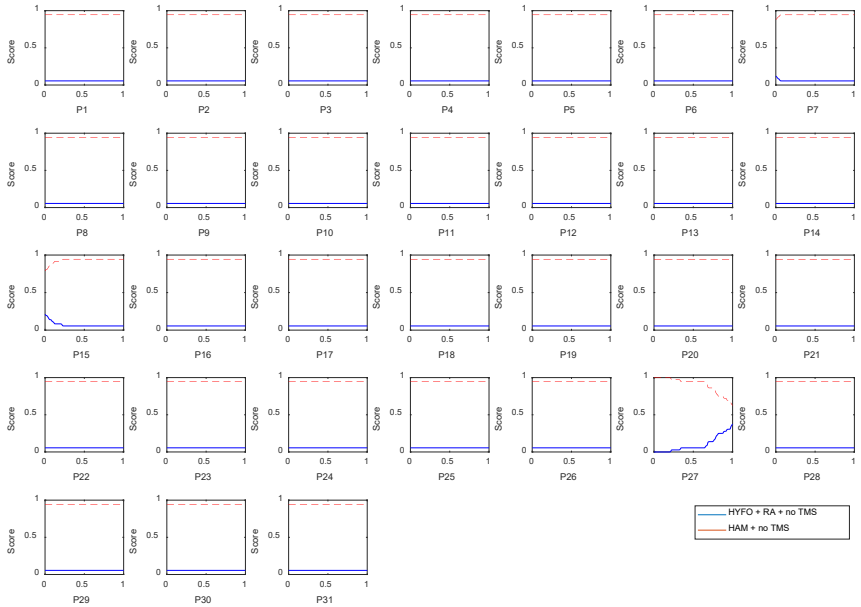


Figure 4: Partial dependence plots for the two discrete designs that were identified as optimal for at least one of the different scenarios considered.

parameter P27 (the activation energy of the side reaction to from n-decene to iso-decene) determines the decision the most. The same result was obtained by PFI where only changing this parameter shows an effect on the predicted class.

Hence, the optimal design of experiments was restricted to determining this parameter. The reaction temperature, the total pressure, the syngas ratio, the catalyst concentration and the sampling times were optimized for one batch experiment. The experiment was replaced by a simulation of the reaction with the nominal values of the parameters corrupted with added white noise with a standard deviation of 5 %. After one simulation experiment, the parameters were updated and the procedure was repeated. After seven iterations, one design could be identified as optimal for all scenarios as it can be seen in Figure 3 (right). In contrast, by a full factorial design with 32 additional experiments one cannot identify one design as optimal for all scenarios. Therefore, the proposed methodology shows an advantage over established techniques for model refinement.

4. Conclusion and Outlook

We presented an integrated methodology that accelerates the process development process by reducing the number of required experiments to find the optimum discrete design decisions. Using a superstructure optimization under uncertainties followed by a discrimination analysis, the parameters that influence the design decision the most can be identified. Hence, efficient experiments can be planned by optimal design of experiments that focus only on determining these parameters until the model is accurate enough to make a design decision. In the case study, we could identify one parameter as strongly influencing the design decision. Simulation studies showed that designing experiments for this parameter can reduce the number of experiments compared to a full-factorial design. In the future, we plan to expand the methodology to uncertain parameters in different process units.

Acknowledgement

Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) - TRR 63 "Integrierte chemische Prozesse in flüssigen Mehrphasensystemen" (Teilprojekt D1) - 56091768.
Funded by DFG in the context of the Transregio SFB InPROMPT.

References

Chen, Q. & Grossmann, I. E. (2017). Recent Developments and Challenges in Optimization-Based Process Synthesis. *Annual Review of Chemical and Biomolecular Engineering*, *8*(1), 249–283.

Fisher, A., Rudin, C. & Dominici, F. (2019). All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research*, *20*, 1–81.

Franceschini, G. & Macchietto, S. (2008). Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science*, *63*(19), 4846–4872.

Greenwell, B. M. (2017). pdp: An R package for constructing partial dependence plots. *R Journal*, *9*(1), 421–436.

Hentschel, B., Kiedorf, G., Gerlach, M., Hamel, C., Seidel-Morgenstern, A., Freund, H. & Sundmacher, K. (2015). Model-based identification and experimental validation of the optimal reaction route for the hydroformylation of 1-dodecene. *Industrial and Engineering Chemistry Research*, *54*(6), 1755–1765.

Kaiser, S. & Engell, S. (2020). Integrating Superstructure Optimization under Uncertainty and Optimal Experimental Design in early Stage Process Development. *Computer Aided Chemical Engineering*, *48*, 799–804.

Kaiser, S., Menzel, T. & Engell, S. (2021). Focusing experiments in the early phase process design by process optimization and global sensitivity analysis. *Computer Aided Chemical Engineering*, *50*, 899–904.

Kirschtowski, S., Jameel, F., Stein, M., Seidel-Morgenstern, A. & Hamel, C. (2021). Kinetics of the reductive amination of 1-undecanal in thermomorphic multicomponent system. *Chemical Engineering Science*, *230*, 116187.

Nentwich, C. & Engell, S. (2019). Surrogate modeling of phase equilibrium calculations using adaptive sampling. *Computers and Chemical Engineering*, *126*, 204–217.

Rinne, H. (2008). *Taschenbuch der Statistik* (4th ed.). Verlag Harri Deutsch.

Skiborowski, M., Wessel, J. & Marquardt, W. (2014). Efficient optimization-based design of membrane-assisted distillation processes. *Industrial and Engineering Chemistry Research*, *53*(40), 15698–15717.

Steimel, J. & Engell, S. (2016). Optimization-based support for process design under uncertainty: A case study. *AIChE Journal*, *62*(9), 3404–3419.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50014-2

Retrosynthesis Pathway Design Using Hybrid Reaction Templates and Group Contribution-Based Thermodynamic Models

Wang W., Liu Q., Zhang L.*, Dong Y., Du J.

Institute of Chemical Process Systems Engineering, School of Chemical Engineering, Dalian University of Technology, Dalian 116024, CHINA
keleiz@dlut.edu.cn

Abstract

Organic synthesis plays an essential role in the pharmaceutical industry. Traditionally, knowledge-based methods are used for the design of synthesis route, which is expensive and time-consuming and thus hinders the high-throughput design of the synthesis route. In this article, a retrosynthetic analysis framework is established using hybrid reaction templates and Group Contribution (GC)-based thermodynamic models. First, a hybrid database consisting of partial atom-mapping and full atom-mapping reaction templates is constructed utilizing well-studied organic reactions from literature. Second, numerous virtual reactions are generated from reaction templates with respect to target molecule, and reaction thermodynamic models based on the GC method are developed to validate the effectiveness of virtual reactions in a timely fashion. Finally, Breadth-First Search (BFS) algorithm is employed to search candidate retrosynthesis pathways which are thermodynamically feasible. In this procedure, five quantitative criteria are used to identify the top-ranked routes, including Fathead Minnow 96-hr LC_{50} ($LC_{50}FM$), flash point (Fp), Natural Product-likeness Score (NPScore), Synthesis Accessibility Score (SAScore), and Synthesis Complexity Score (SCScore). With the help of the developed framework, synthesis routes considering thermodynamic feasibility can be obtained. Two case studies involving Aspirin and Ibuprofen are presented to highlight the feasibility and reliability of the proposed framework.

Keywords: Retrosynthesis pathway design; Reaction template; Reaction equilibrium constant; Group contribution method; Breadth-First Search algorithm

1. Introduction

Organic synthesis is one of the most crucial components of the modern pharmaceutical industry. Traditionally, knowledge-based methods are applied to design synthesis routes for different pharmaceutical products. Nowdays, Computer-Aided Synthesis Design (CASD) techniques have enabled in-silico retrosynthesis and thus received considerable attention from chemists (Law et al., 2008; Szymkuc et al., 2016; Schwaller et al., 2020). Various searching algorithms have been successfully applied to optimize different objectives of synthesis route design. A crucial step in retrosynthetic analysis is to find a method that constructs explicit or implicit relations between reactants and products. Corey and Wipke are forerunners in retrosynthesis for their efforts in introducing the Logic and Heuristics Applied to Synthetic Analysis (LHASA) in the 1960s (Corey & Wipke, 1969). In their work, reaction templates (a certain type of sub-molecular pattern that maps atom connectivity) were used as a bridge linking up reactants with products.

Route Designer (Law et al., 2008) employed extended reaction templates to ensure the validity of synthesis routes. Syntaurus (Szymkuc et al., 2016) used more than 20,000 explicitly defined reaction rules to avoid obtaining absurd synthesis routes. The developers of Chematica (Szymkuc et al., 2016) have manually encoded more than 10,000,000 molecules and reactions to form synthesis networks. Among these methods, the reaction template is more favorable for its concise form in representing chemical reactions. Reaction templates are manually encoded by chemists who are experienced in organic synthesis. Gradually, chemists pay more and more attention to algorithms that automatically extract features as well as generate templates from the known reactions in commercial databases owned by pharmaceutical companies and non-commercial databases like USPTO (Lowe, 2014) and Reaxys (Goodman, 2009). So far, extraction algorithms have made significant progress in theory and practice. Law et al. (2008) focusing on extending the reaction cores to necessary chemical environments. Coley et al. (2017) used a heuristics-driven algorithm to extract reaction templates from the USPTO database. Reaction templates cannot work alone in retrosynthesis. Specialized algorithms written by expert chemists are used to cooperate with reaction templates. As the Artificial Intelligence (AI) develops, many researchers have found that Machine Learning (ML) can solve retrosynthesis when reaction templates are applied. Segler and Waller (2017) proposed a model for retrosynthesis using neural-symbolic ML and 103 hand-coded reaction templates, while Coley et al. (2017) applied ML and rigid reaction templates for the reversed problem. Template-free method is developing rapidly thanks to the new advancement in NLP (Natural Language Process) technology. Reactions written in SMILES (Simplified Molecular-Input Line-Entry System) (Weininger, 1998) notations are used to train RNN (Recurrent Neural Network) or Transformer model. The well-known template-free architecture is Molecular Transformer (Schwaller et al., 2019), which reads the mixed (or separated) strings of reactants, solvents, catalysts, and reagents as the inputs to predict possible product strings.

Although retrosynthesis analysis has been studied for several decades, there are remaining problems unsolved. Increasing the depth of a neural network or applying newly raised neural network architecture may allow us to get more satisfying results for prediction, but the relation between input and output becomes hard to be understood. Other aspects, such as process safety, environmental friendliness of reagents, and the price of raw materials, should also be considered during the process of synthesis route design. This paper presents a retrosynthetic analysis framework using hybrid reaction templates and GC-based thermodynamic models. The curated reaction templates are manually encoded according to available literature (Smith & March, 2001) to ensure the validity. In Section 2, the proposed three-steps framework is discussed in detail. In Section 3, two case studies are presented to highlight the feasibility and reliability of the proposed framework.

2. Retrosynthetic analysis framework using hybrid reaction templates and GC-based thermodynamic models

The proposed retrosynthetic analysis framework is divided into three parts: (1) Generate virtual routes; (2) Verify virtual routes; (3) Rank valid routes, as shown in Figure 1.

2.1. Step 1: Generate virtual routes

Reaction template is a sub-molecular pattern that maps atom connectivity. SMARTS strings are used to encode reaction templates in this framework since they are fully

supported in RDKit (Landrum, 2016). Chemical reactions written in SMARTS strings are reversible by simply switching reactants and products. SMARTS strings are classified into two categories according to the completeness, called partial atom-mapping and full atom-mapping. Partial atom-mapping SMARTS strings have asymmetric numbers labeling, while full atom-mapping SMARTS strings have symmetric numbers labeling. The most of the SMARTS strings in the well-known USPTO 1976-2016 database (Lowe, 2014) are partial atom mapping and thus hardly useful for our framework. As a result, the reaction template database is established manually according to available literature (Smith & March, 2001). In our reaction template database, 347 different reaction templates covering the most used reactions are established. These reaction templates are classified into nine categories. Some partial atom-mapping reaction templates (9.7% of total reaction templates) in our reaction template database are kept maintaining the diversity of the reaction template database.

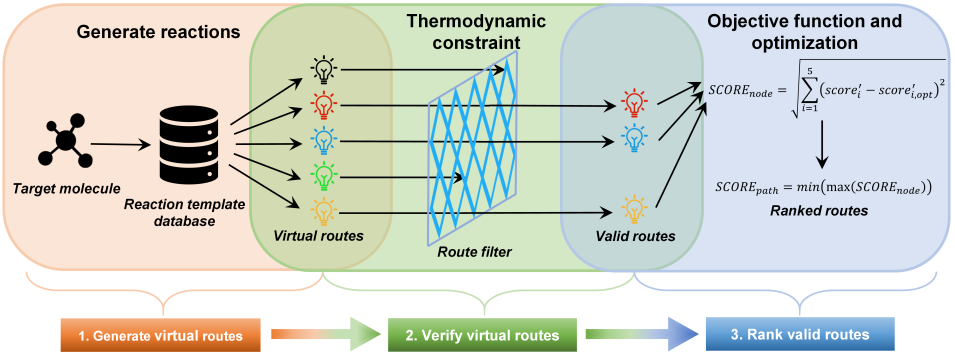


Figure 1. A three-step retrosynthesis analysis framework using hybrid reaction templates and GC-based thermodynamic models

2.2. Step 2: Verify virtual routes

Chemical equilibrium theory provides a convenient way to evaluate whether a reaction is able to occur or not under a given temperature. For isothermal and isobaric reaction, the reaction equilibrium constant is correlated with the change of standard molar Gibbs free energy of the reaction $\Delta_r G_m^\theta$, ideal gas constant R and reaction temperature T, as shown in Eq.(1).

$$K = \exp\left(-\Delta_r G_m^\theta / RT\right) \tag{1}$$

A more flexible formula for calculating reaction equilibrium constant using standard molar Gibbs free energy is expressed as Eq.(2), which is derived under a rational assumption as per the textbook "Principles of Modern Chemistry (7th Ed)" (Oxtoby et al., 2011).

$$\Delta_r G_m^\theta(T) = \sum_j \upsilon_j \Delta_f H_{m,j}^\theta(298.15\ \mathrm{K}) - T\sum_j \upsilon_j S_{m,j}^\theta(298.15\ \mathrm{K}) \tag{2}$$

Here, $\Delta_f H_{m,j}^\theta$ and $S_{m,j}^\theta$ represent the standard molar enthalpy of formation and the standard molar entropy for compound j, respectively; υ_j is the stoichiometric coefficient of compound j. Existing databases such as Lange's Handbook of Chemistry (Speight, 2005) contain thermodynamic parameters at 298.15 K for most common molecules. However, molecules involved in retrosynthesis pathway design are usually

intermediates whose thermodynamic parameters are not readily available from the databases. Therefore, GC methods are introduced here as an alternative way to calculate relevant thermodynamic parameters and other properties that are involved in retrosynthesis pathway design. The relevant data used in GC methods come from our previous work in solvent design (Liu et al., 2019). BFS algorithm is widely used in solving problems like shortest path problems and minimum steps problems. Pre-set reaction equilibrium constant K_0 is used as a criterion to keep the BFS algorithm focusing on the most promising reaction routes. An online database of market-buyable molecule is introduced to improve computational effectiveness and accelerate convergence.

Step 3: Rank valid routes

In this step, a quantitative evaluation system is developed to rank routes that pass the thermodynamic verification objectively. The following criteria are considered: (1) Fathead Minnow 96-hr LC_{50} ($LC_{50}FM$); (2) flash point (Fp); (3) Natural Product-likeness Score (NPScore) (Ertl et al., 2008); (4) Synthesis Accessibility Score (SAScore) (Ertl et al., 2009); (5) Synthesis Complexity Score (SCScore) (Coley et al., 2018). A normalization is applied for each criterion to ensure they are normalized to a fixed range between 0 to 1. Euclidean distance is used as a quantitative method to calculate the distance between a specific molecule and the target (optimal) molecule in chemical space as shown in Eq.(3).

$$SCORE_{node} = \sqrt{\sum_{i=1}^{5}\left(score_i' - score_{i,opt}'\right)^2} \tag{3}$$

In a multi-step synthetic problem, the node with the highest value of $SCORE_{node}$ is deemed to be the synthesis-determining step. After finding all the synthesis-determining steps of corresponding routes, their scores are ranked in ascending order to find out the optimal synthesis route as shown in Eq.(4).

$$SCORE_{path} = \min\left(\max(SCORE_{node})\right) \tag{4}$$

The $SCORE_{path}$ in Eq.(4) is defined as the score of a full synthesis route, representing the synthetic features of the synthesis route. If any additional criterion needs to be considered in the future, Eq.(3) is extensible while Eq.(4) remains valid.

3. Case studies

3.1. Synthesis route design for Aspirin

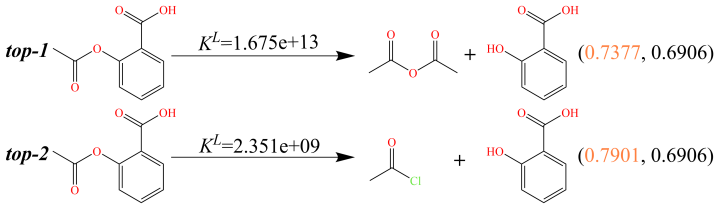


Figure 2. The results of synthesis route design for Aspirin

The SMILES of Aspirin is required, which is CC(=O)OC1=CC=CC=C1C(=O)O. Here, K_0 was set to 100 and search depth was set to 2. The reaction temperature was set to

298.15 K in order to search reactions that are feasible at room temperature. The results of synthesis route design for Aspirin are shown in Fig. 2.

According to Eq. (4), the optimal synthesis route for Aspirin is to use acetic anhydride reacting with salicylic acid, which is consistent with the industry practice. Acetyl chloride gets a higher overall score for its low flash point and thus ranked second.

3.2. Synthesis route design for Ibuprofen

The SMILES of Ibuprofen is required, which is CC(C)CC1=CC=C(C=C1)C(C)C(=O)O. All synthesis constraints were the same as Aspirin except for the search depth which was set to 3. The design results for Ibuprofen are shown in Fig. 3.

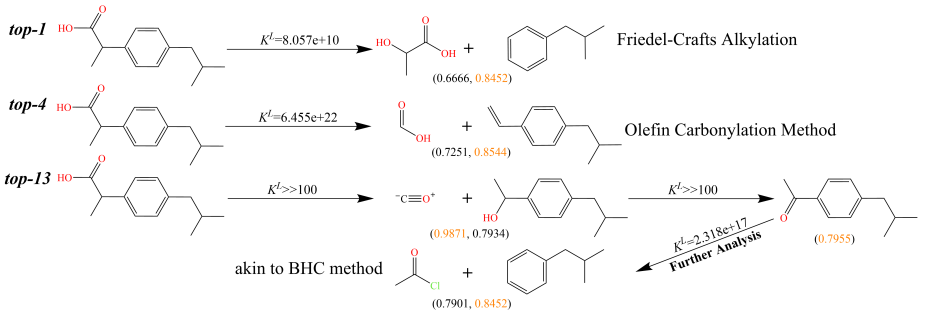


Figure 3. The results of synthesis route design for Ibuprofen

The top-1 route is Friedel-Crafts alkylation which tends to produce multi-substituted products. The catalysts with high shape-selectivity are needed to make this route practical. The well-known olefin carbonylation method is found and ranked 4 while a method alike to BHC method is found and ranked 13.

4. Conclusions

In this paper, a retrosynthetic analysis framework using hybrid reaction templates and GC-based thermodynamic models is proposed. First, a hybrid reaction template database is used to generate various possible routes. Then, reaction equilibrium constant is used to verify the thermodynamic tendency of virtual routes while BFS algorithm is applied to the searching process. Finally, model criteria are proposed as a quantitative and efficient method to evaluate different synthesis routes. The constructed hybrid reaction templates database is reliable and can be updated manually. During the process of ranking synthesis routes, the concept of "synthesis-determining step" is introduced and integrated with the ranking system, and more attentions ought to be paid to the synthesis-determining steps. The synthesis route design results for Aspirin and Ibuprofen are satisfactory and thus highlight the feasibility and effectiveness of the proposed framework. The limitations are clear due to the using of thermodynamic models which only provide the tendency of reaction. Reaction kinetics should also be considered and work together with the thermodynamics. Considering reaction kinetics requires the knowledge of precise kinetic equations which could be a major obstacle for developing reaction kinetics-based models in retrosynthesis. In conclusion, the proposed framework provides a new solution for a rational retrosynthesis by utilizing reaction thermodynamics.

References

Coley, C. W., Barzilay, R., Jaakkola, T. S., Green, W. H., Jensen, K. F., 2017. Prediction of Organic Reaction Outcomes Using Machine Learning. ACS Cent. Sci. 3, 434-443.

Coley, C. W., Rogers, L., Green, W. H., Jensen, K. F., 2018. SCScore: Synthetic Complexity Learned from a Reaction Corpus. J. Chem. Inf. Model. 58, 252-261.

Corey, E. J., Wipke, W. T., 1969. Computer-Assisted Design of Complex Organic Syntheses. Science. 166, 178-192.

Ertl, P., Roggo, S., Schuffenhauer, A., 2008. Natural Product-likeness Score and Its Application for Prioritization of Compound Libraries. J. Chem. Inf. Model. 48 (1), 68-74.

Ertl, P., Schuffenhauer, A., 2009. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. J. Cheminform. 1, 8.

Goodman, J. 2009. Computer Software Review: Reaxys. J. Chem. Inf. Model. 49 (12), 2897-2898.

Landrum, G., 2016. RDKit: Open-source cheminformatics; http://rdkit.org, (accessed March 15, 2020).

Law, J., Zsoldos, Z., Simon, A., Reid, D., Liu, Y., Khew, S. Y., Johnson, A. P., Major, S., Wade, R. A., Ando, H. Y., 2009. Route Designer: A Retrosynthetic Analysis Tool Utilizing Automated Retrosynthetic Rule Generation. J. Chem. Inf. Model. 49, 593-602.

Lowe, D. M., 2014. Patent reaction extraction: downloads. https://bitbucket.org/dan2097/patent-reaction-extraction/downloads, (accessed November 6, 2019).

Liu, Q., Zhang, L., Tang, K., Feng, Y., Zhang, J., Zhuang, Y., Liu, L., Du, J., 2019. Computer-aided reaction solvent design considering inertness using group contribution-based reaction thermodynamic model. Chem. Eng. Res. Des. 152, 123-133.

Oxtoby, D. W., Gillis, H. P., Campion, A., Helal, H. H., Gaither, K. P.,2011. Principles of Modern Chemistry, 7th Ed. CENGAGE Learning, Belmont.

Schwaller, P., Laino, T., Gaudin, T., Bolgar, P., Bekas, C., Lee, A. A., 2019. Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction. ACS Cent. Sci. 5 (9), 1572-1583.

Schwaller, P., Petraglia, R., Zullo, V., Nair, V. H., Haeuselmann, R. A., Pisoni, R., Bekas, C., Luliano, A., Laino, T., 2020. Predicting retrosynthetic pathways using a combined linguistic model and hyper-graph exploration strategy. Chem. Sci. 11, 3316-3325.

Segler, M. H. S., Waller, M. P., 2017. Neural-Symbolic Machine Learning for Retrosynthesis and Reaction Prediction. Chem. Eur. 23, 5966-5971.

Smith, M. B., March, J., 2001. March's Advanced Organic Chemistry, 5th Ed. John Wiley & Sons, New Jersey.

Speight, J. G., 2005. Lange's Handbook of Chemistry, 16th Ed. McGraw-Hill, New York.

Szymkuc, S., Gajewska, E. P., Klucznik, T., Molga, K., Dittwald, P., Startek, M., Bajczyk, M., Grzybowski, B. A., 2016. Computer-Assisted Synthetic Planning: The End of the Beginning. Angew. Chem. Int. Ed. 55, 5904-5937.

Weininger, D., 1998. SMILES, a Chemical Language and Information System 1. Introduction and Encoding Rules. J. Chem. Inf. Comput. Sci. 28, 31-36.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50015-4

Optimization-based Design of Product Families with Common Components

Chen Zhang[a], Clas Jacobson[a], Qi Zhang[b], Lorenz T. Biegler[c], John C. Eslick[d,e], Miguel A. Zamarripa[d,e], David Miller[d], Georgia Stinchfield[c], John D. Siirola[f], Carl D. Laird*[c]

[a]*Carrier Global Corporation, Palm Beach Gardens, Florida 33418*
[b]*Dept. of Chem. Eng. and Materials Science, U. of Minnesota, Minneapolis, MN 55455*
[c]*Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, United States*
[d]*National Energy Technology Laboratory, Pittsburgh, Pennsylvania, USA*
[e]*NETL Support Contractor, Pittsburgh, Pennsylvania, USA*
[f]*Sandia National Laboratories, P.O. Box 5800, Albuquerque, NM 87123, United States*

**claird@andrew.cmu.edu*

Abstract

For many industries addressing varied customer needs means producing a family of products that satisfy a range of design requirements. Manufacturers seek to design this family of products while exploiting opportunities for shared components to reduce manufacturing cost and complexity. We present a mixed-integer programming formulation that determines the optimal design for each product, the number and design of shared components, and the allocation of those shared components across the products in the family. This formulation and workflow for product family design has created significant business impact on the industrial design of product families for large-scale commercial HVAC chillers in Carrier Global Corporation. We demonstrate the approach on an open case study based on a transcritical CO_2 refrigeration cycle. This case study and our industrial experience show that the formulation is computationally tractable and can significantly reduce engineering time by replacing the manual design process with an automated approach.

Keywords: product family design, discrete optimization, product manufacturing

1. Introduction

For many industries, addressing global markets and varied customer needs means producing a family of products that are able to satisfy a range of design requirements. For example, commercial chiller systems for HVAC sold in different regions of the world are subject to different operating and boundary conditions, customer cost and performance expectations, and efficiency regulations. This requires the design and manufacturing of a family of products to meet requirements of different geographical regions and customer needs. Optimizing each of the products independently results in significantly increased manufacturing cost and complexity since each design will include unique sizing for all of the sub-components, ignoring the potential for sharing these components across multiple products within the family. Therefore, manufacturers seek to design the entire family of products simultaneously, determining the optimal design for each product, the designs of

common components, and the assignment of these components to each of the products in the family. This is a highly-combinatorial problem, that is typically performed with heuristics and ad-hoc approaches, takes significant engineering time, and results in sub-optimal designs. Many industries need effective design of product families that can exploit shared components, and this is an active area of research in manufacturing where various heuristics and optimization strategies have been applied (Simpson et al. 2014). Some examples of optimization-based approaches have focused on definition and optimization of a commonality index or degree of commonality (Thonemann & Brandeau 2000) and application of genetic algorithms (Liu et al. 2011). Integer programming techniques have also been used in, for example, the integration of the supply chain with the product family design (Baud-Lavigne et al. 2016). These concepts have applicability to chemical process design. In particular, for decentralized applications where many instances of similar processes with different performance specifications are required, the benefits of well-designed product families allow for significant reduction in engineering and construction costs.

In this paper, we present an mixed-integer programming formulation for product family design with common sub-components developed in collaboration with researchers at Carrier Global Corporation. Instead of manufacturing uniquely specified (e.g., sized) components for each product, we seek to manufacture a small number of component designs and share these across multiple products. The formulation determines the cost optimal designs for each of the products, the optimal sizing for the shared components, and the allocation of these components for each of the products. This formulation and workflow for product family design has created significant business impact on the industrial design of product families for large-scale commercial HVAC chillers in Carrier Global Corporation. In one application, the product family design workflow selected common compressors for a global family of over 200 products, leading to significant direct cost savings (material and labor), indirect cost savings (prototype design, build, and test), and an order of magnitude reduction in R&D time associated with this task. This process is being used and extended within Carrier across several product lines.

We demonstrate the product family design formulation on an open case study considering a family of HVAC products based on a CO_2 refrigeration cycle described in Li & Groll (2005). The model for the system is built using the IDAES process modeling platform (Lee et al. 2021) and the product family design problem is implemented in Pyomo (Bynum et al. 2021). The approach is shown to be computationally tractable for real-world systems, with significantly reduced engineering time, replacing the manual design process with an automated, optimization-based approach.

2. Product Family Design Formulation

We assume that the set of products P and their performance requirements have already been specified (e.g., from market analysis). Product requirements may be captured as boundary conditions that must be matched exactly or as inequalities that provide bounds on the product performance. The set of components where there is opportunity for utilizing shared designs across multiple products is given by C, and the set of candidate designs for each component c is given by S_c. Our goal is to optimally design all of the products $p \in P$ while reducing the overall manufacturing costs by utilizing a (hopefully small) subset of the candidate component designs in these products.

For each product p we consider a set of design alternatives. For each alternative, we specify which candidate component designs are to be utilized in the product. For the initial set of design alternatives, we typically consider all combinations of candidate designs for each of the components (i.e., the Cartesian product of all S_c for all $c \in C$). Then, for each of these alternatives, we can perform simulations (or optimizations) and identify the alternatives that meet the required performance specifications. We define this set of all feasible alternatives for product p as A_p. The set Q_a is a tuple set that captures the specific candidate component designs used within each alternative a.

The proposed formulation for optimal design of product families with common components is shown in Equations (1-6). The binary variables z_{cs} identify which candidate designs s are selected for each component c, and x_{pa} captures which alternative is selected for product p. Equation (1) is the objective function, and the first term captures the expected cost associated with the family design where w_p is the expected sales (or sales fraction) for each product, and α_{pa} is the annualized cost if alternative a is selected for product p. The second term captures the cost required to develop the manufacturing process for each unique component selected. In many industrial examples, the cost of this manufacturing complexity is difficult to capture, and we can also constrain the number of candidate component designs selected with Equation (2).

$$\min_{x,z} \sum_{p \in P} w_p \sum_{a \in A_p} \alpha_{pa} x_{pa} + \sum_{c \in C} \sum_{s \in S_c} \beta_{cs} z_{cs} \tag{1}$$

s.t.

$$\sum_{s \in S_c} z_{cs} \leq N_c \qquad \forall\, c \in C \tag{2}$$

$$\sum_{a \in A_p} x_{pa} = 1 \qquad \forall\, p \in P \tag{3}$$

$$x_{pa} \leq z_{cs} \qquad \forall\, p \in P, a \in A_p, (c,s) \in Q_a \tag{4}$$

$$0 \leq x_{pa} \leq 1 \qquad \forall\, p \in P, a \in A_p \tag{5}$$

$$z_{cs} \in \{0,1\} \qquad \forall\, c \in C, s \in S_c. \tag{6}$$

Equation (3) ensures that only one alternative is selected for each product, and Equation (4) allows alternative a for product p only if the required components have been selected.

3. Process Case Study

For our case study, we consider the design of a family of products for commercial HVAC applications based on the transcritical CO_2 refrigeration cycle described in Li & Groll (2005). The process flow diagram is shown in Figure 1. We developed an IDAES model for this process using the standard unit model library with the exception of the ejector which required a custom model. The compressor model includes an efficiency curve to capture the drop in efficiency when it is operating away from the design flowrate. IDAES also includes a costing framework that was used to capture the equipment capital costs.

We consider two performance criteria when specifying the products P. The cooling capacity is the primary criterion determining the size of the components in the refrigeration

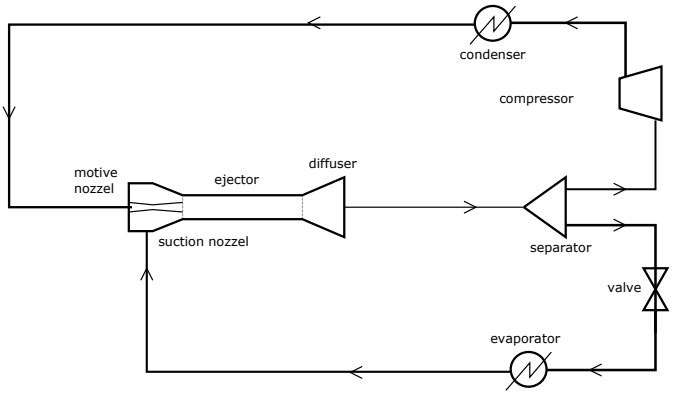


Figure 1: Process flowsheet for CO2 refrigeration cycle. This model was based on Li & Groll (2005)

.

cycle and can vary significantly based on customer needs. The outside air temperature varies significantly by region, and different units are designed for different conditions.

Here, we consider capacities of CAP=$\{80, 100, 120, 140, 160, 180, 200\}$ tons of refrigeration and outside air temperature specifications of OAT=$\{28, 29, 30, 31, 32, 33, 34, 35\}$ degrees Celsius. With these specifications, we have a total of 56 different products to consider, identified as the Cartesian product of all values in CAP and OAT.

The opportunities we consider for shared components across the products include the evaporator, the condenser, and the compressor, defining C=$\{\text{Evap}, \text{Cond}, \text{Compr}\}$. We consider five sizes of evaporator labeled A through E in order of increasing size, seven sizes of condenser labeled A through G in order of increasing size, and four sizes for the compressor, labeled A through D, also in order of increasing size. This gives us a total of 140 alternatives to consider for each product defined by the Cartesian product of the different candidate components specified as, $S_{\text{Evap}} = \{\text{A}, \text{B}, \text{C}, \text{D}, \text{E}\}$, $S_{\text{Cond}} = \{\text{A}, \text{B}, \text{C}, \text{D}, \text{E}, \text{F}, \text{G}\}$, and $S_{\text{Compr}} = \{\text{A}, \text{B}, \text{C}, \text{D}\}$.

We performed simulations for each of these alternatives across all the products (with CAP and OAT specified as boundary conditions) for a total of 7840 simulations. Of these, 3708 were infeasible and not able to meet the desired performance specifications. The feasible alternatives were used to define the remaining data required in the optimization formulation along with recorded capital and operating costs from the IDAES model.

The product family design problem (1-6) was formulated in Pyomo (Bynum et al. 2021) and solved using Gurobi (Gurobi Optimization, LLC 2021). We set the maximum number of candidate components to 2 for each of the evaporator, condenser, and compressor. Gurobi was able to solve this problem in under one second. Results showing the optimal designs are illustrated in Figure 2. The figure on the left shows the solution considering capital cost only (materials and construction). In this case, the optimization selected evaporators C and D, condensers A and B, and compressors A and B for manufacturing. The colors on the figure show unique designs, and the legend on the right indicates which selected components were matched with each design. The optimization selected a larger

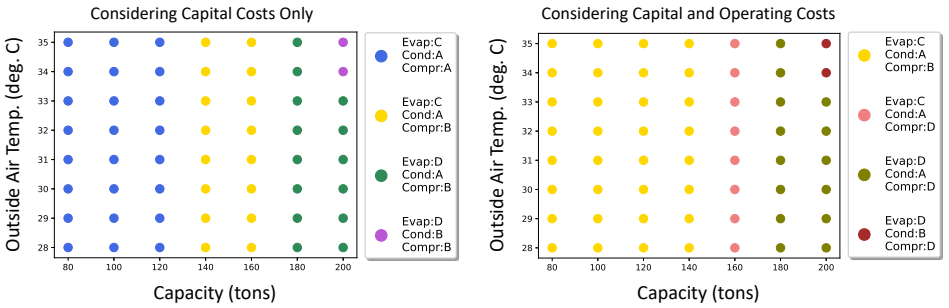


Figure 2: Optimal design of the product family with capital costs only on the left and capital plus operating costs on the right.

compressor when moving from 120 to 140 tons of capacity, and a larger evaporator when moving from 160 to 180 tons. As well, we see a change to a larger condenser for the higher outside air temperatures at the largest capacity. In this case, since we considered capital costs only, the optimization has selected the smallest compressors that are able to guarantee feasibility across the products. However, for most of these products, these compressors are operating off of their design flowrates and not achieving peak efficiency.

The figure on the right shows the optimal product family design considering both capital and operating costs. Here, we notice that the optimization did not select the smallest compressors, but has selected larger compressors so that they are operating closer to their design flowrate for improved operating efficiency.

4. Conclusion

In this paper, we have presented a formulation for optimal product family design. This formulation determines optimal designs across a set of products from a number of defined alternatives while reducing manufacturing costs by exploiting the opportunity for shared components across multiple products. This approach has been used industrially at Carrier Global Corporation with significant reduction in both costs and engineering time. The approach is also easily extended to support optimization of non-shared components by replacing the simulations with optimization problems for each of the alternatives considered.

This formulation can be efficiently solved for large data sets with commercial mixed-integer linear programming solvers. The computational bottleneck is the large number of simulations or optimizations that are required to gather the input data. It can be beneficial to use engineering knowledge to reduce the set of alternatives, and consequently, the total number of simulations that need to be performed.

It is important to note that there are a number of chemical process applications that can benefit from distributed operation of smaller, intensified, modular processes (Baldea et al. 2017). Any application that requires a large number of similar processes with variation in specific process requirements is an excellent candidate for the approaches outlined in this paper. This includes, for example, applications in water treatment, carbon capture from smaller localized sources, direct air capture, and other environmental processes. The

concepts of product family design can be utilized to shift from one-off unique designs for each application to the definition of a suite of products that span the design space while reducing manufacturing costs with shared components.

References

Baldea, M., Edgar, T. F., Stanley, B. L. & Kiss, A. A. (2017), 'Modular manufacturing processes: Status, challenges, and opportunities', *AIChE journal* **63**(10), 4262–4272.

Baud-Lavigne, B., Agard, B. & Penz, B. (2016), 'Simultaneous product family and supply chain design: An optimization approach', *International Journal of production economics* **174**, 111–118.

Bynum, M. L., Hackebeil, G. A., Hart, W. E., Laird, C. D., Nicholson, B. L., Siirola, J. D., Watson, J.-P. & Woodruff, D. L. (2021), *Pyomo—Optimization Modeling in Python*, Vol. 67, Springer Nature.

Gurobi Optimization, LLC (2021), 'Gurobi Optimizer Reference Manual'.
URL: *https://www.gurobi.com*

Lee, A., Ghouse, J. H., Eslick, J. C., Laird, C. D., Siirola, J. D., Zamarripa, M. A., Gunter, D., Shinn, J. H., Dowling, A. W., Bhattacharyya, D. et al. (2021), 'The idaes process modeling framework and model library—flexibility for process simulation and optimization', *Journal of Advanced Manufacturing and Processing* p. e10095.

Li, D. & Groll, E. A. (2005), 'Transcritical co2 refrigeration cycle with ejector-expansion device', *International Journal of refrigeration* **28**(5), 766–773.

Liu, Z., San Wong, Y. & Lee, K. S. (2011), 'A manufacturing-oriented approach for multiplatforming product family design with modified genetic algorithm', *Journal of Intelligent Manufacturing* **22**(6), 891–907.

Simpson, T. W., Jiao, J., Siddique, Z. & Hölttä-Otto, K. (2014), 'Advances in product family and product platform design', *New YorN: Springer* .

Thonemann, U. W. & Brandeau, M. L. (2000), 'Optimal commonality in component design', *Operations Research* **48**(1), 1–19.

Acknowledgements and Disclaimer: Neither the United States Government nor any agency thereof, nor any of their employees, nor the support contractor, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. The authors acknowledge contributions of Dr. Rui Huang who assisted with the problem formulation and workflow definition.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50016-6

Economic evaluation and analysis of electricity and nano-porous silica productions from rice husk

Semie kim[a], and Young-Il Lim [a*]

[a]*CoSPE, Department of Chemical Engineering, Hankyong National University, Gyeonggi-do, Anseong-si, Jungang-ro 327, 17579 Republic of Korea*
limyi@hknu.ac.kr

Abstract

Techno-economic analysis (TEA) was performed for the production of electricity or nano-porous silica (NPS) from 50 t/d of rice husk (RH). The process for electricity generation from RH had a net electricity efficiency of 15%. Using the same amount of RH (approximately 2 t/h), 278 kg/h of NPS was produced. The electricity production process was not profitable because of negative return on investment (ROI). In the case of producing NPS from RH, the total investment cost (TCI) and total production cost (TPC) were 13.7 M$ and 1.5 M$/y, respectively. The ROI and payback period (PBP) were predicted to be 3.7%/y and 17.5 y, respectively.

Keywords: Rice husk; Rice husk ash (RHA); Nano-porous silica (NPS); Electricity production; Techno-economic analysis (TEA)

1. Introduction

About 782 million tonnes of paddy rice were produced in the world in 2020. Asia accounts for 90% (705 million tonnes) of global production (FAO, 2020). During the milling process of paddy rice, rice husk is produced as an agricultural by-product (Peerapong and Limmeechokchai, 2009). 1,000 kg of paddy rice produces 220 kg of rice husk (Mor et al., 2017). 1,000 kg of rice husk (RH) produces 480~1,000 kW_e of electricity (Steven et al., 2021). Along with electricity generation, about 18~20 % of rice husk remains ash (Subbukrishna et al., 2007). The ash is mainly composed of silica (>90%) (Liu et al., 2011; Nayak et al., 2019). Silica with the purity of 95 wt% is used in industrial fields such as reinforced rubber additives (tire), materials (zeolite and polymer), concrete, and semi-conductor (Prasara-A and Gheewala, 2017; Steven et al., 2021). In general, Silica is produced from sand and sodium carbonate at 1400-1500 °C (Munasir and Triwikantoro, 2013). Sodium silicate (SS) is produced from RH ash burned at 650~850 °C (Kim and Kim, 2020). The nano-porous silica (NPS) is finally synthesized from a polymerization of SS in H_2SO_4 solution (Pode, 2016). Therefore, eco-friendly electricity and nano-porous silica (NPS) can be produced using rice husk, which is considered as a carbon-neutral biomass (Bergqvist et al., 2008; Pode, 2016).

In this study, two process flow diagrams (PFDs) were constructed for electricity and nano-porous silica (NPS) productions from rice husk using a commercial process simulator (ASPEN Plus, ASPEN Tech, USA). Based on the PFDs, the economic feasibility of the electricity and NPS production processes was compared using an equal amount of rice husk.

2. Process description

The rice husk (RH) used in this study is a by-product of paddy rice produced in Vietnam. Table 1 shows the proximate and ultimate analyses of RH containing 41.0 wt% carbon, 5.5 wt% hydrogen, 34.9 wt% oxygen, 0.7 wt% nitrogen, 0.1 wt% sulfur, and 17.8 wt% ash. The ash includes 94.20 wt% SiO_2, 0.75 wt% P_2O_5, 2.88 wt% K_2O, 0.97 wt% CaO, and 1.20 wt% others.

Table 1. Proximate and ultimate analyses of rice husk (RH) in this study.

Proximate analysis (wt%)		Ultimate analysis (wt%, dry basis)		Ash	
Moisture	10.00	C	41.0	SiO_2	94.20
Volatile matter	68.30	H	5.5	P_2O_5	0.75
Fixed carbon	13.68	O	34.9	K_2O	2.88
Ash	16.02	N	0.7	CaO	0.97
		S	0.1	others	1.20
		Ash	17.8		
Total	100.00		100.0		100.00
HHV (MJ/kg)	14.80				

To compare the economic values of the electricity or NPS production process, the same amount of RH (50 t/d) was used as a raw material. Two processes using RH were considered: Case 1 (electricity), and Case 2 (NPS).

Case 1 is the electricity production power plant from RH, as shown in Figure 1. The RH is burned with 50% excess air at 750 °C (see Eq. (1)).

$$Rice\ husk + O_2 \rightarrow H_2O + CO_2 + SiO_2 + others \quad (1)$$

The steam turbine generates electricity at 400 °C and 25 bar using the heat of combustion. The cyclone to remove fly-ash, bag-filter and bird blue scrubber to remove fine dust were used.

In Case 2, the NPS is produced using RHA (see Figure 2). RHA reacts with sodium carbonate to produce sodium silica (SS) in the hydrothermal synthesis reactor as shown in Eq. (2).

$$Na_2CO_3 + SiO_2 \rightarrow Na_2O \cdot 3.4SiO_2 + CO_2 \quad (2)$$

The activated carbon (AC), which is an unconverted carbon, is separated in the ultrafiltration. SS reacts with sulfuric acid to produce NPS (see Eq. (3)). Finally, NPS containing 6% moisture is produced through filtering and drying.

$$Na_2O \cdot 3.4SiO_2 + H_2SO_4 \rightarrow Na_2SO_4 + SiO_2 + H_2O \quad (3)$$

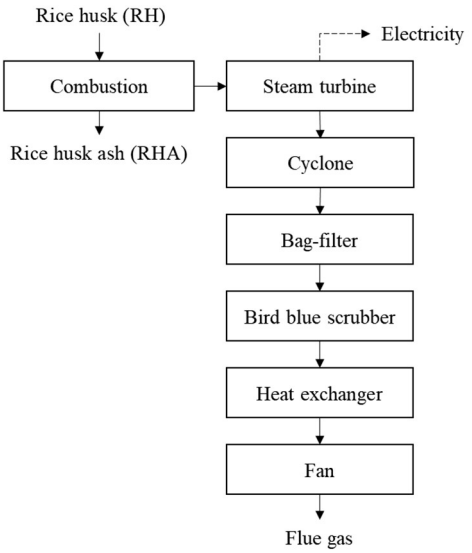


Figure 1. Block flow diagram (BFD) of electricity production from rice husk

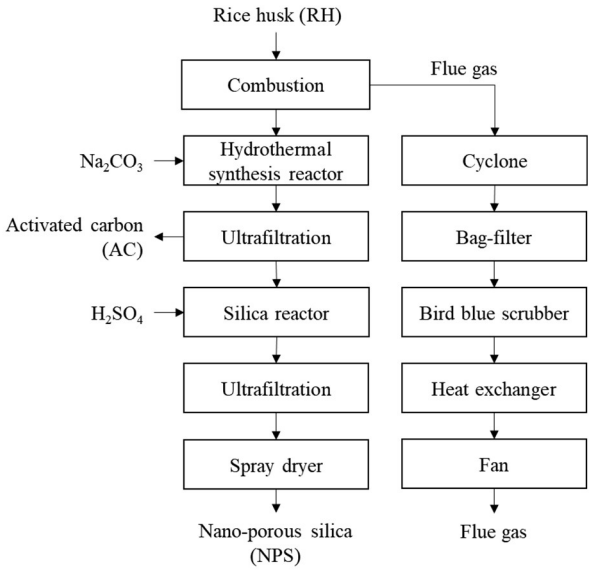


Figure 2. Block flow diagram (BFD) of NPS production from rice husk

3. Methodology of techno-economic analysis

The techno-economic analysis (TEA) is a method for determining the economic feasibility of a process (Do et al., 2014; Lim et al., 2016). The TEA used in this study

investigates the technical feasibility of the process by calculating mass and energy balance using thermodynamic models and equations of state. In addition, the economic values such as the total investment cost (TCI), total production cost (TPC), return on investment (ROI), and payback period (PBP) are examined in the TEA (Kim and Lim, 2021; Kim et al., 2021).

Table 2 shows the assumptions used in economic evaluation. The prices of raw electricity and cooling water are the same as Vu et al. (Vu et al., 2020). For the raw material price, the 2019 market price was applied. The prices of RH, AC and NPS were assumed to be 10, 1,000, and 1,500 $/t, respectively.

Table 2. Economic assumptions for electricity or NPS plants from rice husk.

Parameter		Assumption	Unit
Debt ratio (λ)		0.7	
Plant availability		8000	h/y
Startup time (50% plant performance)		4	month
Plant lifetime (L_p)		20	y
Inflation rate (α)		2	%/y
Corporation tax rate (β)		20	%/y
Interest rate (γ)		6	%/y
Raw material and product price	Rice husk	10	$/t
	Na_2CO_3	200	$/t
	H_2SO_4	143	$/t
	Activated carbon	1,000	$/t
	Nano-porous silica	1,500	$/t
Utility price	Electricity	0.098	$/kWh
	Liquefied natural gas	0.5	$/kg
	Cooling water	0.273	$/m^3

4. Results

The performance and economic feasibility of the electricity or NPS production process from RH were compared.

4.1. Process performance

The process simulation was performed for the two processes to produce electricity and NPS, respectively. Table 3 shows the process performance for the two processes. In Case 1, electricity was 1,343 kW_e from 2,083 kg/h (50 t/d) rice husk. The 90 kW_e was consumed for steam turbine, flue gas treatment, and air compression. The net electricity was 1,254 kW_e, which was a net electricity efficiency of 15% . Case 2 produced 278 kg/hr NPS and 40 kg/hr AC with the same amount of rice husk as Case 1. An electricity of 90.5 kW_e was consumed in NPS production.

Table 3. Process performance for electricity or NPS plants from rice husk.

	Case 1	Case 2	Unit
Feed (Rice husk) flow rate	2,083.3	2,083.3	kg/h
Product			
Nano-porous silica	-	278.1	kg/h
Activated carbon	-	39.5	kg/h
Electricity	1,343.2	-	kW_e
Electricity consumption	88.9	90.5	kW_e
Net electricity	1,254.3	-	kW_e
Net electricity efficiency	14.9	-	%

4.2. Economic values (TCI, TPC, POI, and PBP)

Table 4 shows the economic values of the processes for electricity or NPS production from rice husk. The TCI for electricity generation (Case 1) including combustion, flue gas treatment, and steam turbines was 13,447 k$. In Case 2 including hydrothermal synthesis reactor, filter, and dryer for NPS production, TCI was 13,705 k$. The TPCs of Case 1 and Case 2 were 998 and 1,476 k$/y.
Case 1 was not economically feasible because of negative ROI. The ROI and PBP of Case 2 were 3.7 %/y and 17.5 y, respectively.

Table 4. Economic values for electricity or NPS plants from rice husk.

Economic values	Case 1	Case 2	Unit
Total capital investment (TCI)	13,446.9	13,705.1	k$
Total production cost (TPC)	997.9	1,476.1	k$/y
Return on investment (ROI)	-	3.7	%/y
Payback period (PBP)	-	17.5	y

5. Conclusions

The agricultural by-products are produced from crop production, and an eco-friendly process is required to treat the by-products. In this study, the economic feasibility of electricity or NPS production process was compared using the same amount of rice husk (50 t/d). The electricity production plant included a combustor and steam turbine. The NPS plant included a combustor, reactor, filter, and dryer. An electricity of 1,254 kW_e were produced with a net electrical efficiency of 14.9%. The 278 kg/h NPS was produced using the same amount of RH. The TCI and TPC of the electricity production plant were13,447 k$ and 998 k$/y, respectively, which was not profitable. Those of the NPS production plant were13,705 k$ and 1,476 k$/y, respectively. The ROI was 3.7%/y, which may be attractive as a carbon-neutral technology.

References

M.M. Bergqvist, K. Samuel Wårdh, A. Das, E.O. Ahlgren, 2008, A techno-economic assessment of rice husk-based power generation in the Mekong River Delta of Vietnam, International Journal of Energy Research, 32, 12, 1136-1150.

T.X. Do, Y.-i. Lim, H. Yeo, Y.-t. Choi, J.-h. Song, 2014, Techno-economic analysis of power plant via circulating fluidized-bed gasification from woodchips, Energy, 70, 547-560.

FAO, 2020, World Food and Agriculture - Statistical Yearbook 2020.

J.G. Kim, J.Y. Kim, 2020. Method for Manufacturing Water Glass Using Rice Husk or Rice Husk Ash, no. 10-2097728.

S. Kim, Y.-I. Lim, 2021, Heat Integration and Economic Analysis of Dry Flue Gas Recirculation in a 500 MW e Oxy-coal Circulating Fluidized-bed (CFB) Power Plant with Ultra-supercritical Steam Cycle, Korean Chemical Engineering Research, 59, 1, 60-67.

S. Kim, Y.I. Lim, D. Lee, M.W. Seo, T.Y. Mun, J.G. Lee, 2021, Effects of flue gas recirculation on energy, exergy, environment, and economics in oxy-coal circulating fluidized-bed power plants with CO2 capture, International Journal of Energy Research, 45, 4, 5852-5865.

Y.-I. Lim, J. Choi, H.-M. Moon, G.-H. Kim, 2016, Techno-economic comparison of absorption and adsorption processes for carbon monoxide (CO) separation from Linze-Donawitz gas (LDG), Korean Chemical Engineering Research, 54, 3, 320-331.

Y. Liu, Y. Guo, Y. Zhu, D. An, W. Gao, Z. Wang, Y. Ma, Z. Wang, 2011, A sustainable route for the preparation of activated carbon and silica from rice husk ash, Journal of hazardous materials, 186, 2-3, 1314-1319.

S. Mor, C.K. Manchanda, S.K. Kansal, K. Ravindra, 2017, Nanosilica extraction from processed agricultural residue using green technology, Journal of cleaner production, 143, 1284-1290.

S. Munasir, Z. Triwikantoro, 2013, Book Synthesis of silica nanopowder produced from Indonesian natural sand via alkalifussion route, 1555, Issue, 28-31.

P.P. Nayak, S. Nandi, A.K. Datta, 2019, Comparative assessment of chemical treatments on extraction potential of commercial grade silica from rice husk, Engineering Reports, 1, 2, e12035.

P. Peerapong, B. Limmeechokchai, 2009, Thermodynamic and Economic Analysis of 1.4 MWe Rice-Husk Fired Cogeneration in Thailand, Journal of Renewable Energy and Smart Grid Technology, 4, 2, 35-46.

R. Pode, 2016, Potential applications of rice husk ash waste from rice husk biomass power plant, Renewable and Sustainable Energy Reviews, 53, 1468-1485.

J. Prasara-A, S.H. Gheewala, 2017, Sustainable utilization of rice husk ash from power plants: A review, Journal of Cleaner production, 167, 1020-1028.

S. Steven, E. Restiawaty, Y. Bindar, 2021, Routes for energy and bio-silica production from rice husk: A comprehensive review and emerging prospect, Renewable and Sustainable Energy Reviews, 149, 111329.

D. Subbukrishna, K. Suresh, P. Paul, S. Dasappa, N. Rajan, 2007, Book Precipitated silica from rice husk ash by IPSIT process, Issue, 7-11.

T.T. Vu, Y.-I. Lim, D. Song, T.-Y. Mun, J.-H. Moon, D. Sun, Y.-T. Hwang, J.-G. Lee, Y.C. Park, 2020, Techno-economic analysis of ultra-supercritical power plants using air-and oxy-combustion circulating fluidized bed with and without CO2 capture, Energy, 194, 116855.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50017-8

Future biofuels: A Superstructure-Based Optimization Framework Integrating Catalysis, Process Synthesis, and Fuel Properties

Juan-Manuel Restrepo-Flórez[a], Christos T. Maravelias[b,c*]

[a.]*Department of Chemical and Biological Engineering, University of Wisconsin–Madison, 1415 Engineering Drive, Madison, 53706, USA*
[b.]*Department of Chemical and Biological Engineering, Princeton University, 50-70 Olden St, Princeton, 08540, USA*
[c.] *Andlinger Center For Energy And Environment, Princeton University, 86 Olden St, Princeton, 08540, USA*
maravelias@princeton.edu

Abstract

In this work, we present a process systems engineering framework that allows the integration of catalysis, process synthesis, and fuel property modelling enabling the systematic design of fuels with tailored properties and the biorefineries able to produce them. Methodologically, the proposed framework relies on a superstructure-based formulation in which three hierarchical decisions are made: what chemical products will undergo chemical transformations? what chemistries will be used to transform these chemical products? and which specific catalyst and processes will be used? This optimization framework is coupled with a fuel property model such that the decisions made at the process synthesis level are constrained by the desired fuel properties. We apply this framework to the problem of ethanol upgrading with an emphasis on three specific problems: First, we show how we can design biorefineries for the production of fuels (gasoline, diesel, or jet fuel) with specific properties. Second, we study the interplay between fuel properties and profit, and we show how the constraints imposed on the fuel properties impact both the optimal biorefinery designed and its economics. Finally, we show how the studied framework can be used to find the optimal biorefinery associated with a specific chemistry or catalyst. The results presented constitute the first systematic and comprehensive study of ethanol upgrading in which the simultaneous process and product design are considered.

Keywords: Biorefineries; superstructure; process synthesis; biofuels; product design.

1. Introduction

In the last twenty years, ethanol has been established as the dominant biofuel. However, ethanol has significant limitations: first, it can only be blended at a maximum 10% level with gasoline; and second, it is a poor replacement for middle distillates (Eagan et al., 2019a). These limitations are becoming increasingly problematic. At the same time, it is likely that in the near future there will be a surplus of ethanol, provided that the demand for blending at 10% is satisfied, and the consumption of gasoline will decrease as electric and fuel cell vehicles penetrate the market (Eagan et al., 2019a, 2019b; Fasahati and Maravelias, 2018). These challenges have prompted the search for technologies

enabling the chemical transformation of ethanol into more fungible components. With this goal in mind numerous chemistries, catalysts, and processes have been developed (Eagan et al., 2019a; Sun and Wang, 2014). At first, research was focused on designing ethanol upgrading strategies yielding fuel products with properties similar to those of currently used fossil fuels. In recent years, however, we have come to realize that this vision is short-sighted (König et al., 2020). The diversity of chemistries associated with ethanol, which is known for been a platform chemical, can be exploited to design fuel products with properties not only equal but better than currently used fossil fuels. The realization of this vision requires the integration of such diverse areas as catalysis, process synthesis, and fuel property modelling (Restrepo-Flórez and Maravelias, 2021).

In this work we develop a superstructure-based framework for the automated design of biorefineries for ethanol upgrading. This framework is conceived such that we can simultaneously consider the design of processes (biorefineries) and fuels. Thus, enabling the design of biofuels with tailored properties. We apply this framework three problems (1) the identification of the optimal refinery to produce gasoline, diesel, and jet fuel (2) the characterization of the relation between fuel properties and profit, and (3) the identification of the optimal technology associated with a specific chemistry.

2. Superstructure description

An upgrading strategy can be defined as a sequence of chemical transformations leading to the production of products with desired properties. There are at least three decisions associated with the design of an upgrading strategy (1) which molecules will undergo transformation (2) which chemistries will be used to transform these molecules, and (3) which processes (reaction-separation system) will be used such that these chemical transformations can be accomplished. To represent this sequence of nested decisions we propose a hierarchical superstructure (Figure 1) containing three levels, each of which is associated with one of the aforementioned decisions (1) technology group, (2) technology and (3) module.

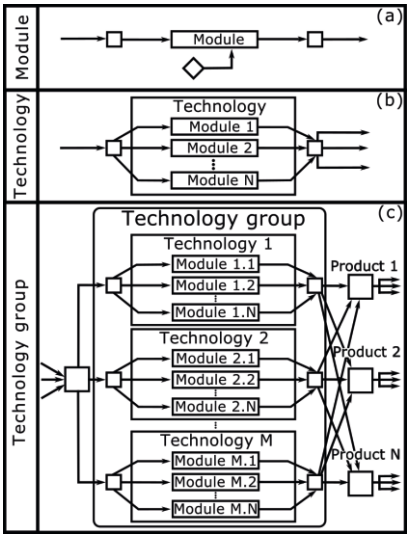


Figure 1. Superstructure architecture proposed and the three decision levels represented (a) Module (b) Technology (c) Technology group

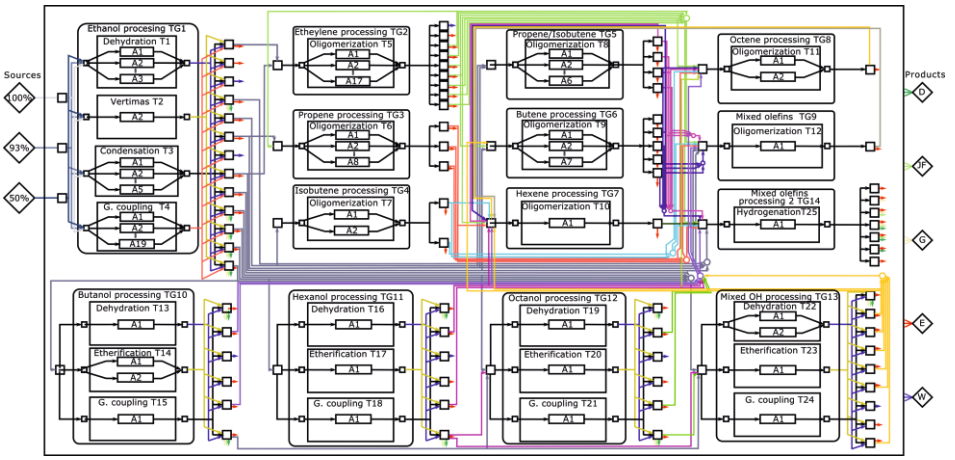


Figure 2. Ethanol upgrading superstructure containing the chemistries that can be used as well as feedstocks and final products. D: Diesel, JF: Jet fuel, G: Gasoline, E: Electricity, W: Waste.

In the case of ethanol, we use the architecture inf Figure 1 to build a comprehensive superstructure capturing the multiple chemistries available for ethanol upgrading (Figure 2). This superstructure is designed to consider three ethanol sources with different purity (50%, 93%, and 99.9%), and five products: gasoline, jet fuel, diesel, electricity, and waste. The superstructure is comprehensive because (1) it contains a representative set of the chemical reactions available for ethanol upgrading, and (2) it is richly connected, allowing to capture all feasible sequences of upgrading steps.

3. Mathematical model

The problem of finding the optimal sequence of upgrading steps to produce a given fuel can be represented as a mixed inter non-linear program (MINLP) model. The simplified mathematical representation is shown in Eq. (1). To write the model, we define three types of binary variables to account for the discrete decisions made at each of the superstructure levels (1) Y_i^{TG} $(\forall i \in \mathbf{I}^{\mathrm{TG}})$ (2) Y_i^{T} $(\forall i \in \mathbf{I}^{\mathrm{T}})$ and (3) Y_i^{M} $(\forall i \in \mathbf{I}^{\mathrm{M}})$, where $\mathbf{I}^{\mathrm{TG}}$ is the set of technology groups considered, $\mathbf{I}^{\mathrm{T}}$ the set of technologies, and $\mathbf{I}^{\mathrm{M}}$ the set of modules. The equations in the mathematical model can be grouped in (1) Process equations, used to model the selection of technology groups, technologies, and catalyst; represent mass balances for the different superstructure units; calculate capital and operating costs associated with the selected units; and enforce the superstructure connectivity (2) A fuel property model, used to estimate the values of the most relevant fuel properties.

$$Max(\text{Profit}) \qquad (1)$$
$$s.t. \begin{cases} \text{Process equations} \\ \text{Fuel property model} \end{cases}$$

The fuel property model consists of: Linear blending rules used to estimate the value of viscosity (υ), density (ρ), cetane number (CN), and octane number (RON) as a function of the fuel composition; a model of the distillation profile constructed based on the true boiling point approximation, according to which the components of a blend boil sequentially based on their boiling points; and a set of constraints limiting the amount of certain components (olefins, aromatics, and ethanol) in the final fuel blend.

4. Results

Optimal refineries to produce gasoline, jet fuel and diesel

The framework that we developed can be used to find the optimal sequence of upgrading operations required to produce a fuel with similar properties to gasoline, jet fuel, or diesel. We show these results by means of a Sankey diagram in Figure 3 (a-c). Additionally, we show the breakdown of capital and operating costs for each of these refineries in Figure 3(d-f). The simplest refinery also yielding the higher economic benefit is the one used to produce gasoline (Figure 3(a)). In this case, a Guerbet coupling module, followed by a hexanol dehydration module is enough to produce a blend satisfying the imposed constraints. In the case of jet fuel, the optimal biorefinery consists of an ethanol Guerbet coupling module followed by a butanol dehydration module and a sequence of oligomerization reactions. Additionally, the refinery also contains a small etherification module, and a hydrogenation module aimed at reducing the olefin content in the final fuel. Finally, in the case of diesel, the optimal biorefinery consists of a Guerbet coupling module, followed by a butanol dehydration module and a sequence of oligomerizations. This biorefinery also contains a hydrogenation unit. We note that in all cases the most important economic driver is the cost of the feedstock. This implies that to improve the biorefinery economics it is important to find strategies to reduce the cost of ethanol, or alternatively to increase the biorefinery's yield.

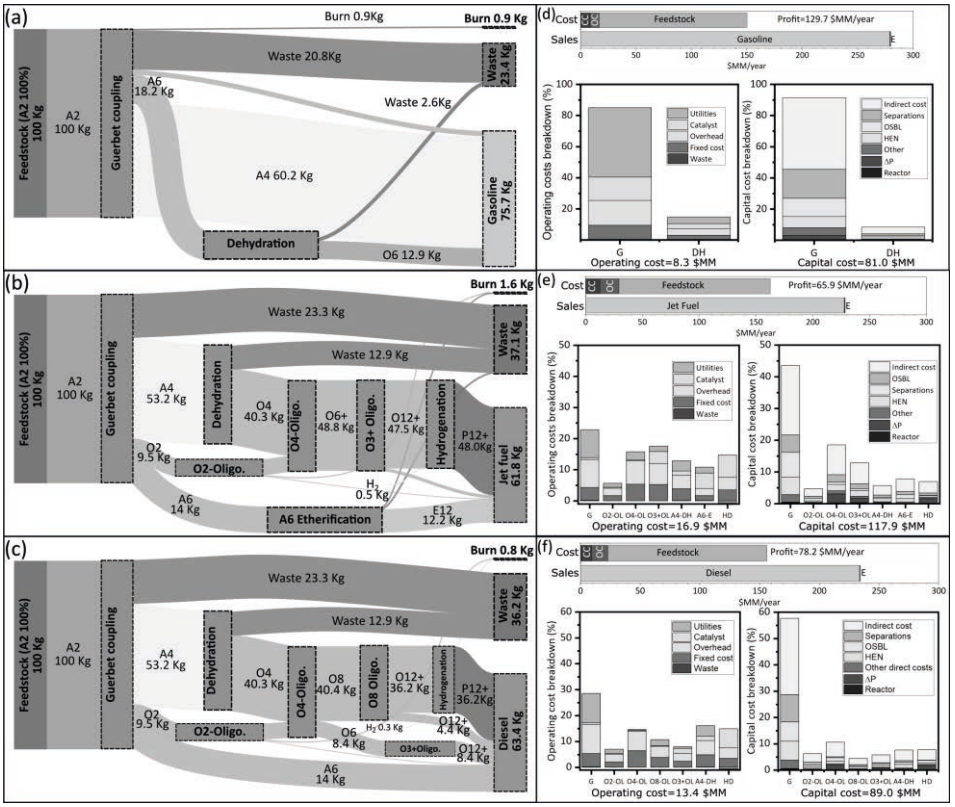


Figure 3. (a)(b)(c) Sankey diagram with the mass flows in the optimal biorefineries. (d)(e)(g) Capital and operating costs breakdown in these biorefineries.

The role of complexity

In Figure 4, we show the relation that exist between the biorefinery complexity, measured as the number of modules, and the profit and fuel composition in a diesel production biorefinery. Increasing complexity leads to a higher profit, but at the same time increases the operational challenges of the system. Understanding this relation is fundamental in the design stage of upgrading refineries.

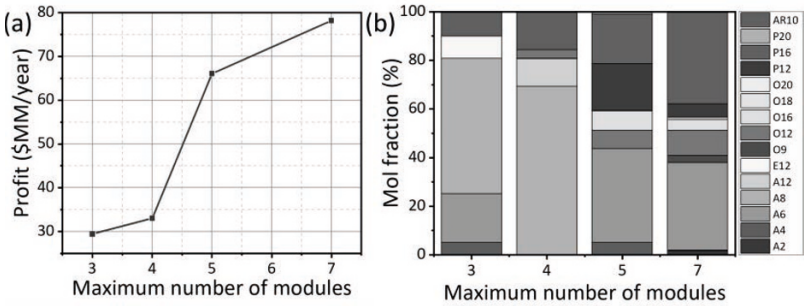


Figure 4. Effect of complexity on (a) profit and (b) fuel composition

The role of properties on the biorefinery economics

The production of high-quality fuels from ethanol is an exciting opportunity that opens the door to a new paradigm in biofuel research. In Figure 5, we study the role of fuel properties on the biorefinery economics (Figure 5(a)) and fuel composition (Figure 5(b)). Particularly, we study the effect cetane number on a diesel production biorefinery. Cetane number (CN) has been identified as a key property to mitigate NO_x emissions. Having a biofuel with high CN can serve two purposes: first, such a fuel burns cleaner; second, it can be used in blends with fossil diesel to raise the overall quality of the fuel. From Figure 5(a), we see that producing fuels with higher CN impacts the refinery economics, the higher the CN the lower the profit. In terms of fuel composition (Figure 5(b)), we note that as the CN increases so does the fraction of ethers (known for having a high CN) in the fuel. It is important to highlight that finding strategies to produce these fuels while simultaneously considering their properties was only possible because we used a framework able to capture the complexity of the problem.

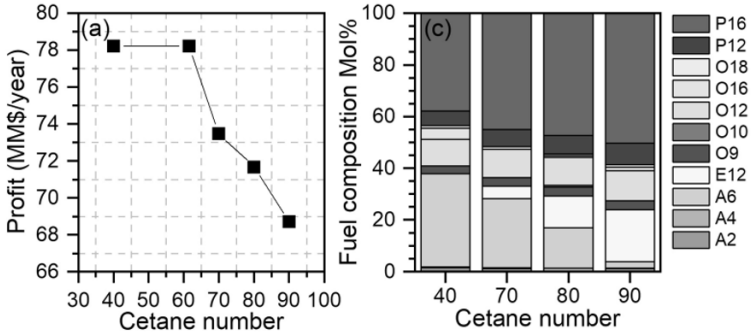


Figure 5. Effect of cetane number on (a) process economics and (b) fuel composition. Components are labelled using a character to identity of the functional group (P: paraffin, O: olefin, E: ether, A: alcohol), and a number to denote the number of carbons in the molecule.

Identification of optimal technologies

Another capability of the developed framework consists in providing insights as to the optimal strategy that can be used to produce a specific fuel by using a specific chemistry. For example, in Figure 6, we show the optimal biorefinery for diesel production obtained when we force the system to use ethanol dehydration. The strategy selected consist in dehydrating ethanol to ethylene, and then use a sequence of oligomerization reactions to increase the molecular weight. A final hydrogenation unit to reduce the olefin content is also employed. This kind of approach is useful to researchers working in the development of a specific chemistry to identify how their work fits into a broader context.

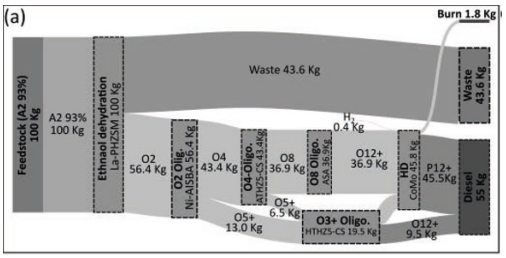


Figure 6. Sankey diagram showing the optimal refinery compatible with ethanol dehydration to produce diesel fuel

4. Conclusions

In this work, we developed a superstructure optimization approach to study the problem of ethanol upgrading toward fuels with tailored properties. We showed optimal ethanol upgrading strategies for the production of gasoline, jet fuel, and diesel. The most important cost driver in all cases was the cost of feedstock. We studied the relation between profit and biorefinery complexity and showed that increasing complexity may lead to improvements in the process economics. Additionally, we explored the relation between fuel properties and profit in the context of a diesel production biorefinery. We proved that it is possible to upgrade ethanol toward diesel fuel with high cetane number, with a superior quality than its fossil counterpart. This contrasts with typical approaches for biofuel production, focused on finding fuels with the same quality than fossil fuels.

References

Eagan, N.M., Kumbhalkar, M.D., Buchanan, J.S., Dumesic, J.A., Huber, G.W., 2019a. Chemistries and processes for the conversion of ethanol into middle-distillate fuels. Nat. Rev. Chem. 3, 223–249.

Eagan, N.M., Moore, B.J., McClelland, D.J., Wittrig, A.M., Canales, E., Lanci, M.P., Huber, G.W., 2019b. Catalytic Synthesis of Distillate-Range Ethers and Olefins from Ethanol through Guerbet Coupling and Etherification. Green Chem. 1–12.

Fasahati, P., Maravelias, C.T., 2018. Advanced Biofuels of the Future: Atom-Economical or Energy-Economical? Joule 2, 1915–1919.

König, A., Neidhardt, L., Viell, J., Mitsos, A., Dahmen, M., 2020. Integrated design of processes and products: Optimal renewable fuels. Comput. Chem. Eng. 134.

Restrepo-Flórez, J.M., Maravelias, C.T., Advanced fuels from ethanol – a superstructure optimization approach. Energy Environ. Sci., 2021, **14**, 493-506

Sun, J., Wang, Y., 2014. Recent advances in catalytic conversion of ethanol to chemicals. ACS Catal. 4, 1078–1090.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50018-X

Superstructure Optimization of Biodiesel Production from Continuous Stirred Tank and Membrane Reactors

Thien An Huynh[a*], Vincent Reurslag[a], Maryam Raeisi[a], Meik B. Franke[a] and Edwin Zondervan[a]

[a]*Sustainable Process Technology, Faculty of Science and Technology, University of Twente, Meander, kamer 216, Postbus 217, 7500 AE Enschede, the Netherlands*
t.a.huynh@utwente.nl

Abstract

This work presents a superstructure model with the objective to maximize the total profit of biodiesel production by reducing the production cost and increasing the value of the by-product glycerol. The heat integration of the superstructure model is a novel feature which allows further reduction of utility costs and energy consumption of the biodiesel separation. The superstructure model is used to optimize two biodiesel production scenarios from a conventional continuous stirred tank reactor (CSTR) and a membrane reactor (MR). The superstructure optimization is solved with Advanced Interactive Multidimensional Modeling System (AIMMS) software. The annual profit of the new optimized production pathway for the conventional reactor is 840,606 $. The biodiesel production pathway with the membrane reactor consumed 70% less energy than the conventional reactor. However, the production cost of the MR is nearly two times higher than the CSTR due to the low biodiesel yield of the membrane reactor. The results show the potential to improve traditional biodiesel production and make intensified production methods more viable with the superstructure optimization.

Keywords: Biodiesel, Superstructure, Optimization, AIMMS, Process, Design.

1. Introduction

Biodiesel is a biofuel which is mainly obtained from chemical reactions between vegetable oil or animal fat with alcohol in the presence of a catalyst (Knothe et al., 2010). Biodiesel has become a potential solution for reducing greenhouse gas (GHG) because it has a lower net carbon dioxide (CO_2) emission than fossil fuels. CO_2 released from biodiesel engines is absorbed by plants which will be the feedstock for biofuel production making this a circular process (Hanaki and Portugal-Pereira, 2018). However, biodiesel is more expensive than fossil fuels, which poses a significant challenge for integrating the biofuel into GHG reduction strategies.

The cost of biodiesel can be reduced by optimizing its production which consists of reaction and purification processes. Intensified reactor designs which combine reaction and separation into one operation unit have been developed to improve biodiesel conversion and purity. A membrane reactor is a process intensification option which integrates a membrane separation into a cross-flow reactor to produce higher quality biodiesel than conventional reactors (Cao et al., 2008). Besides the reaction, the purification process plays an important role in biodiesel production as it accounts for 60-80% of the total production cost (Atadashi et al., 2011). Therefore, the optimization

of a biodiesel purification process has become an important research topic. For example, several biodiesel purification scenarios have been simulated and analysed to identify the optimal biodiesel production process from soybean oil (Myint and El-Halwagi, 2009).

To design an optimal biodiesel production process, two methods are commonly used: the heuristic approach and the superstructure-based approach. The heuristic approach is based on rules derived from experience and understanding of unit operations while the superstructure approach is based on optimization algorithms and mathematical models to identify the optimal process from all possible alternatives (Tula et al., 2017). However, a disadvantage of the heuristic approach is that the interaction between different process stages and levels of detail are difficult to capture. The superstructure approach solves design problems simultaneously as a mathematical programming problem and therefore does not have this disadvantage (Mencarelli et al., 2020).

Superstructure optimization has become more popular in recent researches of biochemical process design. AlNouss et al. (2019) used superstructure optimization to develop an economic and environmentally friendly gasification process, which produces fuels, fertilizers, and power from multiple biomass sources. Galanopoulos et al. (2019) developed a superstructure framework for optimizing the design of an integrated algae biorefinery which can reduce the cost of biodiesel production up to 80%. However, superstructures for biodiesel production are usually generalized with a minimum numbers of operating units and a simplified glycerol purification process.

Therefore, this work proposes a superstructure model for biodiesel production that includes a wide range of operating units, a detailed glycerol purification section and heat integration functions. The model is used to optimize two biodiesel production processes from: a) continuous stirred tank reactor and b) membrane reactor. The results are compared with a conventional biodiesel production process (Zhang et al., 2003a,b).

2. Superstructure development

2.1. Problem statement

Given are the composition of feedstock and products from the transesterification reactor and options of processing equipment which are grouped into tasks and stages, and the technical and economic specifications of processing options. Under conditions that: 1) The possible processing routes are represented by logical constraints where each processing option is associated with a logical decision variable. 2) The flow rates in and out of an option complies with mass balance constraints. 3) The energy requirements are calculated based on the flowrates. 4) A heat integration function which is capable of matching hot and cold streams is integrated for further reduction of heating and cooling requirements. 5) The investment and operating costs are calculated according to according to the flowrates and energy requirement. The superstructure optimization problem decides the optimal biodiesel processing route while complying with logical, mass and energy constraints, and ASTM standards of biodiesel product (Zhang et al., 2003a), while maximizing the total profit of the biodiesel refinery.

2.2. Superstructure topology

The superstructure of the biodiesel purification section has 28 technical options which are relating to 28 binary decision variables and grouped into different tasks including phase separation, methanol removal, neutralization, washing and purification. By

grouping similar options into tasks, the superstructure can be defined easier. The possible processing routes which are combinations of different options over 5 consecutive stages are presented in Figure 1. The input stream of the superstructure is the product stream of the transesterification reactor which converts vegetable oil into biodiesel. The main output is the biodiesel stream with purity of 99.65% according to ASTM standards. The glycerol output of phase separation tasks becomes the input of glycerol treatment superstructure.

In Figure 2, the glycerol superstructure is useful in deciding the numbers of treatment stages depending on the initial purity of glycerol input and the final grade of glycerol output. The final glycerol grades are waste glycerol (~ 50% - 85% wt. glycerol), crude glycerol (~85% - 98%) and technical glycerol (~98% - 99.5%) (Bart et al., 2010).

2.3. Mathematical model

The mathematical model includes mass balances of component k in each option j as shown in Eq.(1) and Eq.(2).

$$m_{j,k}^{P} = m_{j,k}^{F} \cdot SF_{j,k} \cdot y_i \tag{1}$$

$$m_{j,k}^{W} = m_{j,k}^{F} \cdot (1 - SF_{j,k}) \cdot y_i \tag{2}$$

where $m^{F}_{j,k}$, $m^{P}_{j,k}$ and $m^{W}_{j,k}$ are mass flow rates of feed (*kg/h*), product and waste streams of component k in and out option j, respectively. $SF_{j,k}$ is the split factor which indicates how much of component k going to product stream from the feed stream. y_j is the binary decision variable which is 1 if the option is selected and 0 if the option is not selected. The product stream of an option will be the feed stream of the next option on the same process route. The equipment cost (*USD*) of a technical option, EC_j, is presented in Eq.(3) (Seider et al., 2016).

$$EC_j = EC_j^{Ref,year} \cdot \left(\frac{m_j^F}{m_j^{F,Ref}}\right)^E \cdot \left(\frac{CE^{2020}}{CE^{year}}\right) \cdot y_j \tag{3}$$

where $EC^{Ref,year}{}_j$, $m^{F,Ref}$ and CE^{year} are the reference cost of the equipment, the reference capacity and the Chemical Engineering Index of the reference year, respectively. The total capital investment (*TCI*) is shown in Eq.(4) (Seider et al., 2016).

$$TCI = 1.05 \cdot f_{L,TCI} \cdot \sum_j (EC_j) \tag{4}$$

where 1.05 is the delivery cost of equipment to the plant location and $f_{L,TCI}$ is the Lang factor with value of 5.93 (Seider et al., 2016). The total annualized capital investment (*TACI*) is calculated with interest rate (*IR*) (0.1) and total project lifetime (*LT*) (20 years) as shown in Eq.(5).

$$TACI = TCI \cdot \frac{IR \cdot (IR+1)^{LT}}{(IR+1)^{LT} - 1} \tag{5}$$

The objective function is to maximize the total annualized profit (*TAP*) as follows.

$$\max TAP = BDS + GLS - TACI - TAOP \tag{6}$$

where the total annual operating costs (*TAOP*), the annual biodiesel sales (*BDS*) and glycerol sales (*GLS*) are defined from the mass flow rate of the superstructure.

Heat integration of the superstructure optimization model is a function based on Pinch Technology to minimize the heating and cooling requirements of the biodiesel production. First, a series of heat intervals defined from the temperature differences of the product streams which are designated as hot streams or cold streams depending on their heating or cooling requirements. Second, the function selects hot and cold streams based on the decision variable in each product stream. Third, the hot and cold streams are matched with each other according to their temperature to calculate the total heat load of heat intervals and set up the heat cascade. Finally, the minimum hot and cold utility requirements can be predicted by balancing the negative heat interval of the infeasible heat cascade. To reduce the complexity of the model, the heat exchanger network and investment costs are not considered in the heat integration function.

The mathematical model is implemented in the software AIMMS, version 4.82.3.29 64-bit. The AIMMS solver is the Outer Approximation Algorithm, which is an algorithm using CPLEX 20.1 as MIP solver and CONOPT 4.1 as NLP solver. The model includes 1,602 constraints and 1,629 variables with 43 binary variables. The optimization problem is solved in an average of 1.83 s with a CPU Intel(R) Core(TM) i5-8265U CPU @ 1.80 GHz and 8.00 RAM.

3. Results and discussion

The superstructure model is applied for two base cases of biodiesel produced from CSTR and MR. The feedstock is rapeseed oil, infeed flowrate 1000 kg/h. The costs are calculated based on the price of biodiesel, feedstock, chemical and equipment in 2020.

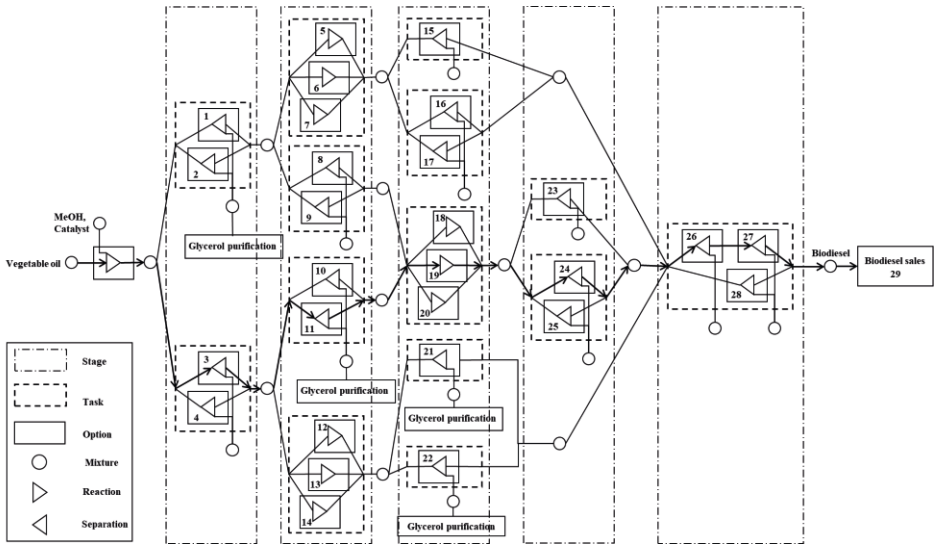


Figure 1: Superstructure of biodiesel purification from the transesterification of vegetable oil. The optimal processing route is the arrow line.

For the case of biodiesel produced in a CSTR, the optimal processing route is presented by the arrow line in Figure 1 and 2. The separation of methanol and glycerol at the first and second stages increases methanol recycle and reduces downstream equipment costs.

The third stage is neutralization of the base catalyst with H_2SO_4, then dry washing the product stream with magnesol. Finally, water, methanol and unreacted oil are removed from the biodiesel stream with vacuum flash evaporators to achieve the purity standard. The glycerol stream from second stage goes through neutralization and decanter to increase the glycerol content to 95%. The glycerol is sold as crude glycerol.

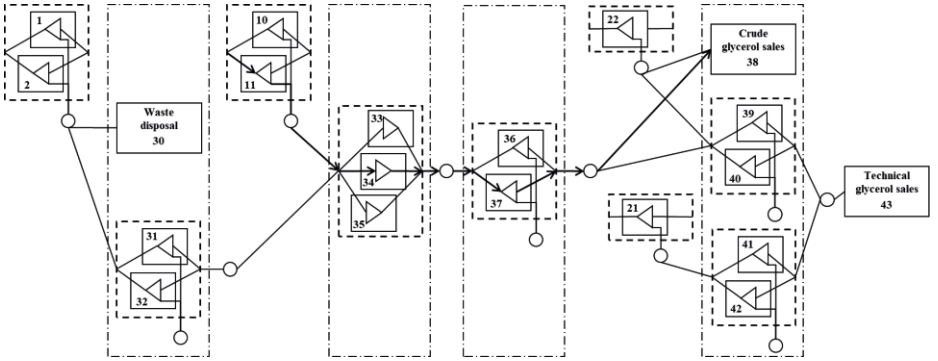


Figure 2: Superstructure of treatment routes for the glycerol separated from biodiesel production process. The arrows show the optimized glycerol processing route.

Options 1, 10, 22 and 36 are centrifuges; 2, 11 and 37 are decanters; 3, 8 and 31 are flash evaporators; 4, 9, 27, 32, 40 and 42 are vacuum distillation columns; 5, 12, 18 and 33 are neutralization with H_3PO_4; 6, 13, 19 and 34 are neutralization with H_2SO_4; 7, 14, 20 and 35 are neutralization with HCl; 15, 21 are 23 are water washing; 16 and 24 are dry washing with magnesol; 17 and 25 are dry washing with ion exchange resins; 26, 27, 39 and 41 are vacuum flash evaporators. Option 30 is treatment of waste glycerol with counting as expense of the process. Options 29, 38 and 43 are selling biodiesel, crude and technical glycerol, respectively.

The total annualized profit of the optimal process is 840,606 $ which is higher than the process proposed by Zhang (2003b). The explanation for finding different profits is that the optimal process has lower production costs by using magnesol dry washing instead of water washing, a system of flash evaporators instead of distillation columns and heat integration to reduce energy consumption.

For the case of MR, the superstructure model gives the same optimal processing route as the case of the CSTR. The difference is within the final purification stage where the MR case uses only one vacuum flash evaporator to remove methanol and water, because the membrane reactor removes the unreacted oil. Therefore, the energy requirement is 70% lower than the case of CSTR. However, the membrane reactor has only 56% biodiesel yield of the conventional reactor making the annualized profit of the process of membrane reactor negative, -2,126,584 $.

4. Conclusions

A superstructure model for optimizing the biodiesel production has been developed. The superstructure can be developed faster and expanded easier by grouping similar options into processing tasks. With a novel heat integration function, the proposed model can be used to identify the best processing route which minimizes the production cost and the energy requirement. The optimization results show the potential for improvement of biodiesel production in terms of economic and environment indicators. The optimized

process of the CSTR shows a good profit and the case of MR has better energy consumption. However, the superstructure only considers one feedstock and two types of reactor. The superstructure will be extended to cover a large range of feedstock and different reaction technologies to further reduce the cost of biodiesel production.

References

A. AlNouss, G. McKay, T. Al-Ansari, 2019, Superstructure Optimization for the Production of Fuels, Fertilizers and Power using Biomass Gasification, Computer Aided Chemical Engineering, 46, 301-306, doi: 10.1016/B978-0-12-818634-3.50051-5.

A.K. Tula, D.K. Babi, J. Bottlaender, M.R. Eden, R. Gani, 2017, A computer-aided software-tool for sustainable process synthesis-intensification, Computers & Chemical Engineering, 105, 74-95, doi: 10.1016/j.compchemeng.2017.01.001.

C. Galanopoulos, P. Kenkel, E. Zondervan, 2019, Superstructure optimization of an integrated algae biorefinery, Computers & Chemical Engineering, 130, doi: 10.1016/j.compchemeng.2019.106530.

G. Knothe, J. Krahl, J.V. Gerpen, 2010, The biodiesel handbook 2nd ed., Urbana: Academic Press and AOCS Press.

I.M. Atadashi, M.K. Aroua, A.R. Abdul Aziz, N.M.N. Sulaiman, 2011, Refining technologies for the purification of crude biodiesel, Applied Energy, 88, 12, 4239-4251, doi: 10.1016/j.apenergy.2011.05.029.

J.C.J. Bart, N. Palmeri, S. Cavallaro, 2010, chap 13 - Valorisation of the glycerol by-product from biodiesel production, Biodiesel Science and Technology, Biodiesel Science and Technology, Woodhead Publishing, Cambridge, UK, 571-624, doi: 0.1533/9781845697761.571.

K. Hanaki, J. Portugal-Pereira, 2018, The Effect of Biofuel Production on Greenhouse Gas Emission Reductions, Biofuels and Sustainability: Holistic Perspectives for Policy-making, Springer Japan, 53-71, doi: 10.1007/978-4-431-54895-9_6.

L.L. Myint, M.M. El-Halwagi, 2009, Process analysis and optimization of biodiesel production from soybean oil, Clean Techn Environ Policy,11, 263–276, doi: 10.1007/s10098-008-0156-5.

L. Mencarelli, Q. Chen, A. Pagot, I.E. Grossmann, 2020, A review on superstructure optimization approaches in process system engineering, Computers & Chemical Engineering, 136, doi: 10.1016/j.compchemeng.2020.106808.

P. Cao, M.A. Dubé, A.Y. Tremblay, 2008, High-purity fatty acid methyl ester production from canola, soybean, palm, and yellow grease lipids by means of a membrane reactor, Biomass and Bioenergy, 32, 11, 1028-1036, doi: 10.1016/j.biombioe.2008.01.020.

T.A. Huynh, E. Zondervan, 2021, Dynamic modeling of fouling over multiple biofuel production cycles in a membrane reactor, Chemical Product and Process Modeling, doi: 10.1515/cppm-2020-0093.

W.D. Seider, D.R. Lewin, J.D. Seader, S. Widagdo, R. Gani, K.M. Ng, 2016, Product and Process Design Principles: Synthesis, Analysis and Evaluation - 4th Editiony, John Wiley & Sons Inc., New York, USA.

Y. Zhang, M.A. Dubé, D.D. McLean, M. Kates, 2003a, Biodiesel production from waste cooking oil: 1. Process design and technological assessment, Bioresource Technology, 89, 1, 1-16, doi: 10.1016/S0960-8524(03)00040-3.

Y. Zhang, M.A. Dubé, D.D. McLean, M. Kates, 2003b, Biodiesel production from waste cooking oil: 2. Economic assessment and sensitivity analysis, Bioresource Technology, 90, 3, 229-240, doi: 10.1016/S0960-8524(03)00150-0.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50019-1

Process Design and Techno-Economic Analysis of Biomass Pyrolysis By-Product Utilization in the Ontario and Aichi Steel Industries

Jamie Rose, Thomas A. Adams II*

McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada, L8S 4L7
**Corresponding Author: tadams@mcmaster.ca*

Abstract

Iron- and steel-making companies throughout the globe have been aiming to reduce emissions. One method to do so is to replace pulverized coal used in blast furnaces with biochar, but biochar is currently far more expensive than coal. To increase the value of biochar, by-products of pyrolysis can be combusted to generate heat and offset fossil fuel usage. In this study, pyrolysis by-product combustion was studied using Aspen Plus and process cost models to offset fuels in both Ontario, Canada, and Aichi, Japan. It was found that each tonne of biochar made produces by-products which save 130 USD and 1.47 t CO_2-e of emissions in Ontario, while in Aichi 96 USD and 2.44 t CO_2-e are saved.

Keywords: Biochar, Pyrolysis, Pulverized Coal Injection, Iron, Steel

1. Introduction

Steel production currently accounts for about 8% of annual anthropogenic carbon emissions (Worldsteel Association, 2021a). One method of reducing emissions is replacing coal used in pulverized coal injection in blast furnaces with biocarbon produced from the pyrolysis of biomass (Ye et al., 2019). However, widespread biochar usage has several hurdles, one of which is that it is prohibitively expensive at present. There is currently little published information on wholesale biochar prices, and the few data points available are not particularly recent. For example, in values of USD_{2021}, wholesale prices were 2400 USD/t (metric tonne) in 2015 (Campbell et al., 2018). Research has suggested production costs may drop to 870 USD/t with small scale production (Keske et al., 2020) or 240 USD/t in a large-scale production facility designed for an economy that uses biochar heavily (Project Drawdown, n.d.). In comparison, steam coal is typically only 70 USD/t(U.S. Energy Information Administration, 2021). Therefore, there is incentive to reduce the net cost of using biochar to match or even go below that of coal.

Another issue with biomass pyrolysis is that it also produces by-products, which are often considered to be waste and are difficult to handle due to toxicity (Bridgwater et al., 1999). The by-products of biomass pyrolysis are separated into two phases, including bio-oil, also known as tar, and light gases (Dunnigan, Ashman, et al., 2018). The light gases generally consist of CO, CO_2, CH_4, H_2, and low carbon fuel gases, while the tar phase consists of water and volatile organic compounds (VOCs) (Amini et al., 2019).

To tackle both of these issues, it is worthwhile to investigate the value of utilization of the by-products of biomass pyrolysis. Although there are studies which looked at tire pyrolysis by-product value (Czajczyńska et al., 2017), usage of by-products for self-sustaining pyrolysis (Xu et al., 2011), the economic value of bio-oil specifically (Badger et al., 2011), and the heating value of biomass pyrolysis by-products (Dunnigan, Morton, et al., 2018), there have not been any comprehensive techno-economic analyses which cover environmental and economic benefits of the utilization of biomass pyrolysis

by-products. Given that the products mostly consist of combustible hydrocarbons, one of the simplest potential methods to use these by-products is heat generation through combustion. This heat can be used for processes such as steam generation, iron production, or even biomass drying and pyrolysis. This allows for fossil fuel usage to be offset, thereby reducing purchase and emission costs, which increases the value of biochar. This value can be used to close the gap between biochar and coal prices.

For this analysis, Aspen Plus chemical process simulation software was used to calculate the thermodynamics and products of combustion of pyrolysis by-products based on experimental compositions and conditions. These results were then compared to fuel and carbon prices used in iron- and steel-making facilities in two locations. The locations investigated were Aichi Prefecture in Japan, which is in the third largest steel producing country in the world (Worldsteel Association, 2021b), and the province of Ontario in Canada. These locations were chosen because Aichi and Ontario both produce a similar amount of steel, at about 10,000,000 t annually (Aichi Prefectural Government, 2017), (Cheminfo Services Inc., 2019), but use different fuels for heat generation. Relevant data were readily available for the most commonly used heating fuels both locations, allowing for a complete cost comparison.

Collaboration with and data sharing from ArcelorMittal Dofasco, Natural Resources Canada (NRCan), and CHAR Technologies has allowed for a realistic determination of the value and feasibility of pyrolysis by-product usage. ArcelorMittal Dofasco is aiming to replace up to 40,000 t of pulverized coal with biochar per year, so this value was used for design calculations.

2. Methods

To determine the financial and environmental value of pyrolysis by-product combustion, the heat generated from combustion was considered to be used to offset the currently most-used non-renewable fuels in local iron- and steel-making facilities. According to data from ArcelorMittal Dofasco, natural gas is generally the only fuel that is purchased for heat generation in their plant. Therefore, by-product value was determined based on offsetting natural gas in the Ontario case. However, in Japan, iron- and steel-making companies tend to use both natural gas and steam coal, but approximately four times more heat is generated with steam coal than natural gas (Japan Iron and Steel Federation, 2020). Therefore, in the Aichi case, steam coal will be assumed to be the main fuel that is offset with by-product combustion. Since the pyrolysis was done with biomass, emissions from by-product combustion are carbon neutral if it is assumed that the biomass would not otherwise be used for carbon sequestration. Therefore, emissions reductions from offsetting fossil fuels with pyrolysis by-product combustion were considered to be direct reductions.

Data on the composition of by-products were received from NRCan's lab-based experiments from the pyrolysis of construction and demolition wood at 600 °C. These data include the ratio of biochar, bio-oil, and light gas produced from pyrolysis, as well as bio-oil and light gas compositions. The distribution of products from wood pyrolysis is shown in Table 1. These ratios are similar to others in literature (Amini et al., 2019).

Table 1: Product distribution of pyrolysis of wood on a dry, ash-free basis

Pyrolysis Product	Mass % of Initial Feedstock
Light Gas	27.6
Bio-oil	44.8
Biochar	27.6

In regards to data used for simulation, the composition of the light gas is given in Table 2, while the composition and ultimate analysis of the bio-oil are given in Table 3 and Table 4, respectively. Although the bio-oil composition given in this paper includes only general categories of compounds, the actual data set used for simulation includes approximately 30 specific compounds. Also, it is common for pyrolysis to be done in a nitrogen-rich atmosphere, but the method used by CHAR Technologies creates positive pressure in the chamber shortly after pyrolysis begins, preventing combustion. This means that the by-products do not contain any nitrogen gas. CHAR Technologies also noted that the pyrolysis process can be considered to be steady state.

Table 2: Composition of light gas by-product of pyrolysis on a dry basis

Light Gas Component	Composition (Volume %)
H_2	9.4
CO	26.2
CO_2	43.0
CH_4	17.2
C_2H_6	1.3
C_2H_4	0.6
Other Light Hydrocarbons	2.3

Table 3: Composition of the bio-oil by-product of pyrolysis

Bio-oil Component	Composition (Mass %)
Water	56.6
Acids	9.6
Other Oxygenates	8.9
Methanol	7.6
Phenols	3.6
Furans	1.6
Other Condensable Compounds	12.1

Table 4: Ultimate analysis of the bio-oil on a wet basis

Ultimate Analysis Element	Mass %
Carbon	26.6
Hydrogen	9.5
Oxygen	63.9

The pyrolysis by-products contain many VOCs, which are gaseous at the 600 °C pyrolysis process outlet temperature but can begin to condense at temperatures below 450 °C, as per data from CHAR Technologies. Therefore, it was imperative that the process was designed so that the by-products can be combusted without condensation. Although it is typical to use thermal oxidizers to destroy gaseous VOCs while recovering a portion of the heat of combustion, thermal oxidizers are used for flue gases which contain up to only 10,000 ppmv organic compounds, with the rest being air (Wang et al., 2020). For destruction of streams without oxygen and that contain VOCs in higher concentrations, a vapour combustor, also known by other names such as enclosed flare (Anguil, n.d.), should be used instead (Gulf Coast Environmental Systems, n.d.). A vapour combustor is essentially a small flue gas stack with the option to recover the heat of combustion, and related operating and capital costs were found for annual usage of 40 kt of biochar using

published correlations (United States Environmental Protection Agency, 1980). A diagram of the process design is shown in Figure 1.

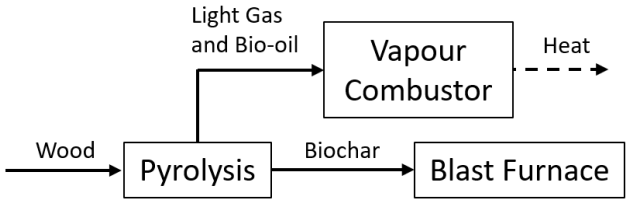


Figure 1: A system based on pyrolysis by-products which generates heat with a vapour combustor

The products of by-product combustion were predicted using an RGIBBS block in Aspen Plus, which calculates the products and enthalpy change of a reaction through minimizing Gibbs free energy based on the parameters and composition of the reactants used. In the model, pyrolysis by-products at 600 °C and 1.01325 bar in a gaseous phase were mixed with air at 25 °C and 1.01325 bar so that the products contained 2 % oxygen by volume after combustion, as per guidelines from ArcelorMittal Dofasco. The property method used was the Peng-Robinson-Boston-Mathias (PR-BM) model, which has been shown in literature to work well for mixtures of CO_2 and hydrocarbons (Li et al., 2019). Peng-Robinson-based methods have also been shown to predict CO_2-H_2O well (Zhao & Lvov, 2016). This simulation model was also used to determine that the by-products are within the flammability envelope when mixed with up to 30 % excess air, as per the calculated adiabatic flame temperature method (Hansel et al., 1992). Aspen Plus was also used to calculate higher and lower heating values of the by-product stream. This was done by adding the known lower heating values of the reactants for the LHV and then adding to this heat of vaporization of product water to determine the HHV.

Cost savings gained from by-product combustion in each location were calculated based on local fuel costs and carbon prices. Specific values used for each situation as well as the equation used for cost calculation are available in the supporting document (Rose & Adams, 2021).

3. Results

Results for the calculated HHV and LHV of the light gas, bio-oil, and weighted by-product mixture are given in Table 5.

Table 5: Calculated heating values for the pyrolysis by-products

By-Product Stream	LHV (MJ/kg)	HHV (MJ/kg)
Light Gases	10.0	10.9
Bio-Oil	10.7	11.4
Mix	10.5	11.2

Given the pyrolysis product ratios in Table 1 and these heating values, it was found that each tonne of biochar made also creates enough by-products to produce 29.4 GJ HHV or 27.4 GJ LHV of heat through combustion. Also, for a vapour combustor system that uses 40 kt of biochar per year, capital and operating costs were found to be 912,000 USD_{2021} total and 271,000 USD_{2021} per year, respectively. Assuming a 20-year project lifetime, these values were then used to calculate specific future value cost savings and carbon emissions reductions from offsetting fossil fuels through vapour combustion, as shown in Table 6. At a rate of 1 tonne of pulverized coal used per 10 tonnes of metal produced (U.S. Department of Energy, 2000), if all pulverized coal for 10 Mt of metal

production were to be replaced with biochar, there would be an annual emissions reduction of 1.50 Mt CO_2-e in Ontario or 2.44 Mt CO_2-e in Aichi, equivalent to taking 625 thousand or one million cars off the road, respectively (Wynes & Nicholas, 2017).

Table 6: Cost savings per tonne of biochar produced from offsetting fossil fuels with by-product combustion in a vapour combustor

Location	Cost Savings/t Biochar 2022 Case	Cost Savings/t Biochar 2030 Case	Emissions Reductions (t CO_2-e/t Char)
Ontario	135 USD_{2021}	280 USD_{2021}	1.50
Aichi	96 USD_{2021}	350 USD_{2021}	2.44

4. Conclusions

Combustion of pyrolysis by-products has been shown to be a viable method for increasing the value of biochar as a replacement for pulverized coal in blast furnaces. Even with the purchase and operation of new equipment, by-product combustion can increase the value of one tonne of biochar by anywhere from 96 to 350 USD_{2021} in Aichi, Japan, and 135 to 280 USD_{2021} in Ontario, Canada. The greater difference in Aichi is due a greater reduction in carbon emissions with a similar increase in carbon taxes. These reductions are up to 1.50 tCO2-e/t char used in Ontario and 2.44 tCO2-e/t char used in Aichi, applicable for up to one million tonnes of biochar used per year in each location.

References

Aichi Prefectural Government. (2017). *全国からみた愛知県の工業 [Industry in Aichi Prefecture as Compared to the Whole of Japan]* (pp. 1–16).

Amini, E., Safdari, M. S., DeYoung, J. T., Weise, D. R., & Fletcher, T. H. (2019). Characterization of pyrolysis products from slow pyrolysis of live and dead vegetation native to the southern United States. *Fuel*, *235*(June 2018), 1475–1491. https://doi.org/10.1016/j.fuel.2018.08.112

Anguil. (n.d.). *Vapor Combustor Unit (VCU)*. https://anguil.com/air-pollution-control-solutions/vapor-combustor-unit-vcu/

Badger, P., Badger, S., Puettmann, M., Steele, P., & Cooper, J. (2011). Techno-economic analysis: Preliminary assessment of pyrolysis oil production costs and material energy balance associated with a transportable fast pyrolysis system. *BioResources*, *6*(1), 34–47. https://doi.org/10.15376/biores.6.1.34-47

Bridgwater, A. V., Meier, D., & Radlein, D. (1999). An overview of fast pyrolysis of biomass. *Journal of Inorganic Biochemistry*, *Organic Ge*, 1479–1493.

Campbell, R. M., Anderson, N. M., Daugaard, D. E., & Naughton, H. T. (2018). Financial viability of biofuel and biochar production from forest biomass in the face of market price volatility and uncertainty. *Applied Energy*, *230*(June), 330–343. https://doi.org/10.1016/j.apenergy.2018.08.085

Cheminfo Services Inc. (2019). *Economic Assessment of the Integrated Steel Industry*. https://canadiansteel.ca/files/resources/Final-Report-Economic-Assessment-of-the-Integrated-Steel-Industry.pdf

Czajczyńska, D., Krzyżyńska, R., Jouhara, H., & Spencer, N. (2017). Use of pyrolytic gas from waste tire as a fuel: A review. *Energy*, *134*, 1121–1131. https://doi.org/10.1016/j.energy.2017.05.042

Dunnigan, L., Ashman, P. J., Zhang, X., & Kwong, C. W. (2018). Production of biochar from rice husk: Particulate emissions from the combustion of raw pyrolysis volatiles. *Journal of Cleaner Production*, *172*, 1639–1645. https://doi.org/10.1016/j.jclepro.2016.11.107

Dunnigan, L., Morton, B. J., Ashman, P. J., Zhang, X., & Kwong, C. W. (2018). Emission

characteristics of a pyrolysis-combustion system for the co-production of biochar and bioenergy from agricultural wastes. *Waste Management*, *77*, 59–66. https://doi.org/10.1016/j.wasman.2018.05.004

Gulf Coast Environmental Systems. (n.d.). *Vapor Combustor Unit*. http://www.gcesystems.com/air-pollution-control/vapor-combustor-unit.html

Hansel, J. G., Mitchell, J. W., & Klotz, H. C. (1992). Predicting and controlling flammability of multiple fuel and multiple inert mixtures. *Plant/Operations Progress*, *11*(4), 213–217. https://doi.org/10.1002/prsb.720110408

Japan Iron and Steel Federation. (2020). *鉄鋼業における発電設備運用の実態 [Actual Conditions of Power Generation Equipment Operation in the Steel Industry]*. https://www.meti.go.jp/shingikai/enecho/denryoku_gas/denryoku_gas/sekitan_karyoku_wg/pdf/002_07_00.pdf

Keske, C., Godfrey, T., Hoag, D. L. K., & Abedin, J. (2020). Economic feasibility of biochar and agriculture coproduction from Canadian black spruce forest. *Food and Energy Security*, *9*(1), 1–11. https://doi.org/10.1002/fes3.188

Li, C., Gao, Y., Xia, S., Shang, Q., & Ma, P. (2019). Calculation of the Phase Equilibrium of CO2–Hydrocarbon Binary Mixtures by PR-BM EOS and PR EOS. *Transactions of Tianjin University*, *25*(5), 540–548. https://doi.org/10.1007/s12209-019-00194-y

Project Drawdown. (n.d.). *Biochar Production*. https://drawdown.org/solutions/biochar-production/technical-summary

Rose, J., & Adams, T. A. II (2021). *Supplemental Data for "Process Design and Techno-Economic Analysis of Biomass Pyrolysis By-Product Utilization in the Ontario and Aichi Steel Industries."* http://psecommunity.org/LAPSE:2021.0800

U.S. Department of Energy. (2000). *Blast Furnace Granulated Coal Injection System Demonstration Project: A DOE Assessment*. *June*, 46.

U.S. Energy Information Administration. (2021). *Quarterly Coal Report April–June 2021*. https://www.eia.gov/coal/production/quarterly/pdf/qcr-all.pdf

United States Environmental Protection Agency. (1980). *Organic Chemical Manufacturing Volume 4: Combustion Control Devices*. https://nepis.epa.gov/Exe/ZyPDF.cgi/2000MF11.PDF?Dockey=2000MF11.PDF

Wang, F., Lei, X., & Hao, X. (2020). Key factors in the volatile organic compounds treatment by regenerative thermal oxidizer. *Journal of the Air and Waste Management Association*, *70*(5), 557–567. https://doi.org/10.1080/10962247.2020.1752331

Worldsteel Association. (2021a). *Climate change and the production of iron and steel*. https://www.worldsteel.org/en/dam/jcr:228be1e4-5171-4602-b1e3-63df9ed394f5/worldsteel_climatechange_policy%2520paper.pdf

Worldsteel Association. (2021b). *Global crude steel output decreases by 0.9% in 2020*. https://www.worldsteel.org/media-centre/press-releases/2021/Global-crude-steel-output-decreases-by-0.9--in-2020.html

Wynes, S., & Nicholas, K. A. (2017). The climate mitigation gap: education and government recommendations miss the most effective individual actions. *Environmental Research Letters*, *12*(7), 74024. https://doi.org/10.1088/1748-9326/aa7541

Xu, R., Ferrante, L., Hall, K., Briens, C., & Berruti, F. (2011). Thermal self-sustainability of biochar production by pyrolysis. *Journal of Analytical and Applied Pyrolysis*, *91*(1), 55–66. https://doi.org/10.1016/j.jaap.2011.01.001

Ye, L., Peng, Z., Wang, L., Anzulevich, A., Bychkov, I., Kalganov, D., Tang, H., Rao, M., Li, G., & Jiang, T. (2019). Use of Biochar for Sustainable Ferrous Metallurgy. *Jom*, *71*(11), 3931–3940. https://doi.org/10.1007/s11837-019-03766-4

Zhao, H., & Lvov, S. N. (2016). Phase behavior of the CO2-H2O system at temperatures of 273-623 K and pressures of 0.1-200 MPa using Peng-Robinson-Stryjek-Vera equation of state with a modified Wong-Sandler mixing rule: An extension to the CO2-CH4-H2O system. *Fluid Phase Equilibria*, *417*, 96–108. https://doi.org/10.1016/j.fluid.2016.02.027

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50020-8

Optimal design of solar-aided hydrogen production process using molten salt with CO_2 utilization for ethylene glycol production

Wanrong Wang, Nan Zhang, and Jie Li*
Department of Chemical Engineering, The University of Manchester, Manchester M13 9PL, UK
jie.li-2@manchester.ac.uk

Abstract

In this work, a machine-learning based optimisation framework is proposed for optimal design of solar steam methane reforming using molten salt (SSMR-MS) with CO_2 capture and utilisation. The computational results demonstrate that significant profit in TAC can be made compared with the existing SSMR-MS. With ethylene glycol (EG) production, the optimal Levelised cost of Hydrogen Production (LCHP) is 0.00 \$ kg^{-1} which is largely reduced compared to the existing process with LCHP of 2.40 \$ kg^{-1}. The captured CO_2 using the amine-based solution is utilized to produce around 33.59 kt y^{-1} EG.

Keywords: Hydrogen; Solar energy; CO_2 utilization; Machine learning

1. Introduction

Hydrogen is an important energy carrier in the transportation sector and essential industrial feedstock for petroleum refineries, methanol, and ammonia production. The global demand for hydrogen is expected to increase 10-fold by 2050, clearly indicating its significant role in the future (Wang et al., 2021). Conventional hydrogen production primarily utilises natural gas and oil-based feedstock for steam reforming, which results in considerable greenhouse gas emissions mainly CO_2, thus contributing to global warming (Voldsund et al., 2016). The damaging consequences of global warming deem further investigation into clean and affordable hydrogen production process using renewable energy sources to be crucial. Meanwhile, research is also ongoing into CO_2 capture and utilisation technology which considers CO_2 as a viable alternative carbon source for the chemical supply chain (Alper et al., 2017) to obtain value-added products such as methanol, ethylene carbonate and ethylene glycol (Yang et al., 2021).

Solar energy for hydrogen production has received significant attention in recent years due to its primary abundance as an energy source (Koumi Ngoh et al., 2012). To effectively use solar energy for large-scale hydrogen production, an optimal design of solar steam methane reforming using molten salt (SSMR-MS) which shows great potential has been studied to reduce TAC and CO_2 emission (Wang et al., 2021). However, the optimal Levelised Cost of Hydrogen Production (LCHP) is still much higher than that of the conventional methane steam reforming. Furthermore, in their work CO_2 removal model is represented using a simple separation block with a constant separation efficiency, which could lead to inaccurate account of annualized cost of CO_2 capture. To further reduce LCHP and improve the model accuracy, an integrated rate-based CO_2 removal model in SSMR-MS along with CO_2 utilization for ethylene glycol (EG) production is investigated in this work. This is the main novelty of this work.

In this work, an optimization framework from (Wang et al., 2021) is extended for such optimal design of SSMR-MS with integration of CO_2 capture and utilization. The artificial neural network (ANN) is employed to establish relationships of total annualised cost (TAC), hydrogen production rate, molten salt duty and gas flowrates from CO_2 capture unit with thirteen independent input variables in SSMR-MS. A hybrid global optimisation algorithm is employed to solve the developed optimisation problem and generate the optimal design, which is then validated in Aspen Plus V8.8 and SAM. The computational results demonstrate that TAC of the SSMR-MS process can be compensated by the profit of selling EG and CO_2 emissions reduction by 68.92 % can be achieved compared to the existing SSMR-MS process in Wang et al. (2021). Captured CO_2 can produce around 33.59 kt yr^{-1} EG.

2. Problem description

Figure 1 illustrates a schematic diagram of integrated system including SSMR-MS for large-scale hydrogen production, CO_2 capture, and EG production. A detailed description of the SSMR-MS process has been made in Wang et al. (2021). The pre-reformer is non-adiabatic and molten salt transfers concentrated solar energy in heat to pre-reformer. The flow scheme in the pre-reformer is in co-current. The process is to produce F_{H_2} hydrogen with η_{H_2} purity to satisfy hydrogen demand in an oil refinery. A rate-based CO_2 removal model using methyl diethanolamine (MDEA) as the solution is built in Aspen Plus V8.8. The reaction for CO_2 absorption and MDEA regeneration process using MDEA are listed in Moioli et al. (2016). The capture CO_2 with coke oven gas is used to produce EG.

The objective is to minimize total annualized cost (TAC) of the integrated system, which includes total annualized capital cost and operating cost.

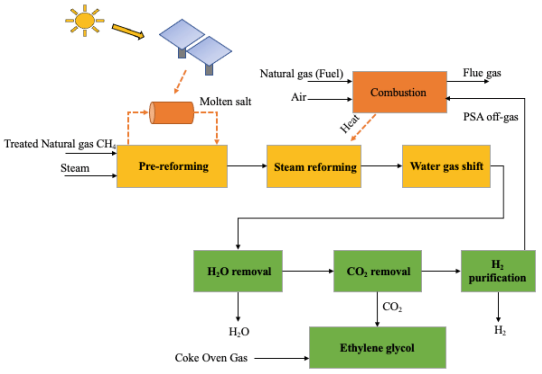


Figure 1: Block diagram of the SSMR-MS process (Adapted from Wang et al., 2021)

3. Mathematical formulation

The integrated process can be modelled using rigorous models in Aspen Plus V8.8. The optimisation problem using these rigorous models is often highly complex. To reduce the complexity of the problem, surrogate model is developed based on machine learning techniques. ANN is used to create surrogate models of the rate-based CO_2 removal process and the entire process. A major advantage of ANN over other statistical techniques is the ability to correlate multiple inputs to multiple outputs, leading to

compact models that can be implemented in an optimisation environment with ease (Ibrahim et al., 2018).

In this work, a surrogated model for the CO_2 capture process is firstly developed and integrated within SSMR-MS process in Aspen Plus V8.8. This is because when the rigorous rate-based CO_2 removal model for CO_2 capture is integrated with the SSMR-MS process in Aspen Plus V8.8, the simulation of the integrated system is extremely difficult to converge. The input variables for the CO_2 capture process include the inlet stream component flowrate of CH_4, H_2O, CO, CO_2, H_2 which are denoted as $F_{in,CH_4,MDEA}$, $F_{in,H_2O,MDEA}$, $F_{in,CO,MDEA}$, $F_{in,CO_2,MDEA}$ and $F_{in,H_2,MDEA}$, respectively and temperature ($T_{in,MDEA}$) obtained from hydrogen production process. In other words,

$$\boldsymbol{z} = [F_{in,CH_4,MDEA}, F_{in,H_2O,MDEA}, F_{in,CO,MDEA}, F_{in,CO_2,MDEA}, F_{in,H_2,MDEA}, T_{in,MDEA}]^T.$$

The outlet stream flowrates of CH_4, H_2O, CO, CO_2 and H_2 in the CO_2 removal process are predicted using ANN surrogate models respectively, as shown in Eqs.1-5.

$$F_{out,CH4,MDEA} = ANN_1(\boldsymbol{z}) \tag{1}$$

$$F_{out,H2O,MDEA} = ANN_2(\boldsymbol{z}) \tag{2}$$

$$F_{out,CO,MDEA} = ANN_3(\boldsymbol{z}) \tag{3}$$

$$F_{out,CO2,MDEA} = ANN_4(\boldsymbol{z}) \tag{4}$$

$$F_{out,H2,MDEA} = ANN_5(\boldsymbol{z}) \tag{5}$$

These surrogate models are then integrated with the rigorous models of SSMR-MS by using user model within Aspen Plus interface with Excel Link (Fontalvo, 2014) for sample generation. Then a new surrogate model representing the entire integrated process is constructed through extending the optimisation framework of Wang et al. (2021). There are usually three steps for the development of a surrogate model, including data generation (i.e., sampling), construction of the surrogate model and construction of feasibility constraints using a support vector machine (Wang et al., 2021). Samples generated using the Latin hypercube sampling method are used as input in Aspen Plus V8.8 to get values for the corresponding output variables.

In the integrated process, the independent input variables including molar flowrate of natural gas into pre-reformer F_{NG}, steam to methane ratio $\gamma_{S/C}$, operating temperature of reformer T_R, high-temperature water gas shift (HWGS) reactor T_{HWGS}, low-temperature water gas shift (LWGS) reactor T_{LWGS}, tube length of pre-reformer L_{PR}, reformer L_R, HWGS reactor L_{HWGS} and LWGS reactor L_{LWGS}, tube number in pre-reformer N_{PR}, reformer N_R, HWGS reactor N_{HWGS} and LWGS reactor N_{LWGS} vary between lower and upper bounds. A vector **x** is used to denote all these variables. In other words,

$$\boldsymbol{x} = [F_{NG}, \gamma_{S/C}, T_R, T_{HWGS}, T_{LWGS}, L_{PR}, L_R, L_{HWGS}, L_{LWGS}, N_{PR}, N_R, N_{HWGS}, N_{LWGS}]^T.$$

$$x^L \leq \boldsymbol{x} \leq x^U \tag{6}$$

The objective function TAC can be calculated as follows,

$$TAC = Ccapital \cdot ACCR + Cproduction \tag{7}$$

where $Ccapital$ is total capital investment. $ACCR$ is annual capital charge ratio. $Cproduction$ is the total production cost per year.

The optimisation problem using the surrogate models is stated as follows,

$$
\begin{aligned}
&(PS) \quad Min \quad && TAC = TAC_1 + TAC_{solar} \\
&\qquad s.t. \quad && TAC_1 = ANN_6(\boldsymbol{x}) + ANN_7(\boldsymbol{x}) \\
&\qquad\qquad && TAC_{solar} = f(Q_{MS}) \\
&\qquad\qquad && \mathrm{Q_{MS}} = \mathrm{ANN_8(\mathbf{x})} \\
&\qquad\qquad && F_{H_2} = ANN_9(\boldsymbol{x}) \geq F_{H_2}^{TA} \\
&\qquad\qquad && \textit{Eq. (6)}
\end{aligned}
$$

where TAC_1 is non-solar related cost, TAC_{solar} is the solar related cost. $ANN_6(\boldsymbol{x})$ is CO_2 removal process related cost. ANN_7 is the non-solar related cost excluding MDEA unit. $\mathbf{x}$ is the set of independent variables in hydrogen production process, Q_{MS} is molten salt duty. The relationship of solar-related equipment cost, and molten salt duty is described using an algebraic linear function $f(Q_{MS})$. The surrogate model comprises 4 artificial neural networks as indicated above in the optimization problem PS and a linear regression model $f(Q_{MS})$. F_{H_2} denotes the predicated flowrate of hydrogen.

4. Solution algorithm

A hybrid optimisation algorithm similar to that of (Wang et al., 2021) is employed to solve the optimisation problem PS, as shown in Figure 2. This hybrid algorithm combines the advantages of the stochastic optimisation algorithm and the deterministic optimisation method. We employ different platforms and data are transferred between them to exploit their strength and reduce the computational complexity. In sample generation process, Matlab is used as the core platform to interact with other programs. Sample points are imported to Aspen Plus. Within Aspen Plus, the process contains a user model which calls Visual Basic Application (VBA) in Excel (Fontalvo, 2014) as a bridge to transfer data between Aspen Plus user model and Matlab (ANN model for composition prediction in CO_2 removal process). The hybrid algorithm is implemented in MatLab R2019a.

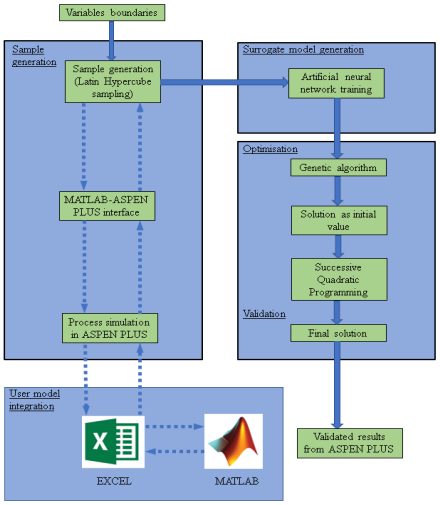


Figure 2: Flowchart of the extended design methodology

5. Computational studies

The extended optimisation framework is used to generate the optimal design of the SSMR-MS process with integration of CO_2 capture and utilisation. The hydrogen production rate is 2,577 kmol h^{-1}. The desired hydrogen purity is 99.9 vol%. Other data can be referred to Wang et al. (2021). The results are given in Table 1. It can be seen that the optimal TAC is 166.50 M$ y^{-1}. The optimal steam to carbon ratio is 2.7.

Table 1: Optimisation results for SSMR-MS from surrogate models

Item	Optimal value
$\gamma_{S/C}$	2.7
T_R (°C)	962.3
T_{HWGS} (°C)	421.1
T_{LWGS} (°C)	200.7
L_{PR} (m)	11.2
L_R (m)	12.0
L_{HWGS} (m)	4.8
L_{LWGS} (m)	4.3
N_{PR}	4,031
N_R	55
N_{HWGS}	1367
N_{LWGS}	2624
F_{NG} (kmol h^{-1})	781.9
Q_{MS} (MW)	14.54
F_{H_2} (kmol h^{-1})	2,577
TAC (M$ y^{-1})	166.50

The optimal values of independent variables in Table 1 are used as input in Aspen Plus V8.8 to generate values of all dependent variables. The validated results for Q_{MS}, F_{H_2} and TAC are 14.31 MW, 2577.2 kmol h^{-1}, 165.22 M$ y^{-1} respectively. The largest difference between actual results and predicted results from the ANN surrogate models is within 1 %, indicating the ANN model has high prediction accuracy.

Then heat integration is conducted to further reduce energy consumption. The final results are provided in Table 2. It can be observed after heat integration, TAC is 155.05 M$ y^{-1}, which is reduced by 6.2 % compared to that before heat integration (165.22 M$ y^{-1}). What is striking is that, with the integration of EG production, the whole hydrogen production process cost can be compensated with a large profit.

We also compare our optimal results with the best results from Wang et al. (2021) and the conventional steam methane reforming (denoted as CSMR), as shown in Table 2. It can be observed that without EG production, TAC in this work is higher than that from Wang et al. (2021) due to CO_2 capture cost increased using the rate-based CO_2 removal model. With an annual production of 33.59 kt EG, LCHP decreases from 2.40 $ kg^{-1} to 0

$ kg^{-1} due to high profit from EG. CO_2 emission reduces by 68.92 %. What is more important is that LCHP (0.00 $ kg^{-1}) is extremely economic attractive and CO_2 emission is reduced by 73.80 % compared to that of CSMR. The economic and environmental benefit obtained by using solar energy and applying CO_2 utilization process show the optimal case in this work is very promising for industrial hydrogen production.

Table 2: Comparative optimization results

Item	Optimal Case	Wang et al. (2021)	CSMR
Q_{MS} (MW)	14.31	10.20	20.00
F_{H_2} (kmol h^{-1})	2,577.2	2577.3	2,577.0
TAC without EG production (M$ y^{-1})	155.05	122.30	90.90
TAC with EG production (M$ y^{-1})	-21142.24	-	-
LCHP ($ kg^{-1})	0.00	2.40	2.00
CO_2 emission (kt y^{-1})	131.74	423.90	502.90
EG (kt y^{-1})	33.59	0.00	0.00

6. Conclusion

In this paper, the existing optimisation-based framework using machine learning techniques is extended for optimal design of solar steam methane reforming using molten salt (SSMR-MS) integrated with CO_2 capture and utilization for large-scale H_2 production. The computational results show that TAC was reduced largely with significant profit generated compared to the existing SSMR-MS. By considering CO_2 conversion, around 33.59 kt EG is produced per year. In the future, more process options for different pre-reformer operating conditions are expected to evaluate.

References

Alper, E. and Yuksel Orhan, O., 2017, CO_2 utilization: Developments in conversion processes. Petroleum, 3, 109-126.

Fontalvo Alzate, J., 2014. Using user models in Matlab® within the Aspen Plus® interface with an Excel® link. Ingeniería e Investigación, 34(2), pp.39-43.

Ibrahim, D., Jobson, M., Li, J. & Guillén-Gosálbez, G., 2018, Optimization-based design of crude oil distillation units using surrogate column models and a support vector machine, Chem. Eng. Res. Des., 134, 212-225.

Koumi Ngoh, S. & Njomo, D., 2012, An overview of hydrogen gas production from solar energy, Renewable and Sustainable Energy Reviews, 16, 6782-6792.

Moioli, S., Pellegrini, L., Picutti, B. and Vergani, P., 2013, Improved Rate-Based Modeling of H_2S and CO_2 Removal by Methyldiethanolamine Scrubbing. Industrial & Engineering Chemistry Research, 52, 2056-2065.

Voldsund, M., Jordal, K. & Anantharaman, R., 2016, Hydrogen production with CO_2 capture, International Journal of Hydrogen Energy, 41, 4969-4992.

Wang, W., Ma, Y., Maroufmashat, A., Zhang, N., Li, J. and Xiao, X., 2021, Optimal design of large-scale solar-aided hydrogen production process via machine learning based optimisation framework. Applied Energy, 305, 117751.

Yang, Q., Zhang, J., Chu, G., Zhou, H. and Zhang, D., 2021b, Optimal design, thermodynamic and economic analysis of coal to ethylene glycol processes integrated with various methane reforming technologies for CO_2 reduction. Energy Conversion and Management, 244,114538.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50021-X

Design of a novel hybrid process for membrane assisted clean hydrogen production with CO_2 capture through liquefaction

Donghoi Kim[ab*], Zhongxuan Liu[b], Rahul Anantharaman[a], Luca Riboldi[a], Lars Odsæter[a], David Berstad[a], Thijs A. Peters[c], Jonathan M. Polfus[c], Harald Malerød-Fjeld[d], and Truls Gundersen[b]

[a]*SINTEF Energy Research, Trondheim 7465, Norway*
[b]*Department of Energy and Process Engineering, Norwegian University of Science and Technology (NTNU), Trondheim 7491, Norway*
[c]*SINTEF Industry, Oslo 0314, Norway*
[d]*CoorsTek Membrane Sciences AS, 0349 Oslo, Norway*
donghoi.kim@sintef.no

Abstract

This work introduces a novel hybrid concept to produce H_2 from natural gas by using the protonic membrane reformer (PMR) with liquefaction based CO_2 capture system. For process intensification, recycling of the off-gas from the capture process and a water gas shift reactor for the retentate gas from the PMR are applied to the hybrid configuration. The suggested concept achieves around 99 % of system H_2 and CO_2 recovery rates even when the PMR is operated at relatively low hydrogen recovery, resulting in energy efficient H_2 production with a low carbon intensity.

Keywords: Low emission hydrogen production, proton conducting membrane, membrane reactor, CO_2 capture, hybrid process.

1. Background

Hydrogen is a clean fuel and is thus expected to play an important role in a future decarbonized energy scenario. Currently, 48 % of the world's hydrogen is produced by steam reforming (Voldsund et al., 2016), where natural gas and steam react to form hydrogen rich syngas. The focus on low-carbon hydrogen production from natural gas has been predominantly on CO_2 separation technologies. However, CO_2 separation does not contribute significantly to the energy penalty of the process (Voldsund et al., 2016). The largest losses are in the reforming of natural gas to hydrogen and subsequent separation to produce high purity hydrogen. The key focus area for cost-efficient low emission hydrogen production is an intensified process for hydrogen production and separation from natural gas with suitable CO_2 separation technology. Here we investigate an innovative hybrid technology for H_2 production with CO_2 capture combining H_2 production from natural gas by a protonic membrane reformer (PMR) technology with subsequent CO_2 separation by liquefaction in a novel integrated process. The technology enables a high carbon capture rate with high purity CO_2 and H_2 and a hydrogen cost comparable to conventional technologies without CO_2 capture.

The PMR technology produces high-purity hydrogen from steam methane reforming (SMR) in a single-stage electrochemical membrane reactor process with near-zero

energy loss (Malerød-Fjeld et al., 2017). The tubular membrane reformer comprises a BaZrO3-based proton-conducting electrolyte deposited as a dense film on a porous Ni composite electrode with a dual function as a reforming catalyst. Methane is steam-reformed to CO and H_2 over Ni particles inside the ceramic tube. Hydrogen is electrochemically transported as protons to the outer side, and CO is thereby converted to CO_2 as the water gas shift (WGS) equilibrium is shifted due to the extraction of H_2. The hydrogen produced is of high purity and electrochemically compressed in situ. The H_2 recovery in the PMR is proportional to the electricity input (Malerød-Fjeld et al., 2017). At high hydrogen recovery, the outlet composition is mainly CO_2 and steam. The retentate gas from the PMR has a relatively high fraction of CO_2, which makes CO_2 separation by liquefaction the most competitive technology for this application (Berstad et al., 2013). Liquefaction based CO_2 capture technologies have also been well tested for a wide range of syngas compositions with hydrogen (Kim et al., 2020).

Thus, in this work, different process configurations are developed in an analytical manner to combine the two technologies. One of the process concepts considers the appropriate placement of recycle streams to improve overall H_2 and CO_2 recovery when the PMR is operated at low H_2 recovery of around 90 % (for example reduced current density) for less energy intensive unit operation. Such operating conditions are also expected to lower stress on the material leading to prolonged life. Detailed process models of the different unit operations including the protonic membrane reactor are included in the hybrid system to analyse the different process options.

2. Hybrid process concepts

High recovery rates of H_2 and CO_2 are required on the plant level to achieve energy efficient low carbon hydrogen production for the PMR based system. This requires the development of optimal integration between the PMR and CO_2 liquefaction processes where additional process steps are considered. Figure 1 shows one of the process concepts for the PMR based hydrogen production with carbon capture. In this configuration, natural gas and water are heated by the hot temperature H_2 product and the retentate gas from the PMR. The mixture of natural gas and steam is then sent to a pre-reformer to convert heavier hydrocarbons in natural gas to hydrogen, CO, and CO_2 to supply a pre-reformed feed to the PMR. The pre-reformer outlet stream is set to have a fixed steam carbon ratio.

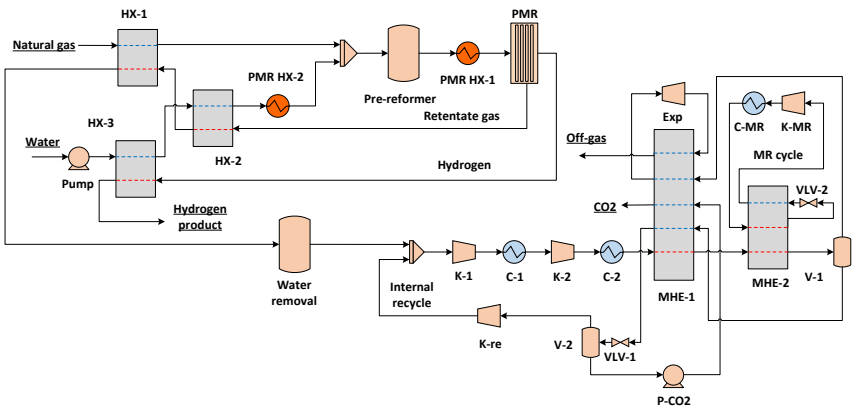


Figure 1. Process flow diagram of the simplified hybrid system for clean hydrogen production.

The temperature of the PMR feed is further increased to the operating temperature of the PMR by using the heat produced from the PMR, which is assumed to be operated isothermally. Then, the compressed pure hydrogen and the retentate gas are produced from the PMR. The SMR and WGS in the PMR result in a net endothermic reaction. However, the heat requirement can be covered by the heat generated by electricity used for the separation and compression of H_2 in the membrane, which is also enough to increase the temperature of the feed streams via PMR HX-1 and 2. The remaining PMR heat after the heat integration could be further used to produce steam.

The retentate gas from the heat recovery unit is fed to the CO_2 liquefaction process, after dehydration, to capture high purity liquid CO_2 while removing impurities in the liquid product through off-gas venting. In the CO_2 capture process (CCP), the dehydrated retentate gas is compressed before being liquefied by a hydrocarbon based mixed refrigerant (CH_4, C_2H_6, C_3H_8, and C_4H_{10}). The cold energy of the incondensable gas (off-gas) from the liquefier (MHE-2) is then used to pre-cool the compressed retentate gas. The off-gas from the pre-cooler is further utilized to supply the cold duty of heat exchanger MHE-1 by depressurizing it via a turbo expander. The liquid CO_2 product from the liquefier is also sent to the pre-cooler to cover the cold duty after being pressurized to the transport pressure. The off-gas leaving the CO_2 capture process could be vented or used as fuel to produce steam in the system.

Hydrogen production of this configuration is, however, dependent on the performance of the PMR as it is the only place where H_2 is extracted. If the hydrogen recovery rate (HRR) of the PMR is low with reduced electric power input, a considerable amount of H_2 left in the PMR is sent to the liquefaction process through the retentate gas. Since the hydrogen is not condensed in the CCP, it is lost through the off-gas, resulting in a low system HRR. Thus, when the PMR is operated at lower hydrogen recovery, the system HRR is also reduced, showing limited flexibility of the process. Another issue of the simplified concept with the PMR operating at low hydrogen recovery is the relatively high CO fraction in the retentate gas that causes poor performance of the CO_2 liquefaction system. The high fraction of CO in the feed to the CCP results in a deeper purification of the liquid CO_2 to achieve high purity. For the purification of the liquid CO_2, a larger amount of off-gas is produced, containing traces of CO_2, hence reducing the system carbon capture rate (CCR). The large flow rate of the off-gas stream will also require an extra facility to treat the CO and H_2 mixture. The high CO fraction, and thus a lower CO_2 fraction in the retentate gas, also means larger power consumption for the liquefaction process where the energy efficiency of the system is proportional to the CO_2 purity of feed gas (Kim et al., 2020).

To maintain H_2 production performance high at a low HRR of the PMR, the off-gas from the liquefaction system can be recycled (see Figure 2). This recycle allows collecting the valuable H_2 in the off-gas through the PMR, achieving a high system HRR. However, some off-gas venting will still be required to avoid N_2 accumulation in the system, which is assumed to be 10 % in this work. The off-gas recycle, however, will not reduce the CO fraction in the retentate gas, resulting in poor carbon capture performance of the hybrid concept. The improvement of the CCP can be achieved by a WGS reactor for the retentate gas as illustrated in Figure 2. The WGS reactor will convert the CO in the retentate gas to CO_2 and H_2, giving a low CO content and simultaneously increasing CO_2 content in the feed to the liquefaction process. Thus, this configuration can achieve high HRR and CCR while producing liquid CO_2 with negligible impurities even when the PMR is operate at low hydrogen recovery.

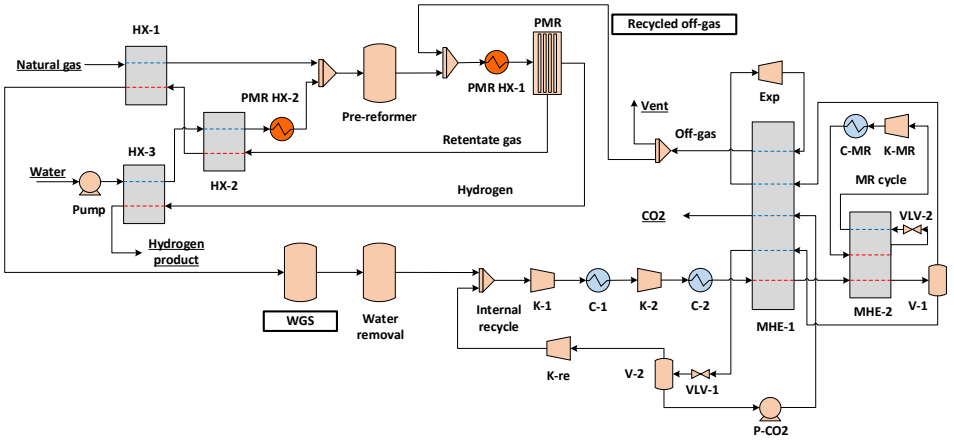


Figure 2. Process flow diagram of the modified hybrid system with off-gas recycle and a WGS reactor for the PMR operating at a low H_2 recovery (See text boxes for the modifications).

3. Modelling approach and design basis

In order to simulate the hybrid system, the PMR is modelled in C to represent the data from Malerød-Fjeld et al. (2017) and connected to Aspen HYSYS where all the other process units are built. In this work, the two process concepts neither include a vent gas utilization nor a steam cycle for the PMR surplus heat left after the heat integration. The PMR operating conditions that give 91 % of HRR are selected for the comparison of the two hybrid configurations assuming the membrane reformer is operated at relatively low H_2 recovery. However, it is worth mentioning that the PMR operating conditions such as temperature and current density will certainly impact process performance. While this has been analysed as part of this work, is not included in the paper. The hybrid system is assumed to have a natural gas feed rate of 3,000 kmol/h (lower heating value of 50 MJ/kg) to produce about 500 t/d hydrogen. CO_2 is assumed to be delivered at 150 bar with 99 mol% purity while allowing CO level lower than 0.5 vol%, assuming pipeline transport (Harkin et al., 2017). Other design conditions are listed in Table 1.

Table 1. Design basis for the PMR and the CO_2 capture process.

Parameters	Unit	Value
PMR feed steam to carbon ratio	-	2.5
PMR operating pressure	bar	26
PMR operating temperature	°C	800
PMR current density	A/m^2	7000
PMR H_2 product and retentate pressure	bar	26
PMR H_2 product and retentate temperature	°C	800
Pre-reformer inlet temperature	°C	450
WGS reactor inlet temperature	°C	200
ΔT_{min} for gas/gas heat exchanger	°C	30
ΔT_{min} for gas/liquid heat exchanger	°C	20
ΔT_{min} for low temperature heat exchanger	°C	3
Isentropic efficiency of compressor	%	80
Isentropic efficiency of gas expander	%	85
Isentropic efficiency of pump	%	75

4. Key performance indicators (KPIs)

Various key performance indicators are selected to evaluate the thermodynamic performance of the hybrid systems, such as specific power consumption (SPC) of the PMR, the CO_2 capture process, and the overall system. The SPC of the PMR is based on the electricity input to the PMR per unit mass of hydrogen produced. The SPC of the CCP is the net power consumption in the CCP per unit mass of CO_2 captured. The SPC of the hybrid system is estimated by the total net power consumption per unit mass of H_2 produced. The CCR of the CCP is the molar flow rate of the CO_2 captured per unit molar flow rate of CO_2 in the retentate gas. The system CCR is defined as the molar flow rate of CO_2 captured divided by the total molar flow rate of carbon in natural gas. Other KPIs such as hydrogen recovery rate (HRR) are as follows (it is worth mentioning that CH_4 conversion of the PMR is always kept high in this work):

$$HRR_{\mathrm{PMR}} = \frac{\dot{n}_{\mathrm{H_2,product}}}{\dot{n}_{\mathrm{H_2,PMR\ feed}} + \dot{n}_{\mathrm{H_2,generated\ in\ PMR}}} \tag{1}$$

$$HRR_{\mathrm{sys}} = \frac{\dot{n}_{\mathrm{H_2,product}}}{\dot{n}_{\mathrm{H_2,produced\ in\ pre-ref}} + \dot{n}_{\mathrm{H_2,produced\ in\ PMR}} + \dot{n}_{\mathrm{H_2,produced\ in\ WGS}}} \tag{2}$$

5. Results and discussion

The simulation results in Table 2 indicate that compared to the simplified hybrid concept, the process with off-gas recycle and a WGS reactor has a larger H_2 production capacity and a lower system SPC. Besides, the configuration with the off-gas recycle gives very high system HRR and CCR at around 99 %, verifying that this concept can produce H_2 with a low carbon intensity even when the PMR operating conditions are set for a relatively low HRR (91 %). As presented in Table 2, due to the recycle of the H_2 rich off-gas, the HRR and the hydrogen production rate of the PMR are improved compared to the simplified hybrid system. The recycled stream also makes the PMR feed richer in hydrogen, and it is advantageous to extract and compress H_2 in the membrane reformer, reducing its SPC. Besides, the WGS reactor effectively shifts CO to CO_2 in the retentate gas, increasing the CO_2 content of the feed to the liquefaction process and the efficiency of the CO_2 capture system (higher CCR and lower SPC).

The simplified hybrid concept has a low system carbon capture rate although the process has a similar CCR in the CCP compared to the hybrid process with off-gas recycle. This is because only a part of the natural gas supplied to the system is shifted to CO_2 in the PMR while the rest becomes CO, which is not captured through the liquefaction process. Thus, significant amounts of carbon are lost through the CO rich off-gas from the CO_2 capture process. However, the hybrid concept with off-gas recycle has a WGS reactor where almost all CO in the retentate gas is shifted to CO_2, thus allowing the liquefaction system to reduce the carbon loss via the vented off-gas.

It is worth noting that the heat from the PMR is more effectively utilized in the hybrid concept with off-gas recycle as it has a smaller amount of heat left from the PMR compared to the simplified hybrid system. Although the remaining of the PMR surplus heat is assumed to be used to produce electricity and supplied to the hybrid concepts with a 50 % conversion rate, the configuration with off-gas recycle will still have a lower system SPC (43.0 MJ/kg H_2) compared to the simplified scheme (45.6 MJ/kg H_2).

Table 2. Performance of the two hybrid concepts for clean hydrogen production with the PMR.

Parameter	Unit	Simplified hybrid	Modified hybrid
PMR H_2 production	t/d	475	560
PMR heat leftover	MW	29.38	13.00
SPC_{PMR}	MJ/kg H_2	46.39	42.19
HRR_{PMR}	%	91.06	93.99
$x_{CO2,CCP\ feed}$	(dry basis)	0.53	0.65
$x_{CO,CCP\ feed}$	(dry basis)	0.22	0.01
Captured CO_2	t/d	1965	3374
SPC_{CCP}	MJ/kg CO_2	0.45	0.30
CCR_{CCP}	%	83.44	89.27
HRR_{sys}	%	91.06	98.75
CCR_{sys}	%	57.80	99.30
SPC_{sys}	MJ/kg H_2	48.26	43.99

6. Conclusions

In this work, a novel hybrid concept is developed to produce H_2 from natural gas using an innovative proton membrane reformer followed by a liquefaction based CO_2 capture system. The hybrid concept with off-gas recycle and a WGS reactor effectively recovers H_2 produced in the PMR while capturing almost all CO_2 from the process even when the PMR is operated at relatively low H_2 recovery with less energy input. Thus, this hybrid scheme will be a promising option for H_2 production with a low carbon intensity. This process design can be further improved by optimal heat integration with the PMR surplus heat and the utilization of the vent stream as fuel.

Acknowledgements: This work was performed within the CLIMIT-KPN MACH-2 project (294629) with support from the NCCS Centre, performed under the Norwegian research program Centres for Environment-friendly Energy Research (FME). The authors acknowledge the following partners for their contributions: Aker Solutions, Ansaldo Energia, Baker Hughes, CoorsTek Membrane Sciences, EMGS, Equinor, Gassco, Krohne, Larvik Shipping, Lundin, Norcem, Norwegian Oil and Gas, Quad Geometrics, Total, Vår Energi, and the Research Council of Norway (257579).

References

D. Berstad, R. Anantharaman, P. Nekså, 2013, Low-temperature CO2 capture technologies – Applications and potential, International Journal of Refrigeration, 36, 5, 1403-1416.

T. Harkin, I. Filby, H. Sick, D. Manderson, R. Ashton, 2017, Development of a CO2 Specification for a CCS Hub Network, Energy Procedia, 114, 6708-6720.

D. Kim, D. Berstad, R. Anantharaman, J. Straus, T. Peters, T. Gundersen, T, 2020. Low Temperature Applications for CO2 Capture in Hydrogen Production, Computer Aided Chemical Engineering, 48, 445-450.

H. Malerød-Fjeld, D. Clark, I, Yuste-Tirados, R. Zanón, D. Catalán-Martinez, D. Beeaff, S.H. Morejudo, P.K. Vestre, T. Norby, R. Haugsrud, J.M. Serra, C. Kjølseth, 2017, Thermoelectrochemical production of compressed hydrogen from methane with near-zero energy loss, Nature Energy, 2, 12, 923-931.

M. Voldsund, K. Jordal, R. Anantharaman, 2016, Hydrogen production with CO2 capture, International Journal of Hydrogen Energy, 41, 9, 4969-4992.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50022-1

Analysis and design of integrated renewable energy and CO_2 capture, utilization, and storage systems for low-cost emissions reduction

Mohammad Lameh, Dhabia M. Al-Mohannadi, Patrick Linke*

Department of Chemical Engineering, Texas A&M University at Qatar, Education City, PO Box 23874, Doha, Qatar
patrick.linke@qatar.tamu.edu

Abstract

CO_2 capture, utilization, and storage (CCUS) as well as renewable energy (RE) technologies are key options for the decarbonization of economies. The high cost of such pathways makes it important to develop a strategic screening approach that yields the optimal implementation of CO_2 reduction pathways while ensuring the economic viability of such projects. This work proposes a Process Systems Engineering approach to develop minimum cost CO_2 reduction pathways. The approach implements a systematic analysis methodology to understand key decisions of the optimal design. After that, a detailed network portfolio can be obtained by solving a reduced optimization problem. The method is demonstrated in a case study which shows how the high-level analysis can be used to guide the detailed design of CO_2 reduction networks, resulting in an efficient systematic planning.

Keywords: cost-optimal CO_2 reduction, marginal abatement cost, economic analysis, optimization.

1. Introduction

Process Systems Engineering methods have been developed to optimize the planning of CO_2 emissions mitigation (Manan et al., 2017). The general process engineering approach consists of analysing the problem to develop high-level insights and targets based on which the designs of integrated systems are assessed (Klemes, 2013). The early applications of such approach were focused on developing pinch analysis methodologies for optimizing heat integration (Linnhoff et al., 1979). The minimum heating and cooling targets developed allowed the validation and the understanding of optimal designs of heat exchanger networks (Linnhoff & Hindmarsh, 1983). The problem of cost-optimal CO_2 reduction has been addressed through designing an integrated system considering all available CCUS and RE options that achieve the desired CO_2 emissions reduction at the lowest possible cost (Al-Mohannadi et al., 2020). The solution is obtained through implementing an optimization model which yields the integrated processing system with the minimum cost. However, such solutions are not usually easily understood, and they require further analysis and interpretation. Recently, a cost analysis methodology for CO_2 reduction pathways was developed based on the Marginal Abatement Cost (MAC) of the different considered options (Lameh et al., 2021). This methodology allows the development of low-cost CO_2 reduction solutions using basic high-level information about the reduction technologies, but it lacks the level of detail that the design optimization models have.

To our knowledge, none of the exiting studies show a comprehensive Process Systems Engineering approach with analysis and design methodologies that systematically identify optimal pathways for CO_2 reduction. This work addresses the gap by presenting a two-step approach to support the decisions of planners and policy makers to achieve optimal CO_2 reduction. In the first step, the analysis method uses technical and economic factors of the different possible CO_2 reduction pathways to develop quick insights into the minimum cost solutions based on high level overview of the defined problem. These insights would simplify the optimization so that a global optimal solution is achieved. In the second step, the integrated CCUS-RE network optimization is performed to design a detailed CO_2 reduction configuration through physical CO_2 emissions processing and storage, and through applying renewable energy technologies. The solution of the optimization problem can then be understood based on the high-level insights obtained from the analysis tool applied in the first step. A case study is presented to illustrate the application of the method.

2. Methods

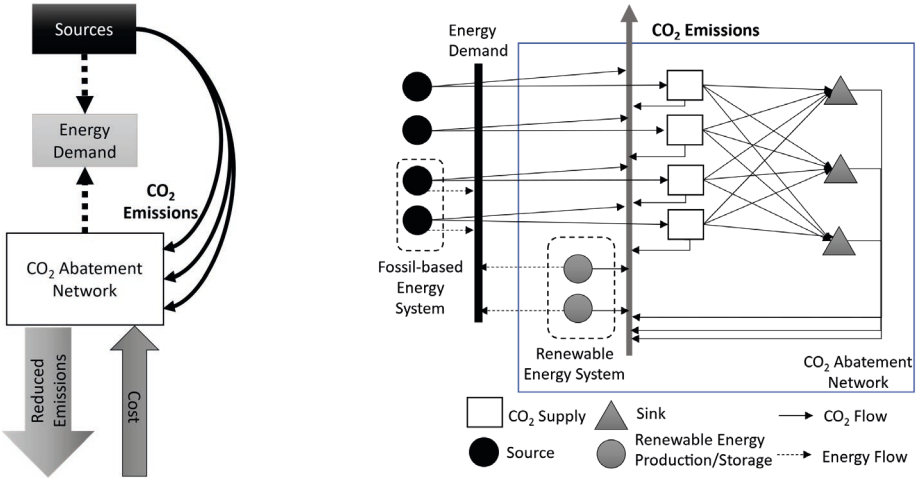
The aim of the proposed approach is to identify cost-optimal transitions to achieve a set target for CO_2 reduction. Different CO_2 emitting sources exist, among which are fossil-based energy production plants that cover a defined demand. The set target for CO_2 emissions reduction can be achieved by implementing a CO_2 abatement network which consists of CCUS and RE pathways. In the CCUS pathways, the emissions from the sources are captured and allocated to CO_2 sinks which can either store the CO_2 or utilize the emissions to produce value-added products. The RE pathways involve the implementation of renewable energy options to replace some of the existing fossil-based energy sources to cover the demand. The problem is addressed at two stages: analysis and design. At the first stage, CO_2 reduction analysis is conducted to determine the expected cost of the optimal CO_2 abatement network which achieves a set level of reduced emissions (Figure 1 (a)). This approach allows the identification of major insights corresponding to the total cost of CO_2 reduction through a simple illustrative procedure. The detailed design of the optimal network that achieves the CO_2 reduction target is addressed in the second stage in which the exact layout of the network with the flowrates and allocations is identified (Figure 1(b)). The analysis is conducted through developing the marginal abatement cost (MAC) curve considering the different available options (Lameh et al., 2021).

$$MAC_{CCUSij} = \frac{C_{si} - R_{dj}}{\eta_{dj} - \gamma_{si}} \tag{1}$$

$$MAC_{Eij} = \frac{C_{ej} - C_{exi}}{\varepsilon_{ei} - \varepsilon_{exj}} \tag{2}$$

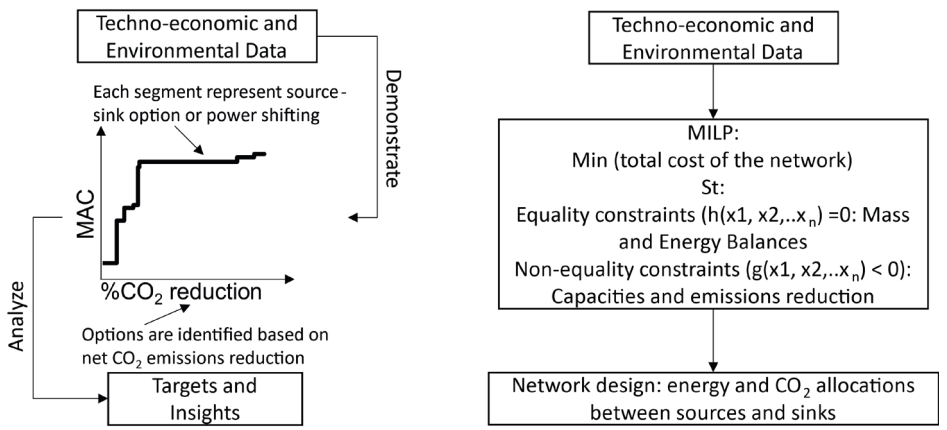
Each CCUS option is characterized by the cost (C_{si}) of CO_2 supply from each source (capture, compression, and transport), the profit (R_{dj}) generated by each sink, the secondary emissions associated with supplying CO_2 from the sources (γ_{si}), and CO_2 fixation efficiency of each sink (η_{dj}). The MAC for the CCUS options can be calculated as shown in equation (1). The RE options are considered as energy-shifting pathways which are characterized by the cost of the RE source (C_{ei}), the cost of the existing source that is phased out (C_{exi}), and their corresponding emissions levels (ε_{ei} and ε_{exi}).

The MAC for the RE options can be determined as shown in equation (2). The CO_2 reduction potential for each option is determined based on prioritizing the cheapest pathways. The different options in the CO_2 abatement network are demonstrated on a minimum MAC (mini-MAC) profile from which the cost of a set level of CO_2 reduction can be determined.



(a) (b)

Figure 1 The CO_2 abatement network as considered through system analysis (a) and network design (a)



(a) (b)

Figure 2 The framework applied in the analysis (a) and design (b) of the CO_2 abatement network

The network optimization problem is decomposed in a two-step approach (Al-Mohannadi et al., 2020) with the main network synthesis problem being formulated as a mixed integer linear program (MILP). The MILP determines the overall design of the CO_2 reduction portfolio which corresponds to the minimum cost for the set level of CO_2 reduction. The model is formulated by setting the equations for the energy and mass balances and the costs for the different components in the system (CO_2 capture,

compression, piping, sinks, and RES). The optimization problem is defined by setting an objective function (minimizing the total cost of the system), and the constraints that ensure that the capacities of the CO_2 abatement technologies are not exceeded, and the set CO_2 emissions reduction is achieved. Figure 2 describes the methodologies followed in conducting the analysis and design of the CO_2 reduction network.

3. Case Study

The system studied in this work consists of four major CO_2 emitting sources which emissions flowrates is estimated based on the energy and industrial sectors existing in Qatar (Alfadala & El-Halwagi, 2017). Table 1 shows the data used to characterize each of the considered sources. The capture costs and secondary emissions are estimated based on Metz et al. (2005) and von der Assen et al. (2016). The cost of transportation is based on linearized compression, pumping, and piping cost models (Al-Mohannadi et al., 2020). Four potential CO_2 sinks are considered for implementation as CO_2 utilization and storage technologies (Table 2). The data is estimated based on existing technoeconomic studies (Hepburn et al., 2019) for enhanced oil recovery (EOR) , CO_2 storage (GCCSI, 2011), chemicals (Pérez-Fortes et al., 2016), and fuels (Tremel et al., 2015). For the renewable energy contribution, the demand for electric power can be covered either by an existing natural gas power plant (NG PP) or by introducing solar photovoltaic system (Solar PV). The solar PV can cover up to 13% of the electric power demand without requiring energy storage. The levelized cost of electricity for the solar PV is assumed to be 0.017 $/kWh (BELLINI, 2020). The cost of operating the NG PP was neglected (assuming very low cost of natural gas).

Table 1 The data for the sources

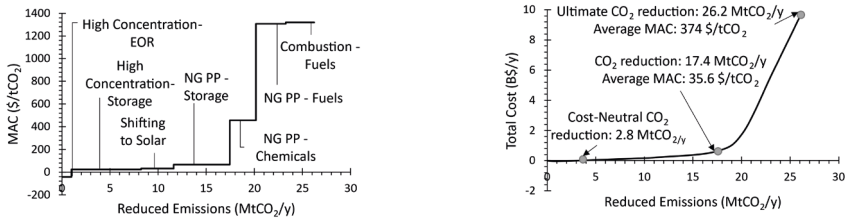
Source	CO_2 Produced ($MtCO_2/y$)	Capture Cost ($/$tCO_2$)	Transportation Cost ($/$tCO_2$)	Secondary emissions ($tCO_2/tCO_{2\text{-captured}}$)
High Concentration	8.32	0.00	3.00	0.01
Combustion	48.65	31.12	3.48	0.24
Cement	1.99	56.85	3.15	0.24
NG PP	25.88	27.33	3.27	0.24

Table 2 The data for the sinks

Sink	Capacity ($MtCO_2/y$)	CO_2 Breakeven Cost ($/$tCO_2$)	CO_2 Fixation Efficiency ($tCO_2/tCO_{2\text{-utilized}}$)
EOR	1	45	1
Storage	15	-20	1
Chemicals	4	-280	0.92
Fuels	17	-440	0.6

The collected data was used to analyze the cost of economic CO_2 reduction by generating the mini-MAC profile of the considered options (Figure 3 (a)). The mini-MAC profile identifies the promising pathways for economic CO_2 reduction: these are the pathways represented by the segments forming the MAC curve. The total cost of CO_2 reduction can be determined by integrating the mini-MAC profile (area under the curve) as shown in Figure 3 (b). This outcome can guide the network design by showing the expected total cost for different levels of CO_2 reduction. Each single point on the

total cost profile corresponds to an optimal network design that can achieve the CO_2 reduction target with minimum cost. Instead of generating random designs through an exhaustive procedure of running the optimization model multiple times, key targets from the total cost profile can be determined. The high-level analysis shows that the maximum CO_2 reduction potential that can be achieved is 26.2 $MtCO_2/y$, which is around 31% of the considered emissions (84.8 $MtCO_2/y$). This is due to the limited capacity of the considered options (the capacity of the sinks is 37 $MtCO_2/y$), and to the secondary emissions associated with CO_2 supply and CO_2 sinks. The analysis identifies a cost neutral CO_2 reduction with a flowrate of 2.8 $MtCO_2/y$ (11% of the maximum reduction potential). The mini-MAC profile shows a high cost for the options that require CO_2 supply from NG PP and combustion for utilization in chemicals and fuels production. Implementing a hybrid network consisting of renewable energy (shifting to solar) and CO_2 integration (capturing from high concentration sources and NG PP and allocating the emissions to EOR and storage) can reduce the emissions by 17.4 $MtCO_2/y$ (66% of the maximum reduction potential) at a relatively low cost (35.6 $\$/tCO_2$). Beyond that abatement level, the expensive pathways are required, and the average MAC would rise to 374 $\$/tCO_2$ at the maximum achievable level of CO_2 reduction.



(a) (b)

Figure 3 The mini-MAC profile (a) and the estimated cost of optimal CO_2 reduction (b)

To verify the results of the analysis, the design is performed for: a network that achieves a cost-neutral CO_2 reduction, a network corresponding to the maximum CO_2 reduction before the aggressive rise in the cost, and a network with the ultimate CO_2 reduction that can be achieved with the considered options. The CO_2 reduction levels for the three different targeted networks were determined from the total cost profile, and they were used as the CO_2 reduction constraints in the described optimization model. The optimization model minimizes the cost by determining the optimal CO_2 allocation between sources and sinks and the energy contribution of each power option. The design results for the three targeted systems are shown in Figure 4. The costs of the detailed designs for the three cases validated the results obtained from the analysis, with a slight marginal error (up to 2%). Hence, the proposed approach provides a systematic methodology for identifying cost-optimal CO_2 reduction by implementing simple high-level analysis to determine the expected costs, and to plan the designs of the optimal networks, and validate their outcomes.

4. Conclusions

This work presented a comprehensive Process Systems Engineering approach for planning and designing cost-optimal CO_2 abatement networks considering CCUS and RE options through analysis and design. The application of the method to a case study showed how the analysis of the system can be used to validate the results obtained from the design procedure and to understand the optimization solutions in the context of

achieving affordable CO_2 emissions reduction. Future work will analyze the errors and deviations in both approaches and their impact under various uncertainties.

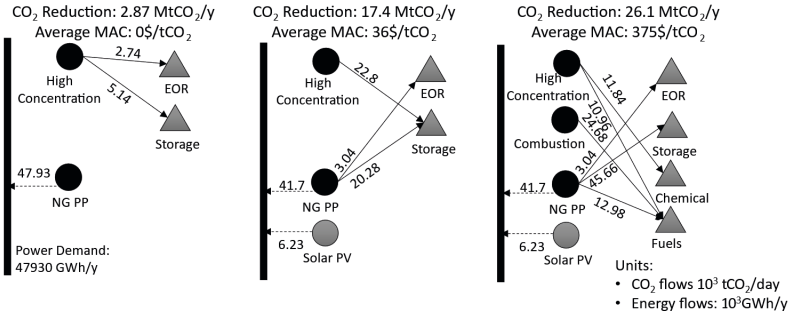


Figure 4 Three designs for the CO_2 abatement network with different levels of CO_2 reduction

References

Al-Mohannadi, D. M., Kwak, G., & Linke, P. (2020). Identification of optimal transitions towards climate footprint reduction targets using a linear multi-period carbon integration approach. Computers & Chemical Engineering, 140, 106907.

Al-Mohannadi, D. M., Kwak, G., & Linke, P. (2020). Identification of optimal transitions towards climate footprint reduction targets using a linear multi-period carbon integration approach. Computers & chemical engineering, 106907.

Alfadala, H. E., & El-Halwagi, M. M. (2017). Qatar's chemical industry: Monetizing natural gas. Chemical Engineering Progress, 113, 38-41.

BELLINI, E. (2020). Qatar's 800 MW tender draws world record solar power price of $0.01567/kWh. In (Vol. 2021): PV Magazine.

GCCSI. (2011). The costs of CO2 storage: post-demonstration CCS in the EU, Global CCS Institute. In: Global CCS Institute.

Hepburn, C., Adlen, E., Beddington, J., Carter, E. A., Fuss, S., Mac Dowell, N., Minx, J. C., Smith, P., & Williams, C. K. (2019). The technological and economic prospects for CO 2 utilization and removal. Nature, 575, 87-97.

Klemes, J. J. (2013). Handbook of process integration (PI): minimisation of energy and water use, waste and emissions: Elsevier.

Lameh, M., Al-Mohannadi, D. M., & Linke, P. (2021). Minimum marginal abatement cost curves (Mini-MAC) for CO 2 emissions reduction planning. Clean Technologies and Environmental Policy, 1-17.

Linnhoff, B., & Hindmarsh, E. (1983). The pinch design method for heat exchanger networks. Chemical Engineering Science, 38, 745-763.

Linnhoff, B., Mason, D. R., & Wardle, I. (1979). Understanding heat exchanger networks. Computers & Chemical Engineering, 3, 295-302.

Manan, Z. A., Nawi, W. N. R. M., Alwi, S. R. W., & Klemeš, J. J. (2017). Advances in Process Integration research for CO_2 emission reduction–A review. Journal of cleaner production, 167, 1-13.

Metz, B., Davidson, O., & De Coninck, H. (2005). Carbon dioxide capture and storage: special report of the intergovernmental panel on climate change: Cambridge University Press.

Pérez-Fortes, M., Schöneberger, J. C., Boulamanti, A., & Tzimas, E. (2016). Methanol synthesis using captured CO2 as raw material: Techno-economic and environmental assessment. Applied Energy, 161, 718-732.

Tremel, A., Wasserscheid, P., Baldauf, M., & Hammer, T. (2015). Techno-economic analysis for the synthesis of liquid and gaseous fuels based on hydrogen production via electrolysis. International Journal of Hydrogen Energy, 40, 11457-11464.

von der Assen, N., Müller, L. J., Steingrube, A., Voll, P., & Bardow, A. (2016). Selecting CO2 Sources for CO2 Utilization by Environmental-Merit-Order Curves. Environmental Science & Technology, 50, 1093-1101.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50023-3

Techno-economic-environmental assessment for optimal utilisation of CO_2 in the Fischer-Tropsch gas-to-liquid process

Ali Attiq Al-Yaeeshi, Ahmed AlNouss and Tareq Al-Ansari*

College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar
*talansari@hbku.edu.qa

Abstract

Considering the un-declining emissions of CO_2, which is a major contributor to global warming, carbon capture and utilisation (CCU) has been promoted as a potential CO_2 reduction pathway, generating economic benefits and reduced environmental burdens. The integration of CCU with power plants and chemical industries drives the potential of adapting a CO_2 capture and utilisation scheme. Chemical synthesis such as gas-to-liquids (GTL) process using the Fischer-Tropsch technology is a promising pathway in this configuration. The objective of this study is to assess the techno-economic-environmental viability of maximising the production of wax, diesel, gasoline and LPG in an FT-GTL plant, while optimizing the utlisation of different variables such as steam, oxygen, CO_2, and the syngas recycle to purge ratio. The effect of reforming techniques and syngas recycle ratio on the production capacity are analysed upon supplementing the process with additional CO_2 at a range of 1000-2000 t/d. The methodology is based on the maximum production of syngas in the reforming units, which include steam-based methane reforming (SMR) and oxygen/steam-based auto-thermal reforming (ATR). Aspen HYSYS is used to model the GTL production flowsheets. The results demonstrate a significant improvement in the total refined products capacity for all scenarios based on variable function of raw material flow rate of CO_2, steam, oxygen and split ratio of syngas to the purge. The sensitivity analyses demonstrate the feasibility of the ATR and SMR options to provide significant enhancement when integrated with CO_2. The total refined product of hydrocarbons increase significantly when the decision variables are optimized.

Keywords: SMR, ATR, CO_2 Utilisation, GTL, CAPEX, OPEX, Optimisation

1. Introduction

Greenhouse gas (GHG) emissions are one of the most considerable environmental concerns of the recent era and are a leading cause for global warming, where CO_2 is a major contributor (IEA, 2018). The concentration of CO_2 in the atmosphere can be reduced through applying carbon capture and utilisation (CCU) processes. The Gas-to-Liquid (GTL) process is one example of an application that can accept CO_2 as a feedstock to enhance its product output (Al-Yaeeshi et al., 2020). Incidentally, McGregor (2019) stated that the CO_2 can replace the CO product in Fischer–Tropsch synthesis within the GTL process. Although, there are challenges for the introduction of CO_2 as feedstock in the FTS process, there are economic and environmental benefits in utilising the otherwise waste CO_2.

The main function of the GTL process is the conversion of natural gas (NG) into liquid refined products using the intermediate carbon monoxide (CO) and hydrogen (H_2) rich

syngas. The long chain hydrocarbon products from the FT reactor include wax, diesel, gasoline and LPG at a reduced aromatic and sulphur content, thereby enhancing environmental compliance. Therefore, the synthetic fuels are considered a relatively more environmentally friendly array of products (Shell, 2019). The GTL production line comprises of five major units: pre-reforming, reforming, FT synthesis, product upgrading and fractionation. It begins by converting the natural gas into synthesis gas via various possible reforming reactions. The long chain hydrocarbons are then synthesised in the FT reactor, and subsequently treated in the upgrading section using H_2. Finally, in the fractionation column the refined products are separated. The economics of the GTL plant is high due to the costs of the FTS, and the efficiency required to produce a high stability ratio of H_2 and CO (syngas) (Al-Sobhi et al., 2021).

The CO_2 can be introduced as a feedstock or as recycled stream to influence the chemical equilibrium in the reforming unit and enhance syngas quality. Accordingly, the key parameters required in an optimisation problem include; reactor model design, operating conditions, and the total feed of CO_2 quantity, where by each component directly contributes to the enhancement of the product and the syngas H_2:CO ratio. Ekwueme et al. (2019) assessed the economics of a GTL plant considering an autothermal reforming model (ATR) and a steam/CO_2 reforming model, demonstrating positive economics of GTL process relative to other gas conversion technologies. Moreover, the steam/CO_2 reforming model is better performing from an economic perspective than the ATR in a small scale plant. Marchese et al. (2021) assessed the economic performance of direct air capture to the FT model, and maximised CO_2 conversion into synthetic chemicals, with a focus on wax. Furthermore, the recirculation of the FT off-gas was studied to enhance the performance, demonstrating a high system efficiency with a maximum carbon dioxide conversion at approximately 68.3 %.

CO_2 utilisation within GTL process has been studied by Al-Yaeeshi et al. (2019) and Al-Yaeeshi et al. (2020) to evaluate the efficiency of integrating the CO_2 into the steam methane reformer (SMR) and Auto-Thermal reformer (ATR). With the objective of maximising the production of wax, diesel, gasoline and LPG, this study analyses from a techno-economic-environmental perspective, the effect of reforming techniques, steam and oxygen demands and syngas recycle to purge ratio on the production capacity upon supplementing the FT-GTL process with CO_2 at the range of 1000-2000 t/d. A model is developed to assess different ATR and SMR reforming techniques within the GTL process. The CO_2 sink considered in this study is the Oryx GTL plant located in state of Qatar, which is configured with an ATR reformer with a natural gas feed of 330,000 cubic feet per day (QP, 2018). The feedstock consists of natural gas and steam with oxygen in the case of ATR and steam only in the case of SMR. The oxygen enters a Gibbs reactor, where natural gas is reformed to mainly CO and H_2. The purification process of syngas occurs prior to the FT unit to ensure high production of hydrocarbon molecules. Further purification is applied for the effluent from the FT unit to separate water and reprocess the unreacted CO and H_2. Subsequently, the hydrocarbon flow is sent to the upgrading unit, in which hydrogen is used to crack the longer-chain carbon molecules into smaller-chain hydrocarbons. Finally, the hydrocarbons are fractionated into wax, diesel, gasoline and LPG while the remaining stream is recycled to hydrocracking section.

2. Methodology

This study introduces a simulation flowsheet of the FT-GTL process integrated with CO_2 as a feedstock to maximise the production of wax, diesel, gasoline and LPG. It assumes the raw feed CO_2 is pure at the required operating parameters of GTL plant. The model is developed based on the raw data and ranges listed in Table 1 by using Aspen HYSYS-V9.

Table 1. Model raw data

Parameter	**NG Feed**	**CO_2**	**O_2 (ATR)**	**Steam (ATR)**	**Steam (SMR)**
Flow (T/d)	1.54 x 10^4	1-2 x 10^3	1.5-2 x 10^4	2-5 x 10^3	1-9 x 10^4
T (C)	25	150	144	500	500
P (bar)	1	25	25	25	25
	Hydrocracking	**FT**	**Reformer**	**Fractionator**	**Split Ratio**
T (C)	345	250	1050	Top P: 1 bar	(0.7-1)%
P (bar)	80	25	23	Bottom P :1.5 bar	

Various scenarios are studied depending on the GTL plant capacity, reformer type and CO_2 feedstock rate. The plant capacity is designed to process 15372 ton/d of NG feedstock to produce 34000 bbl/day of liquid hydrocarbons through dual trains. The molar oxygen to carbon ratio (O_2/C) is 0.6, and steam to carbon ratio (S/C) is 0.4 in the ATR reformer base scenario. In the case of the SMR reformer, the steam to carbon ratio is 3. The feed rate range of CO_2 for both reformers is varied between 1000-2000 ton/d to demonstrate the enhancement profile in the production of liquid hydrocarbons. Aspen HYSYS is used to assess the variations in each operating parameter. Results of the sensitivity analyses are used to construct regression models relating total refined products and H_2:CO ratio in syngas to changes in each operating parameter. The functions of the total refined products and H_2:CO ratio in syngas for each operating parameter are employed, where the weighted average of these functions is calculated to characterise the objective functions of the proposed optimisation model. The singular objective function of total refined products and H_2:CO ratio in syngas is then maximised to produce the optimal decision variables for each objective.

Variables:

$\dot{m}_{Total}$: Total hydrocarbons production rate (t/d)
$\dot{m}_{LPG}$: LPG production rate (t/d)
$\dot{m}_{Gasoline}$: Gasoline production rate (t/d)
$\dot{m}_{Diesel}$: Diesel production rate (t/d)
$\dot{m}_{Wax}$: Wax production rate (t/d)
$\frac{H_2}{CO}$: Molar ratio of hydrogen to carbon monoxide (°C)

Decision variables:

$\dot{m}_{steam}$: Steam flowrate (t/d)
$\dot{m}_{Oxygen}$: Oxygen flowrate (t/d)
$\dot{m}_{CO_2}$: CO_2 flowrate (t/d)
SR: split ratio of syngas to the purge (%)

Objective function:

Maximise: $\dot{m}_{Total} = \dot{m}_{LPG} + \dot{m}_{Gasoline} + \dot{m}_{Diesel} + \dot{m}_{Wax}$ (1)

Maximise: $\frac{H_2}{CO}$ (2)

Constraints: *As illustrated in Table 1*

3. Results

The output of sensitivity analyses demonstrates a variation in the rates of refined products with the changes in CO_2, O_2, steam, and recycle ratio. Figure 1 illustrates the effect of steam variation for ATR and SMR cases. The total refined products decreases with the increase in steam rate indicating optimum values at 21,500 and 2500 t/d for SMR and ATR, respectively while the H_2:CO ratio increases.

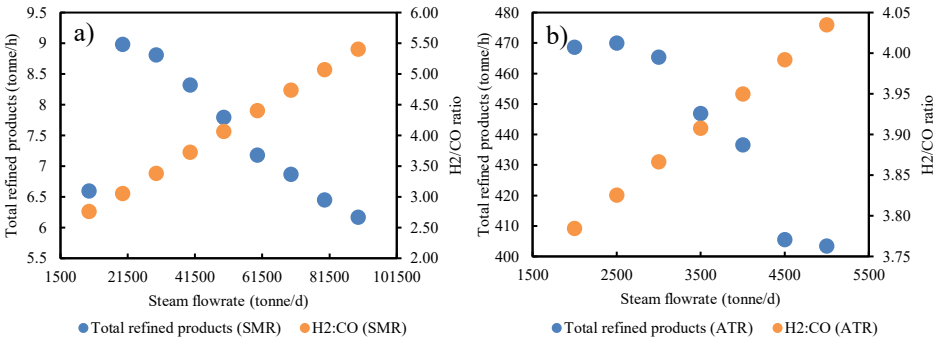


Figure 1: Effect of steam on total refined products and H_2:CO ratio for (a) SMR and (b) ATR.

The effect of CO_2 injection rate illustrated in Figure 2, demonstrates that the total refined products increases with the increase in CO_2 injection rate indicating an optimum value at 1900 t/d for ATR and continuous increasing trend for SMR, while the H_2:CO ratio decreases indicating more generation of CO through the equilibrium shift reaction.

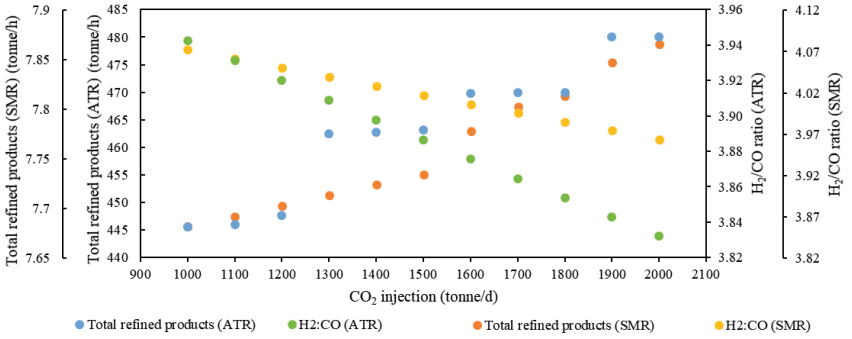


Figure 2: Effect of CO_2 injection on total refined products and H_2:CO ratio for SMR and ATR.

The variation of the split ratio of syngas recycle to FT reactor and purge stream illustrated in Figure 3a indicates that the total refined products are maximised at around 0.95 ratio with no pressure build up issues. The variation on oxygen rate (Figure 3b) applicable to the ATR indicates an optimum value at 18,000 t/d.

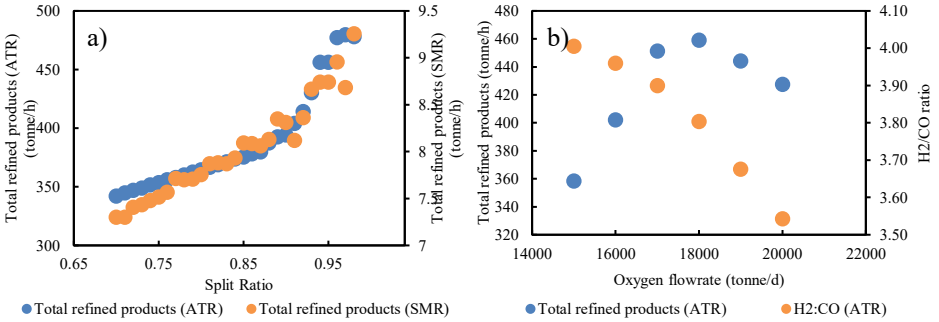


Figure 3: Trends of (a) split ratio variation for SMR and ATR and (b) oxygen flowrate for ATR.

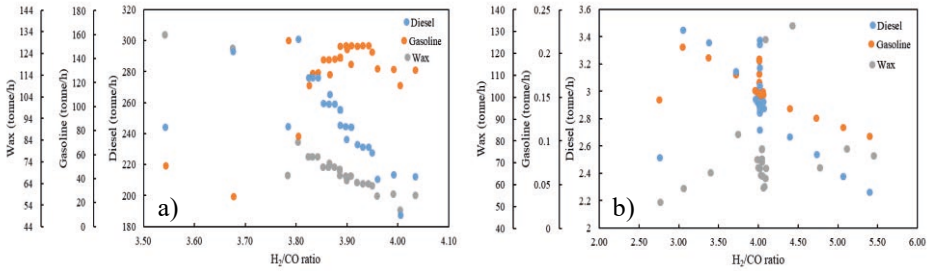


Figure 4: Optimal CO_2 source-sink allocation for (a) scenario 3 and (b) scenario 4.

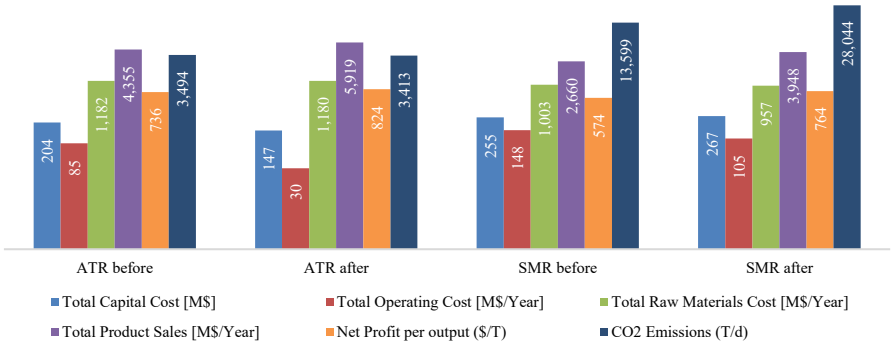


Figure 5: Results of the techno-economic-environmental assessment.

Plotting the two competing functions together as in Figure 4 indicates a high production for gasoline at higher H_2:CO ratio of approximately 3.9-4 compared to 3.8 for diesel and wax in the ATR case. Whereas, the SMR case revealed different schemes where the wax is maximised at a higher H_2:CO ratio of approximately 3.9-4.5 compared to 3 for diesel and gasoline. In summary, the optimization problem for the ATR case indicates an oxygen requirement rate of 18000 t/d, steam rate of 2500 t/d, CO_2 injection rate of 1900 t/d, a 21 bar FT pressure and a 0.97 split ratio. Whereas, in the case of the SMR, the optimization problem reveals a steam rate requirement of 20000 t/d, continuously increasing CO_2 injection trend, a 25 bar FT pressure, and a 0.98 split ratio. Furthermore, in both the ATR and SMR, the steam rate has a significant impact on the system. The techno-economic-environmental assessment demonstrates an improvement in the net profit per products for both ATR and SMR cases, and a reduction in the environmental emissions for the ATR case as illustrated in Figure 5 compared to the study conducted by Al-Yaeeshi et al.

(2020). This is associated with a slight increase in the capital cost due to the increased capacity and a reduction in operating costs, thus indicating a reduction in the requirement of utilities. This provides a substantial assessment to enhance the entire process efficiency and optimise the total capacity of refined products.

4. Conclusions

The assessment in this study is applied to two different GTL reforming technologies, which are the ATR and SMR. The integration of captured CO_2 with the FT-GTL process demonstrates a significant enhancement in the production of refined products. The results of the sensitivity analysis demonstrate that the wide range of variables impact the total refined hydrocarbon products, namely LPG, Gasoline, Diesel, and Wax. The variables applied in the assessment are steam, CO_2, oxygen, and split ratio of syngas recycle to FT reactor and purge stream for the ATR and SMR cases. The output in both cases detail optimal operating values that result in the significant enhancement in the final hydrocarbon production. The highest impact in both cases is with steam injection. Future studies should include the computation of the energy output / input, impact of CO_2 price on the techno-economic recycle of hydrogen and economic viability of hydrogen integration from different sources.

References

S.A. Al-Sobhi, A. AlNouss, and M. Alhamad, 2021, Techno-economic and environmental assessment of Gasoline produced from GTL and MTG processes, Computer Aided Chemical Engineering, 50, 1827-32.

A.A. Al-Yaeeshi, A. AlNouss, G. McKay, and T. Al-Ansari, 2019, A Model based analysis in applying Anderson–Schulz–Flory (ASF) equation with CO2 Utilisation on the Fischer Tropsch Gas-to-liquid Process, Computer Aided Chemical Engineering, 46, 397-402.

A.A. Al-Yaeeshi, A. AlNouss, G. McKay, and T. Al-Ansari, 2020, A simulation study on the effect of CO2 injection on the performance of the GTL process, Computers & Chemical Engineering, 136, 106768.

S. Ekwueme, I. Nkemakolam Chinedu, U. Julian, A. Kerunwa, N. Ohia, J. Princewill, and O. Emeka, 2019, Economics of Gas-to-Liquids (GTL) Plants, Petroleum Science and Engineering, 3, 85-93.

IEA, 2018, Global Energy & CO2 Status Report 2018 - The latest trends in energy and emissions in 2018, https://iea.blob.core.windows.net/assets/23f9eb39-7493-4722-aced-61433cbffe10/Global_Energy_and_CO2_Status_Report_2018.pdf

M. Marchese, G. Buffo, M. Santarelli, and A. Lanzini, 2021, CO2 from direct air capture as carbon feedstock for Fischer-Tropsch chemicals and fuels: Energy and economic analysis, Journal of CO2 Utilization, 46, 101487.

J. McGregor, 2019, 21. Fischer–Tropsch synthesis using CO2, Volume 2 Transformations, 413-32.

QP, 2018, Subsidiaries and join venture details Accessed 14/11/2018, https://www.qp.com.qa/en/QPActivities/Pages/SubsidiariesAndJointVenturesDetails.aspx?aid=3

Shell, 2019, Shell qatar project and sites, https://www.shell.com.qa/en_qa/about-us/projects-and-sites.html

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50024-5

Machine Learning-based Hybrid Process Design for the Recovery of Ionic Liquids

Yuqiu Chen*, Xiaodong Liang, Georgios M. Kontogeorgis

Department of Chemical and Biochemical Engineering, Technical University of Denmark, Lyngby, DK-2800, Denmark
yuqch@kt.dtu.dk

Abstract

Recycling ionic liquids (ILs) from dilute aqueous solutions is essential for their applications in both labs and industries. In this work, an efficient hybrid process scheme that combines aqueous two-phase extraction (ATPE) and distillation operating at their highest efficiencies is proposed for the recovery of ILs from dilute aqueous solutions. To find high performance salting-out agents for ATPE, an optimal IL-based aqueous biphasic systems (ABS) design method is employed. In this optimal design method, a machine learning (ML)-based model, i.e., artificial neural network (ANN)-group contribution (GC) model, is applied to predict the phase equilibrium behaviours of IL-based ABS. As a proof of the concept, results of the recovery of two hydrophilic ILs from their aqueous solutions are presented.

Keywords: IL recovery, Hybrid process scheme, ATPE, Machine learning, ABS.

1. Introduction

Ionic liquids (ILs) as innovative fluids have received wide attention in both academia and industries due to their unique properties such as negligible vapor pressure, non-flammability, wide electrochemical windows, excellent catalytic activities. Great efforts have been made to facilitate their applications in industry. However, currently there are little industrial processes employing ILs mostly because their relatively high costs in comparison with conventional solvents and our limited understanding of their environmental impacts. For example, large volumes of dilute aqueous IL solutions will be produced during the dissolution and regeneration of cellulose when using ILs as solvents. The disposal of these aqueous IL solutions will directly cause the loss of these high value solvents and this may even result in severe environment issues due to the toxicity and degradation of the disposed ILs. Both economic and environmental concerns of using ILs can be offset to some extent if they are efficiently recycled.

To date, various technologies including distillation, extraction, adsorption, membrane separation, aqueous two-phase extraction (ATPE), crystallization, electrodialysis and external force field separation have been proposed for the recovery/recycling of ILs after their application (Zhou et al., 2018). Each separation technology described above has its own advantages and shortcomings. Currently, distillation and extraction are two of the most widely studied separation approaches for the recovery/recycling of ILs. Due to the fact that most ILs have very low volatility, distillation is usually used for the recovery of ILs from volatile substances, while extraction is preferred in the case of separating ILs from non-volatile or thermally sensitive components. When recovering ILs from dilute aqueous solutions, however, distillation method has an extremely low

thermal efficiency since large volumes of water need to be evaporated, while extraction approach demands a large amount of solvent for achieving a high recovery yield. On the other hand, ATPE that based on the formation of the aqueous biphasic systems (ABS), provides an alternative pathway for recovering ILs from dilute aqueous solutions. This is due to the fact that it allows the ILs to be efficiently concentrated or recovered in the IL-rich phase with the addition of a small amount of salting-out agent (Ventura et al., 2017). However, further purification process such as distillation is generally still required after aqueous two-phase extraction due to the fact that the IL purity in the IL-rich phase is not high enough. On the other hand, hybrid process schemes, which combine processing units operating at their highest efficiencies to perform one or more process tasks, are being considered as promising innovative and sustainable processing options (Chen et al., 2018). With this concept, a hybrid process scheme combining ATPE and distillation method is proposed for the recovery of ILs from their dilute aqueous solutions. In this hybrid process scheme, salting-out agents with high ABS forming ability are identified for ATPE through an optimal design method integrating a machine learning (ML)-based model into the computer-aided design technique. Results of two case studies are presented to highlight the hybrid process design method proposed in this work.

2. Design method

2.1. ANN-GC model

The ability to predict phase equilibrium behaviours of IL-based ABS is essential for its early design. However, thermodynamic models that can provide such predictions are still not available for these aqueous systems due to their high complexity. Fortunately, a machine learning-based nonlinear model proposed in our recent work provides the possibility of describing IL-based ABS (Chen et al., 2021a). This model combines the artificial neural network (ANN) algorithm and the group contribution (GC) method. Together with the system's' temperature and the mass fraction of IL, 34 IL functional groups and 37 salting-out agent functional groups are used as inputs (size of 73 x 1) in this ANN-GC model. The input layer reads the structure information of IL-ABS and then the hidden layer transfers and delivers this input information to the output layer where the phase composition of IL-based ABS is quantified, as shown in Figure 1. A combination of tansig transfer function (Eq. (1)) in the hidden layer and purelin transfer function ((Eq. (2)) in the output layer was applied.

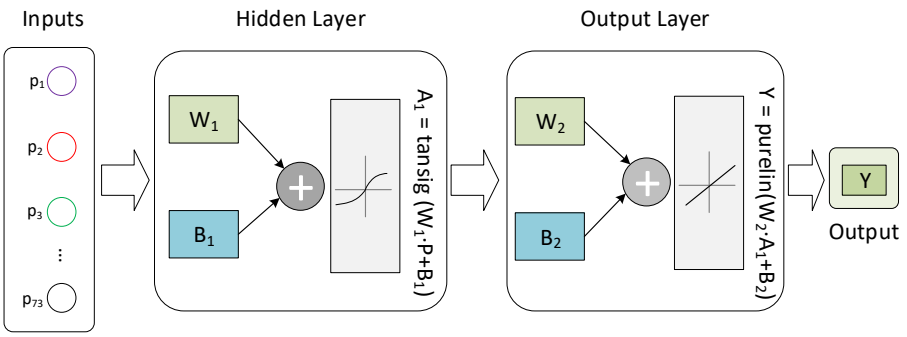


Figure 1. Structure of the three-layer artificial neural network (ANN) with an input vector size of 73 x 1.

$$f_1(x) = \frac{2}{1+e^{-2x}} - 1 \tag{1}$$

$$f_2(x) = x \tag{2}$$

To train and test this ANN-GC model, 17,449 experimental data points covering 171 IL-based ABS at different temperatures (278.15K-343.15K) from numerous literatures were collected. This ANN-GC model gives a mean absolute error (MAE) between the experimental and model-calculated mass fraction of salting-out agent of 0.0175 and a coefficient of determination (R^2) of 0.9316 for the 13,789 training data points, and for the 3,660 test data points they are 0.0177 and 0.9195, respectively. These results show that this ANN-GC model can well describe the IL-based ABS.

2.2. Optimal salting-out agent design

The separation performance of an ATPE largely depend on the ABS forming ability of the used salting-out agents. The formation and stability of IL-based ABS is not only dependent on the structures of ILs such as cation types, lengths of alkyl chain and the anions, they are also highly associated with the type of salting-out agents.36 ABS with different ILs and salting-out agents at different temperature generally present different phase behaviors, it is challenging to find optimal ABS for the recovery of various ILs. Due to the number of potential IL-ABS being so large, it would be time consuming and expensive to use the trial-and-error approach to search for the optimal ABS. On the other hand, the optimal design of compounds/systems through manipulating properties at the molecular level is often the key to considerable scientific advances and improved process systems performance (Alshehri et al., 2020). For IL-based ABS , the optimal design method that integrates the ANN-GC model into the computer-aided design technique proposed in our most recent work is ideally suited as salting-out agents with high ABS forming ability can be rapidly and reliably identified for different IL aqueous solutions (Chen et al., 2021b).

When tailoring an ABS for the recovery of ILs from aqueous solutions, the IL should be as much as possible to be concentrated in the IL-rich aqueous phase, while the salting-out agent and water should be at the other phase. As we know, the closer to the axis origin a binodal curve is, the greater is the ability of a salting-out agent to phase split, the tie-line length (*TLL*) and slope of the tie-lines (S_{TL}) are able to verify trends in the partition coefficients or recovery efficiencies of ABS.

$$TLL = \sqrt{\left(x_S^T - x_S^B\right)^2 + \left(x_{IL}^T - x_{IL}^B\right)^2} \tag{3}$$

$$S_{TL} = \frac{x_{IL}^T - x_{IL}^B}{x_S^T - x_S^B} \tag{4}$$

where the supercrits *T* and *B* designate the top phase (IL-rich phase) and the bottom phase (salt-rich phase), respectively, while the subscripts *S* and *IL* denote the mass fraction *x* of the salting-out agent and of the IL.

With the use of an objective function combining *TLL* and S_{TL}, the optimal design of IL-based ABS is formulated as a MINLP optimization problem mathematically descried by Eq. (5). In IL-based ABS, the specific IL is denoted by a vector $H^{IL} = [H_1^{IL}, H_2^{IL} \ldots H_{34}^{IL}]$. The first 5 elements $H_{1\text{-}5}^{IL}$ are integer variables describing the number of cation substituents. and the rest elements $H_{6\text{-}34}^{IL}$ are binary variables denoting the existence of cations and anions. On the other hand, each generated salting-out agent is represented by a vector $y = [y_1, y_2 \ldots y_{37}]$. The first 26 elements $y_{1\text{-}26}$ are binary variables describing the existence of salt anions, carbohydrates, amino acids. The other elements $y_{27\text{-}37}$ are integer variables denoting the number of salt cations. The best performance salting-out agent and its inputs can be determined for specific IL aqueous solutions (e.g., specific temperature T^{as}, IL structure H^{IL} and IL mass fraction in IL-water mixture G^{as}) by maximizing f $(z, y, T^{as}, H^{IL}, G^{as})$ that subjects to a series of constraints on salting-out agent structure, mass balance and phase equilibria.

$$\max_{z,y} TLL \cdot (-S_{TL})^r = f\left(z, y, T^{as}, H^{IL}, G^{as}\right)$$

$$\begin{aligned} s.t. \quad & \text{salting-out agent structural constraints} \\ & \text{mass balance constraints} \\ & \text{phase equilibria constraints} \end{aligned} \tag{5}$$

where vector *z* represents a continuous variable describing the ratio of added salting-out agent to the original IL-water mixture and *r* is an adjustable parameter describing the degree of influence from *TLL* and S_{TL}.

2.3. Hybrid recovery process scheme

The novelty of a hybrid process scheme is that each involved processing unit can operate at their highest efficiencies. The result is same task performed at much less energy inputs and/or lower cost. As mentioned above, distillation method has an extremely low thermal efficiency for recovering ILs from their dilute aqueous solutions, while ATPE approach cannot meet the final product specification. In such a case, hybrid process scheme combining ATPE and distillation is ideally suited as most water can be easily removed by adding a certain amount of salting-out agents and the rest of water can be distilled with a low energy input, as shown in Figure 2.

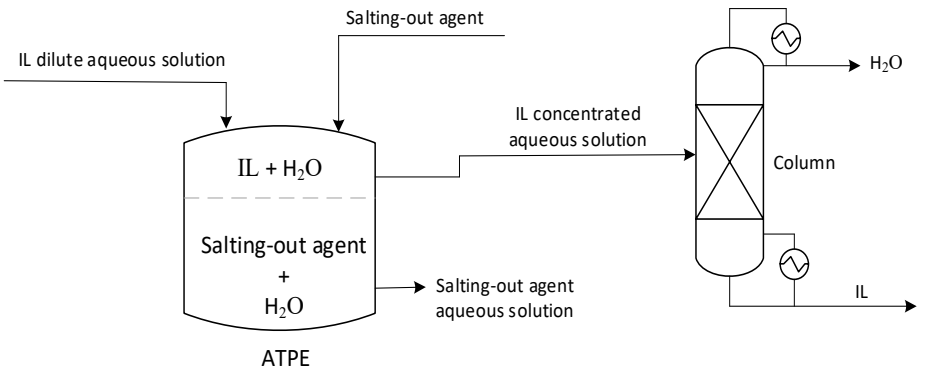
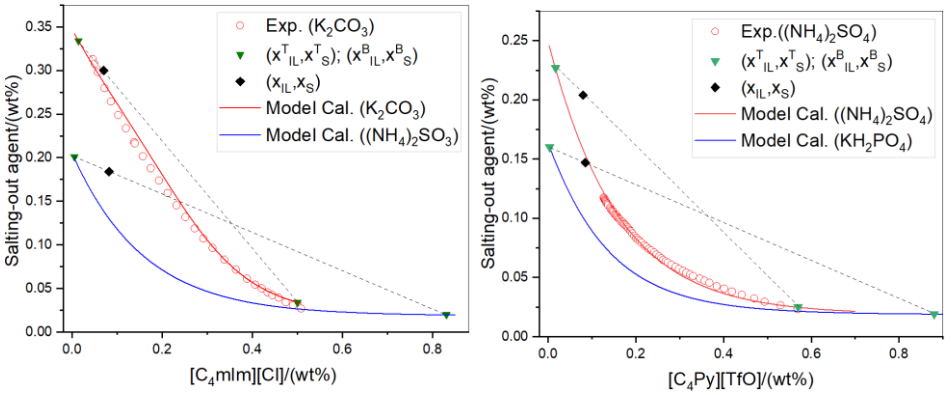


Figure 3: Hybrid process scheme for the recovery of ILs from dilute aqueous solutions.

3. Case studies

1-butyl-3-methylimidazolium chloride ([C_4mIm][Cl]) is a highly efficient direct solvent for the dissolution and regeneration of cellulose and large volumes of dilute IL aqueous solutions are produced during the precipitation of the regenerated cellulose.12,13 Therefore, efficient recycling of [C_4mIm][Cl] from these aqueous solutions is a critical step for the commercialization of this IL-based pretreatment technology. n-butylpyridinium trifluoromethanesulfonate ([C_4Py][TfO]) is another well-known hydrophilic IL that has potential industrial applications and it's also important to recover them from aqueous solutions during these applications. In this section, the proposed hybrid process design method will be used to recover [C_4mIm][Cl] and [C_4Py][TfO] from their aqueous solutions. First, two salting-out agents $(NH_4)_2SO_3$ and KH_2PO_4 are, respectively, identified for [C_4mIm][Cl]- and [C_4Py][TfO]-based ABS formation by solving the MINLP problems in the modelling system GAMS 24.4.6, where a deterministic global optimization solver, Lindoglobal, is applied. As shown in Figure 3a and 3b, both $(NH_4)_2SO_3$ and KH_2PO_4 have better ABS forming ability than their counterparts K_2CO_3 (Zafarani-Moattar et al., 2010) and $(NH_4)_2SO_4$ (Guo et al., 2020) reported in the literature, indicating the availability of this optimal salting-out agent design method.



(a) (b)

Figure 3: Ternary phase diagrams for ABS composed of (a) [C_4mIm][Cl]-H_2O-K_2CO_3/$(NH_4)_2SO_3$ and (b) [C_4Py][TfO]-H_2O-$(NH_4)_2SO_4$/KH_2PO_4.

For the recovery of 10 wt% [C_4Py][TfO] from aqueous solutions, the ABS of [C_4mIm][Cl]-H_2O-$(NH_4)_2SO_3$ gives an IL recovery efficiency of 95.0 wt% and a salting-out agent input of 2.36 kg/kg IL recovery, and for the ABS of [C_4Py][TfO]-H_2O-KH_2PO_4 they are 95.6 and 1.81, respectively.

After removing most water by APTE, the IL concentrated aqueous solution is sent to the distillation column, where purified IL can be obtained at the bottom and the rest of water is distilled from the top. In this work, the detailed process simulations of distillation column are performed in Aspen Plus. By far, ILs are still not included to the component database in Aspen Plus and therefore they should be defined as pseudo-components. To do this, properties of ILs such as molecular weights, physical properties and critical properties need to be specified. Likewise, information of the thermodynamic method for the IL containing system should also be specified. In this work, the physical property models are taken directly from our previous work (Chen et al., 2019) and

critical properties are calculated from the fragment contribution-corresponding states method proposed by Huang et al. (2013). On the other hand, UNIFAC model is selected as the thermodynamic method and model parameters including group volume parameters, surface area parameters and interaction parameters are taken from the published works (Song et al., 2020). Table 1 presents the process performance of both hybrid process scheme (Scheme 1) and pure distillation process (Scheme 2). Clearly, the hybrid process scheme demands much less energy input than that of the pure distillation process. However, a certain amount of salting-out agent is needed for ATPE in the hybrid process scheme. Nonetheless, the hybrid process scheme provides a good alternative for recovering ILs from dilute aqueous solutions due to its excellent energy performance.

Table 1: Energy performance of hybrid process scheme and pure distillation process.

IL aqueous solutions	$[C_4mIm][Cl]$ solution		$[C_4Py][TfO]$ solution	
Process scheme	Scheme 1	Scheme 2	Scheme 1	Scheme 2
Salting-out agent input (kg/kg IL recovery)	2.36	0	1.18	0
Energy input (kW/kW IL recovery)	0.16	6.86	0.082	6.62

4. Conclusions

A hybrid process scheme that combines ATPE and distillation method has been proposed for the recovery of hydrophilic ILs from their dilute aqueous solutions. In this hybrid process scheme, salting-out agents with high ABS forming ability are identified for ATPE through an optimal design method integrating the ANN-GC model into the computer-aided design technique. Two case studies are performed to test this hybrid design method. In both cases, the salting-out agents identified by the optimal design method have better ABS forming ability than their counterparts reported in the literature, and the hybrid process scheme present much better energy performance than the recovery process only using distillation unit.

References

A.S. Alshehri, R. Gani, F.You, 2020, *Computers & Chemical Engineering*, 141 (4), 107005.
J. Guo, S. Xu, Y. Qin, Y. Li, X. Lin, C. He, S. Dai, 2020, *Fluid Phase Equilibria*, 506 (15), 112394.
J. Zhou, H. Sui, Z. Jia, Z. Yang, L. He, X. Li, 2018, *Rsc Advances*, 8 (57), 32832-32864.
M.T. Zafarani-Moattar, S. Hamzehzadeh, 2010, Journal of Chemical & Engineering Data, 55 (4), 1598-1610.
S.P. Ventura, F.A. e Silva, M.V. Quental, D. Mondal, M.G. Freire, J.A. Coutinho, 2017, *Chemical reviews*, 117 (10), 6984-7052.
Y. Chen, E. Koumaditi, J.M. Woodley, G.M. Kontogeorgis, R. Gani 2018, *Computer Aided Chemical Engineering. Elsevier*. 43, 851-856.
Y. Chen, E. Koumaditi, R. Gani, G.M. Kontogeorgis, J.M. Woodley, 2019, *Computers & Chemical Engineering*, 130 (2), 106556.
Y. Chen, X. Liang, J.M. Woodley, G.M. Kontogeorgis, 2021a, *Chemical Engineering Science*, 247 (16), 116904.
Y. Chen, X. Meng, Y. Cai, X. Liang, G.M. Kontogeorgis, 2021b, *Industrial & Engineering Chemistry Research,* 60 (43), 15730-15740.
Y. Huang, H. Dong, X. Zhang, C. Li, S. Zhang, 2013, *AIChE Journal*, 59 (4), 1348-1359.
Z. Song, T. Zhou, Z. Qi, K. Sundmacher, 2020, *AIChE Journal*, 66 (2), e16821.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50025-7

A Short-Cut Method for Synthesis of Solvent-based Separation Processes

Shuang Xu[a], Toby Crump[a], Selen Cremaschi[a], Mario R. Eden[a] and Anjan K. Tula[b]*

[a]*Department of Chemical Engineering, Auburn University, Auburn, AL, 36849, USA*
[b]*College of Control Science & Eng., Zhejiang University, Hangzhou 310027, China*
anjantula@zju.edu.cn

Abstract

In the process industry, non-ideal mixtures are mainly separated by solvent-based separation, such as extraction, extractive distillation, and azeotropic distillation. For these separation methods, the separation barrier is overcome by adding an external component (solvent/entrainer) to the system. Much effort has been devoted to optimally design/select the solvent through screening different solvents' physical properties. It is also necessary to account for separation process properties such as energy consumption, number of stages, etc., during solvent selection. In this work, a short-cut evaluation model that can quickly assess the solvents' physical/mixture properties and process properties has been applied for designing an optimal separation-based process. Four case studies (acetone/chloroform, acetone/methanol, benzene/cyclohexane, and methanol/methyl acetate) have been considered. The results reveal that given a list of potential solvents, the short-cut evaluation model can correctly predict the process performance.

Keywords: Solvent-based separation; optimization; solvent selection.

1. Introduction

Solvent-based separation is a class of processes where non-ideal mixtures are purified based on their solubility difference (extraction) or vapor-liquid equilibrium difference (extractive distillation and azeotropic distillation). Typically, a third component (solvent/entrainer) is added to bypass the separation barrier and facilitate the separation. The effectiveness of this separation is highly dependent on the solvent. Different solvents lead to different process designs and eventually influence the overall capital/operating cost. Many solvent selection methods have been proposed to select the optimal solvent. Shen *et al*. (2015) proposed a solvent evaluation and ranking algorithm, which selects the solvents based on the summation of five important physical properties, such as boiling point, selectivity, molecular weight, etc. Cignitti *et al*. (2019) presented an optimization model to design the solvent by maximizing the separation driving force. Kossack *et al*. (2008) pointed out that solvent screening based on physical properties alone may result in unfavorable solvent choices. A more comprehensive solvent selection method, rectification body method (RBM), was proposed by Kossack *et al*. (2008), which can accurately calculate the process properties, like minimum solvent flowrate and minimum energy demand. However, this method is computationally demanding.

It is necessary to develop a fast and reliable solvent selection method so that a large number of solvents can be evaluated efficiently. A short-cut solvent evaluation model is presented in this paper, which takes both solvent physical properties and separation

process properties such as minimum energy consumption, the minimum number of stages, etc., into account during solvent selection. Given a list of potential solvents, the model can quickly evaluate the performance of the different solvents and give recommendations on the best option. The paper includes two parts: 1) solvent evaluation, 2) evaluation results validation. In the solvent evaluation, the model is applied to rank a list of solvents. The ranking results are validated by rigorous process simulation models where the operating/design variables are identified via derivative-free optimization.

2. Methodology

2.1. Solvent short-cut evaluation method

The solvent evaluation model aims to quickly and reliably assess different solvents based on various performance indicators. This model considers both the process properties and physical properties in the evaluation process. Firstly, a short-cut calculation model, which is based on Underwood and Fenske equations, is applied to calculate the process properties like minimum reflux ratio and the number of stages of a column. The Underwood and Fenske equations assume that the system has constant relative volatility. A typical extractive distillation system includes two columns, where the second column is simply solvent recovery distillation. Therefore, for the second column, the Underwood and Fenske equations can be applied to calculate the minimum number of stages and reflux ratio. However, these two equations cannot be directly applied to the extractive distillation column. Figure 1 shows the vapor-liquid equilibrium curve across the extractive distillation column. The extractive distillation column is divided into three sections: rectification, extraction, and stripping, and the relative volatility is different for each section, which means this change in relative volatility across the sections has to be accounted for. Here, we assume that the relative volatility is constant for each section, so the Fenske equation can be used in each section separately to calculate the minimum number of stages. The minimum reflux ratio is calculated when the operating line intersects with the VLE curve. In this way, the column minimum reboiler duty can be calculated by using the stage enthalpy balance. The extractive distillation column is described by Equations (1) – (6), where Equation (1) is only applied for ternary systems that do not have a separation boundary (Gerbaud and Rodriguez-Donis, 2014). For ternary systems with a separation boundary, such as the acetone/chloroform/ethylene glycol system, one can assume the distillate/bottom composition and calculate the minimum solvent flowrate through mass balance.

$$\left(\frac{F_E}{F_{AB}}\right)_{min} = \frac{(RR+1)D}{F_{AB}} \times \frac{(x_{PA}-y^*_{PA})}{(x_E-x_{PA})} + \frac{D(x_D-x_{PA})}{F_{AB}(x_E-x_{PA})} \tag{1}$$

$$N_{min,i} = \frac{lg\left[\left(x_{i,l}/x_{i,h}\right)/\left(x_{i+1,l}/x_{i+1,h}\right)\right]}{\lg(\alpha_{lh,i})}, i = 0,1,2 \tag{2}$$

$$N_{min} = \sum_{i=0}^{2} N_{min,i} \tag{3}$$

$$R_{min,j} = \frac{F_E x_{j,l} + D(x_D - y_{j,l})}{D(y_{j,l} - x_{j,l})}, j = 1,2 \tag{4}$$

$$R_{min} = max\{R_{min,1}, R_{min,2}\} \tag{5}$$

$$Q_{reboiler,min} = (R_{min}+1)DH_{D,V} + WH_W - H_E F_E - H_{AB}F_{AB} - R_{min}DH_{D,L} \tag{6}$$

In the model, F_E is the solvent/entrainer flowrate, F_{AB} is the raw material flowrate, RR is the predefined reflux ratio for solvent flowrate calculation. D and W are the distillate and bottom flowrates. x_D and x_W are the distillate and bottom compositions. x_E is the inlet solvent composition. x_{PA} is the minimum solvent composition that breaks the azeotropes after adding the solvent, and component A is the lightest component in the system. y_{PA}^* is the vapor composition in equilibrium with x_{PA}. $N_{min,i}$ is the minimum number of stages in the i^{th} section. $x_{i,l}$ and $x_{i,h}$ are the light and heavy component compositions in the i^{th} section. $\alpha_{lh,i}$ is the geometric relative volatility of the i^{th} section. $x_{j,l}$ and $y_{j,l}$ are the light compound liquid and vapor compositions at the first (j=1) and last (j=2) stage of the extractive section. R_{min} is the minimum reflux ratio. $H_{D,V}$ are $H_{D,L}$ are the vapor and liquid enthalpies of the distillate product. H_E and H_{AB} are the enthalpies of entrainer and binary raw materials. The inlet is assumed to be at boiling point (q = 1). After analyzing different extractive distillations systems, $x_{1,E}$ is usually between 60 % to 80%, $x_{2,E}$ is close to $x_{1,E}$, and the difference is in the range of 5%.

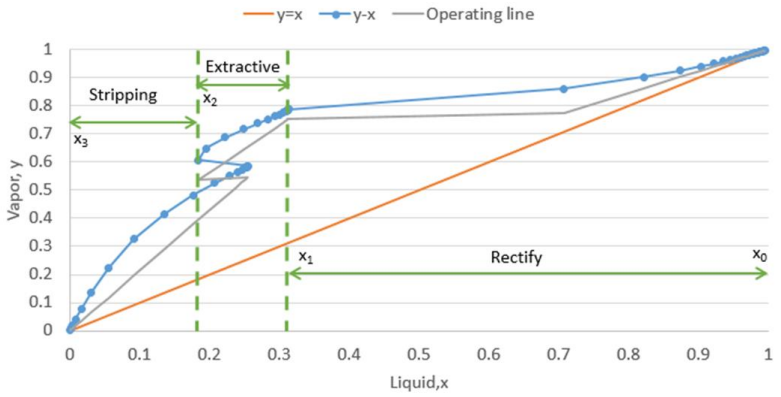


Figure 1. VLE plot of acetone/methanol/water extractive distillation. (x_0, x_1, x_2, x_3 – boundary composition in each section).

For each solvent, we calculate the minimum reboiler duty, number of stages, and reflux ratio by applying the proposed calculation method. Eight properties, including six process properties, e.g., the minimum number of stages, reboiler duty and reflux ratio, and two physical properties, e.g., solvent flowrate and boiling point, are considered in the evaluation model. These properties are selected because they directly influence the process capital and/or utility cost. Given N number of potential solvents, for each property, a value from one to N is assigned (where one is given to the best solvent with that property). Finally, the solvents are ranked based on summation scores, and the best solvent has the overall lowest score.

2.2. Validation model

To validate the rankings given by the solvent evaluation method, simulation-based optimization is employed to optimally design the extractive distillation setups. The total annualized cost, which considers both capital and utility costs, is the objective function. The number of stages, inlet stage, and solvent flowrate are the decision variables. A generalized form of the optimization problem solved by the simulation-based optimization algorithms is given in Equations (7)-(15).

$$min: TAC = \frac{i(i+1)^n}{(i+1)^n - 1} \times IC + AUC \quad (7)$$

$$st. IC = \sum_j Cost_j(q_{s,j}) \quad (8)$$

$$AUC = 24 \times 300 \times \sum_j Utility_j(q_{o,j}) \quad (9)$$

$$q = \Theta(F_S, N_k, N_{k,feed}, N_{1,S}), k = 1,2 \quad (10)$$

$$x_m \geq purity \quad (11)$$

$$F_{S,L} \leq F_S \leq F_{S,U} \quad (12)$$

$$N_{k,feed,L} \leq N_{k,feed} \leq N_{k,feed,U} \quad (13)$$

$$N_{k,L} \leq N_k \leq N_{k,U} \quad (14)$$

$$N_{1,S,L} \leq N_{1,S} \leq N_{1,S,U} \quad (15)$$

Here, TAC is the total annualized cost, IC is the investment cost, AUC is the annualized utility cost, i is interest, n is plant life (n=5 is used in this model), $Cost_j$ is the investment cost of equipment j, $Utility_j$ is the utility cost of equipment j, $q_{s,j}$ is the sizing variable for equipment j, $q_{o,j}$ is the operating variable for equipment j, x_m is the product purity of component m, $\Theta(F_S, N_k, N_{k,feed}, N_{1,S})$ is the process simulation model, F_S is the solvent flowrate rate, N_k is the number of stages of column k, $N_{k,feed}$ is the materials feed stages of column k, $N_{1,S}$ is the solvent feed stage of the first column, and L and U represent the lower and upper bounds.

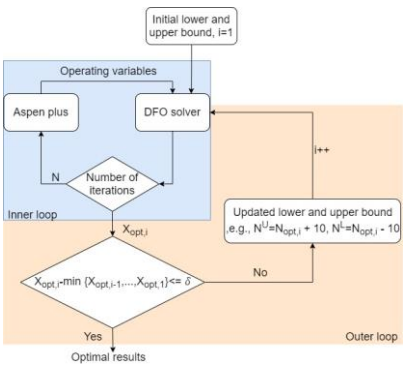


Figure 2. DFO solving algorithm for validation model.

To generate stable results that are not influenced by the initial lower and upper bounds, an algorithm (Figure 2) is developed to solve the optimization problem. The algorithm has two parts: 1) inner loop: given the initial upper and lower bounds of the decision variables, the inner loop runs the process simulation, and using a derivative-free optimization (DFO) solver, the best design parameters are identified. 2) outer loop: according to the identified best design parameters from the inner loop, the lower and upper bounds are updated and sent back to the inner loop. The process terminates after reaching a stable objective value.

3. Results and Discussion

Four separation systems, acetone/chloroform, acetone/methanol, benzene/cyclohexane, and methanol/methyl acetate, with their potential solvents, were selected from the review

paper by Gerbaud *et al*. (2019). The short-cut evaluation model and DFO results are listed in Table 1. The identified design parameters of the best solvent are listed in Table 2.

Table 1. Tested evaluation and DFO results.

Solvents	Score	TAC, 10^6\$	Solvents	Score	TAC, 10^6\$
Acetone/Chloroform			Acetone/Methanol		
EG	16	0.77	Water	20	3.32
DMSO	14	0.66	2-Proponal	35	7.6
o-Xylene	28	1.05	Ethanol	27	6.64
Benzene	33	1.29	DMSO	14	2.87
Chlorobenzene	29	0.94	EG	24	4.63
Benzene/Cyclohexane			Methanol/Methyl Acetate		
Dimethyl phthalate	22	0.95	DMSO	13	1.05
NMP	24	0.76	EG	18	1.11
Aniline	34	0.97	2-Methoxyethanol	17	0.98
Sulfolane	19	0.69	-	-	-
Furfural	21	0.72	-	-	-

For acetone/chloroform separation, the short-cut evaluation method ranks the five solvents in the following order: DMSO (Dimethyl sulfoxide) > EG (ethylene glycol) > chlorobenzene > o-xylene > benzene. The DFO gave a similar TAC order except for chlorobenzene and o-xylene. This is because the system has a separation boundary, and we have to approximate this separation boundary using calculated residue curves from process simulation software. The solvent flowrate of o-xylene is 1.22 times larger than chlorobenzene, but it is 1.61 times larger from the DFO results. So, the o-xylene system has higher reboiler duty and higher TAC value. The short-cut and DFO results give the same order: DMSO > water > EG > ethanol > 2-proponal for acetone/methanol separation. For the benzene/cyclohexane case, the short-cut model predicts the following order: sulfolane > furfural > dimethyl phthalate > N-Methyl-2-pyrrolidone (NMP) > aniline, but the DFO results show that the NMP has better performance than dimethyl phthalate. Although the evaluation model correctly represents that the dimethyl phthalate system has a lower number of stages, but its high boiling point results in higher column temperatures, which requires a furnace. Due to this, the capital cost of dimethyl phthalate system is higher than the NMP system. For the methanol/methyl acetate case study, only three solvents were selected because of the lack of experimental phase equilibrium data. Among these three solvents, the short-cut evaluation method predicts that DMSO has the best performance, but the DFO results show that 2-methoxyethanol has better performance. The evaluation model shows that the 2-methoxyethanol has a smaller number of stages for the extractive column and a higher number of stages for the second column, which results in similar capital costs. However, the predicted minimum reboiler duty does not correctly represent the utility cost. Two reasons may cause this deviation: 1) Boiling point, DMSO has higher boiling point than 2-methoxyethanol, so different types of utilities have to be used. However, in this model, we rank the properties only based on their relative heat duty, and the different types of utilities are not considered. 2) Boundary composition (x_1, x_2). The boundary composition influences the sections' relative volatility and thus influences the calculated reboiler duty. The same $x_{1,E}$ value is used in all cases, but DFO proves that DMSO $x_{1,E}$ is equal to 60%, while 2-methoxyethanol, EG has similar $x_{1,E}$ values around 70%.

Table 2. Design parameters of the identified best solvent for the four separation systems.

		Acetone/ chloroform	Acetone/ methanol	Benzene/ cyclohexane	Methanol/ methyl acetate
Best identified solvent		DMSO	DMSO	Sulfolane	2-methoxyethanol
F_{AB} (equimolar), kmol/h		100	540	100	100
T1	N_1	36	41	25	49
	$N_{1,f}$	13	27	13	37
	$N_{1,s}$	3	3	2	6
T2	N_2	13	19	13	23
	$N_{2,f}$	4	14	4	8
F_s, kmol/h		111	399.6	88	263

4. Conclusions

Solvents can alter the relative volatility of mixtures and therefore the selection of the optimal solvent impacts extractive distillation design and operation. The best solvent has to balance the process capital and utility cost so that the annualized cost is minimized. This work presents a simple and reliable short-cut evaluation method to assist in solvent selection for solvent-based distillation. The proposed method was applied to four different extractive distillation systems. By including the process properties in the solvent ranking algorithm, the solvent with the best process performance (lower capital/utility cost) is identified. The evaluation results were validated by a rigorous design approach where the key operating parameters are optimally designed. Both the solvent evaluation and the optimized process results demonstrated that DMSO, DMSO, sulfolane, are the best solvents for separating acetone/chloroform, acetone/methanol, and benzene/cyclohexane azeotrope systems, respectively. The methanol/methyl acetate azeotrope system results indicated that the solvent boiling point and the choice of boundary composition might highly impact the ranking results, so the ranking algorithm will need to be further improved by taking these factors into account. The proposed approach can be applied as a first screening of potential solvents with low computational cost and decent screening results.

References

Cignitti, S., Rodriguez-Donis, I., Abildskov, J., You, X., Shcherbakova, N. and Gerbaud, V., 2019. CAMD for entrainer screening of extractive distillation process based on new thermodynamic criteria. *Chemical Engineering Research and Design*, 147, 721-733.

Gerbaud, V. and Rodríguez-Donis, I., 2014. Extractive distillation. In *Distillation* (pp. 201-245). Academic Press.

Gerbaud, V., Rodriguez-Donis, I., Hegely, L., Lang, P., Denes, F. and You, X., 2019. Review of extractive distillation. Process design, operation, optimization and control. *Chemical Engineering Research and Design*, 141, 229-271.

Kossack, S., Kraemer, K., Gani, R. and Marquardt, W., 2008. A systematic synthesis framework for extractive distillation processes. *Chemical Engineering Research and Design*, 86(7), 781-792.

Shen, W., Dong, L., Wei, S. A., Li, J., Benyounes, H., You, X., & Gerbaud, V. (2015). Systematic design of an extractive distillation for maximum-boiling azeotropes with heavy entrainers. *AIChE Journal*, 61(11), 3898-3910.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50026-9

Modeling and Optimization of Ionic-Liquid-Based Carbon Capture: Impact of Thermal Degradation Kinetics

Kyeongjun Seo[a], Zhichao Chen[a], Thomas F. Edgar[a], Joan F. Brennecke[a], Mark A. Stadtherr[a], Michael Baldea[a*]

[a]*McKetta Department of Chemical Engineering, The University of Texas at Austin, Austin, Texas 78712, United States*

mbaldea@che.utexas.edu

Abstract

Ionic liquids (ILs) have recently been considered as alternatives to conventional amine-based absorbents in post-combustion CO_2 capture processes. However, solvent losses by IL thermal degradation could be more significant than in the case of conventional amine solvents. In this paper, we propose an advanced process design that uses a thin-film unit under vacuum to minimize the thermal degradation of solvent during regeneration. We employ rigorous thermodynamics and rate-based mass transfer models, with robust simulation and optimization capabilities implemented using a pseudo-transient modeling technique. The impact of solvent thermal degradation on the economic performance of the IL-based carbon capture process is studied. A comparison to a conventional process design is presented.

Keywords: carbon capture, flowsheet optimization, ionic liquids, process design, thermal degradation kinetics

1. Introduction

Recently, ionic liquids (ILs) have gained attention as promising solvents for post-combustion carbon capture due to desirable properties such as negligible volatility, high CO_2 absorption capacity and low heat of regeneration (Aghaie et al., 2018).

In this work, we consider triethyl-(octyl)phosphonium 2-cyanopyrrolide ([P_{2228}][2-CNPyr]) as an IL chemical absorbent for CO_2 capture because of its high CO_2 absorption capacity, moderate reaction enthalpy, superior reversibility, and relatively low viscosity (Seo et al., 2014). Although it is a promising candidate IL for carbon capture in terms of these properties, its thermal stability should also be considered because solvent (thermal) degradation could result in economic losses as well as operational problems (Rao and Rubin, 2002). Our experiments indicate that the thermal degradation rate of [P_{2228}][2-CNPyr] is comparable to or greater than that of monoethanolamine (MEA), a conventional amine-based solvent. In addition, given that the bulk price of ILs is expected to be higher (estimated \$10/kg) than that of conventional amine solvents (e.g., MEA at \$1.5-2/kg) (Ramdin et al., 2012), the cost associated with solvent loss is particularly important.

Solvent thermal decomposition occurs mainly during solvent regeneration, where the solvent is exposed to high temperatures. A conventional solvent regeneration system (stripper and reboiler) has a relatively long liquid residence time (5-10 minutes) (Walters et al.,

2016; Jung et al., 2018), exposing the solvent to high temperatures for prolonged periods of time. The thermal degradation of the solvent can be reduced by using a short residence time with a small liquid hold-up (Alhusseini et al., 1998) at the regeneration temperature. Thin-film technology can replace such a conventional regeneration system, diminishing residence times and thus the thermal degradation issue. When thin film units are operated under vacuum pressure, the regeneration temperature can be lowered even further.

Based on these considerations, we propose a novel flowsheet design for an IL-based carbon capture process. We then perform economic optimization of the proposed process flowsheet using a pseudo-transient optimization framework (Pattison and Baldea, 2014) with a focus on the economic impact of solvent thermal degradation. The regeneration temperature and associated liquid residence time are reduced using a thin-film column integrated with a vacuum compressor. This can significantly reduce the thermal degradation of IL absorbent and the associated make-up costs compared to the conventional regeneration system using a reboiler.

2. Flowsheet description

Figure 1 shows the proposed process design for IL-based CO_2 capture. The conventional regeneration system that consists of a stripper and reboiler is replaced with a thin-film unit. A thin-film unit is composed of a bundle of tubes in a shell. A liquid film flows downward on the interior vertical surface of each tube and the tube walls are heated by steam on the shell side. This unit can provide high heat and mass transfer rates due to the large surface area created by the liquid film. In the case of the proposed carbon capture plant, the liquid is the rich (i.e., high CO_2 concentration) IL solvent. The CO_2 is desorbed from the liquid and leaves at the top of the unit.

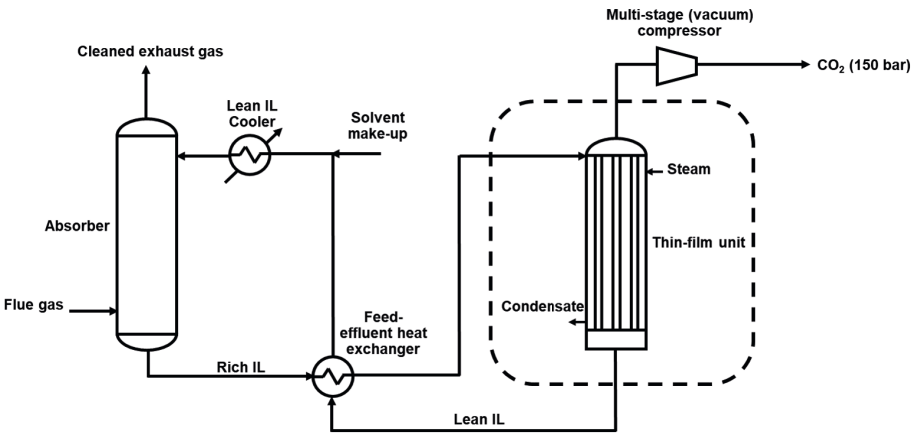


Figure 1: Proposed IL-based carbon capture process flowsheet. A conventional system stripper and reboiler system is replaced with a thin-film unit (shown by the dashed line) for solvent regeneration.

The CO_2 absorption model and other physical properties for the IL solvent used in this work ([P_{2228}][2-CNPyr]) are presented in our previous work (Seo et al., 2020). The experimentally measured thermal degradation of this IL solvent can be described by apparent zero-order kinetics, with an Arrhenius temperature dependence of the rate constant:

$$\frac{m(t)}{m(0)} = 1 - \left(k_0 \exp\left(-\frac{E_\mathrm{a}}{RT} \right) \right) t \tag{1}$$

where $m(0)$ is the initial mass of solvent and $m(t)$ is the mass after time t. The pre-exponential factor k_0 and activation energy E_a are estimated from experimental data to be $k_0 = 305.5\ \mathrm{h}^{-1}$ and $E_\mathrm{a} = 41.8$ kJ/mol.

We use a rate-based mass transfer model to describe kinetically limited transport phenomena. A detailed description of the rate-based model for the absorber column and the related material and energy balances can be found in our previous work (Seo et al., 2020). However, we modify the mass transfer coefficient and effective area correlations (Song et al., 2018) to be more suitable for viscous IL solvent flow. For the thin-film unit, the same rate-based model is used with some modifications. The liquid phase mass transfer coefficient is estimated using the correlation of Yih and Chen (1982). The mass transfer resistance in the vapor phase is assumed to be negligible. The mass transfer area per volume is calculated by dividing the total tube surface area by the overall thin-film column volume. For the heat transfer rate in the thin-film unit, an additional heat transfer term between the steam and liquid film is introduced. The heat exchanger model is also based on Seo et al. (2020). However, the overall heat transfer coefficient and the associated pressure drop models are modified to use empirical correlations for viscous liquids (Talik et al., 1995). The flooding point for the thin-film unit is determined using an empirical correlation from Mouza et al. (2005). Finally, the liquid residence time for each unit (only residence times in the heat exchanger and the regeneration unit where the solvent operates at high temperature are considered) is estimated from the ratio of the total hold-up volume to the liquid flowrate.

3. Process economic optimization

The optimization problem is formulated as:

$$\begin{aligned} \min_{\pi} \quad & \phi(\chi, \pi, \xi) \\ \text{s.t.} \quad & f(\chi, \pi, \xi) = 0 \\ & c(\chi, \pi, \xi) \leq 0 \end{aligned} \tag{2}$$

where the objective function, ϕ, is the sum of the annualized capital cost (for the absorber, heat exchanger, compressor, cooler, gas blower, solvent pump, and thin-film unit) and the operating cost (for heating, cooling, electricity, and solvent make-up) of the IL-based CO_2 capture process, f is the flowsheet model described above, and c are process operating constraints. π are process decision variables, χ are process state variables, and ξ are process parameters. The decision variables and constraints are summarized in Table 1. The resulting CO_2 capture process flowsheet model is difficult to solve because of its large size and coupled nonlinear equations. We improve the initialization and convergence of this complex flowsheet optimization problem using a pseudo-transient modeling technique (Pattison and Baldea, 2014).

Table 1: Decision variables and process constraints for process optimization

Component	Relevant variables and equations
Decision variables[a]	F^{L}, L, D, L_{T}, D_{T}, $T_{\text{absorber,in}}$, $T_{\text{thin-film,in}}$, T_{steam}, P_{T}
Process constraints[b]	$\frac{F^{\mathrm{V}}_{\mathrm{CO_2,in}} - F^{\mathrm{V}}_{\mathrm{CO_2,out}}}{F^{\mathrm{V}}_{\mathrm{CO_2,in}}} \geq 0.9$
	$\Delta T_{\text{appr, min}} \geq 1\,^{\circ}\mathrm{C}$
	$T_{\mathrm{S}} \leq 150\,^{\circ}\mathrm{C}$
	$0.08\,\mathrm{bar} \leq P_{\mathrm{T}} \leq 1\,\mathrm{bar}$
	$\mathrm{Fr_V} \leq 0.8\mathrm{Fr_V^*}$

[a] F^{L} is IL solvent circulation flowrate, L and D are height and diameter of the absorber, L_{T} and D_{T} are height and diameter of the thin-film unit, $T_{\text{absorber,in}}$ and $T_{\text{thin-film,in}}$ are inlet temperatures of the absorber and thin-film unit, T_{steam} is regeneration steam temperature, and P_{T} is pressure of the thin-film unit.

[b] The CO_2 removal rate is constrained to be at least 90%, the minimum approach temperature of heat exchanger is constrained to be not lower than 1 °C, the suction pressure of the vacuum is limited to equal or greater than 0.08 bar, and the thin-film unit is restricted to operate below 80% of the flooding point.

Figure 2 shows a comparison of optimal process costs between the proposed (thin-film unit operated under vacuum pressure) and conventional (stripper with a reboiler system operated under atmospheric pressure) regeneration systems. The flue gas conditions correspond to a natural gas combined cycle power plant (case B31B in James et al. (2019)). The same cost correlations are used for both systems. In the conventional process, the residence times in the stripper and reboiler are assumed to be 0.1 and 5 minutes, respectively (Walters et al., 2016).

The absorber cost of the conventional system ($74.4 M/year) is higher than that of the thin-film system ($38.7 M/year). Also, the optimal absorption temperature is 15 °C for the conventional system whereas it is 30 °C for the proposed system. This is because a smaller mass transfer area for CO_2 absorption and higher absorption temperature would be sufficient to capture the same level of CO_2 since the solvent regeneration is more effective under reduced pressure. The optimal operating pressure for the proposed regeneration system is found to be 0.41 bara. Therefore, the compressor equipment cost is much higher for the thin-film case ($51.0 M/year) compared to the conventional case ($18.4 M/year) because a larger compressor is required to accommodate increased CO_2 gas volume at the reduced operating pressure. However, the CO_2 loading in the regenerated IL solvent can be much lower for the proposed system (0.097 CO_2/mol IL) compared to the conventional system (0.161 mol CO_2/mol IL).

A key comparison in the operating costs is the solvent replacement cost. The residence time in the regeneration system is much smaller in the proposed configuration (thin-film: 1.2 min vs. conventional: 5.1 min). This can be attributed to a small liquid hold-up volume in the thin-film unit. As a result, the solvent make-up cost related to thermal degradation is significantly reduced (thin-film: $31.7 M/year vs. conventional: $90.2 M/year). Electric-

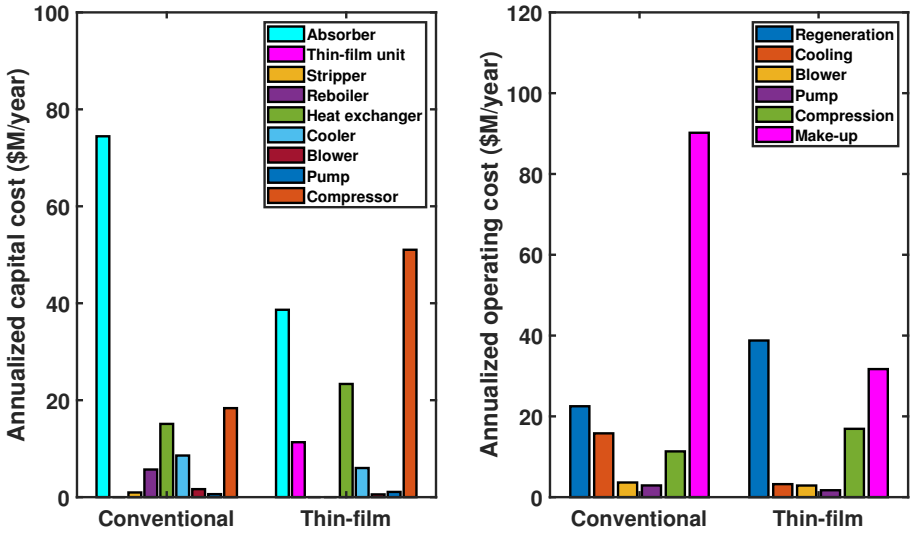


Figure 2: Comparison of process economics between proposed and conventional systems.

ity cost for the compression would be higher for the thin-film system because of operating under vacuum (thin-film: \$16.9 M/year vs. conventional: \$11.3 M/year). However, the solvent cooling cost is higher for the conventional system because of lower absorption temperature (thin-film: \$3.2 M/year vs. conventional: \$15.8 M/year). Overall, the process economic cost is significantly reduced for the proposed system (\$227.3 M/year) relative to the cost for the conventional setup (\$271.6 M/year).

4. Conclusions

Solvent thermal degradation can be a significant concern in an IL-based CO_2 capture process. In this work, we propose a flowsheet configuration that utilizes a thin-film solvent regeneration unit under vacuum for reducing the regeneration temperature and the associated liquid residence time. We determine the optimal annualized process cost of this advanced IL-based carbon capture process (for a a natural gas combined cycle power plant) and find a significant reduction in process cost (in particular, the solvent make-up cost associated with thermal degradation is reduced by about a factor of three) compared to the conventional case, in which a stripper and reboiler are used for the solvent regeneration.

5. Acknowledgments

This work was supported by the University of Texas Energy Institute under the Fueling a Sustainable Energy Transition program. Partial support from ExxonMobil is acknowledged with gratitude.

References

M. Aghaie, N. Rezaei, S. Zendehboudi, 2018, A systematic review on CO_2 capture with ionic liquids: Current status and future prospects, Renew. Sustain. Energy Rev., 96, 502-525.

A. A. Alhusseini, K. Tuzla, J. C. Chen, 1998, Falling film evaporation of single component liquids, Int. J. Heat and Mass Transf., 41, 12, 1623-1632.

B. E. Gurkan, T. R. Gohndrone, M. J. McCready, J. F. Brennecke, 2013, Reaction kinetics of CO_2 absorption in to phosphonium based anion-functionalized ionic liquids, Phys. Chem. Chem. Phys., 15, 7796-7811.

R. E. James, D. Kearins, M. Turner, M. Woods, N. Kuehn, A. Zoelle, 2019, Cost and Performance Baseline for Fossil Energy Plants Volume 1: Bituminous Coal and Natural Gas to Electricity, NETL-PUB-22638.

H. Jung, D. Im, S. H. Kim, J. H. Lee, 2018, Dynamic modeling and analysis of amine-based carbon capture systems, IFAC-PapersOnLine, 51, 18, 91-96.

A. Mouza, M. Pantzali, S. Paras, 2005, Falling film and flooding phenomena in small diameter vertical tubes: The influence of liquid properties, Chem. Eng. Sci., 60, 18, 4981-4991.

R. C. Pattison, M. Baldea, 2014 Equation-oriented flowsheet simulation and optimization using pseudo-transient models, AIChE J., 60, 4104-4123.

M. Ramdin, T. Loos, T. Vlugt, 2012, State-of-the-art of CO2 capture with ionic liquids, Ind. Eng. Chem. Res., 51, 8149-8177.

A. B. Rao, E. S. Rubin, 2002, A technical, economic, and environmental assessment of amine based CO_2 capture technology for power plant greenhouse gas control, Environ. Sci. Technol., 36, 20, 4467-4475.

S. Seo, M. Quiroz-Guzman, M. A. DeSilva, T. B. Lee, Y. Huang, B. F. Goodrich, W. F. Schneider, J. F. Brennecke, 2014, Chemically tunable ionic liquids with aprotic heterocyclic anion (AHA) for CO_2 capture, J. Phys. Chem. B, 118, 21, 5740-5751.

K. Seo, C. Tsay, B. Hong, T. F. Edgar, M. A. Stadtherr, M. Baldea, 2020, Rate-based process optimization and sensitivity analysis for ionic-liquid-based post-combustion carbon capture, ACS Sustain. Chem. & Eng., 8, 10242-10258.

D. Song, A. F. Seibert, G. T. Rochelle, 2018, Mass transfer parameters for packings: Effect of viscosity, Ind. Eng. Chem. Res., 57, 2, 718-729.

A. C. Talik, L. S. Fletcher, N. K. Anand, L. W. Swanson, 1995, Heat transfer and pressure drop characteristics of a plate heat exchanger using a propylene-glycol/water mixture as the working fluid, 30th National Heat Transfer Conference, 83-88.

M. S. Walters, T. F. Edgar, G. T. Rochelle, 2016, Regulatory control of amine scrubbing for CO_2 capture from power plants, Ind. Eng. Chem. Res., 55, 16, 4646-4657.

S.-M. Yih, K.-Y. Chen, 1982, Gas absorption into wavy and turbulent falling liquid films in a wetted-wall column, Chem. Eng. Commun., 17, 1-6, 123-136.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50027-0

Process Design of Formic Acid and Methanol Production from CO2 Promoted by Ionic Liquid: Techno-Economic Analysis

Taofeeq O. Bello*, Antonio E. Bresciani, Claudio A.O. Nascimento, Rita M.B. Alves

Department of Chemical Engineering, Escola Politecnica, Universidade de São Paulo, Av. Prof. Luciano Gualberto, travessa do Politécnico, nº 380, São Paulo 05508-010, Brazil
tbello@usp.br

Abstract

Carbon dioxide conversion technologies have been extensively investigated as a viable pathway for lowering greenhouse gas emissions. However, due to thermodynamic and product separation limitations, numerous routes have been proposed. This work presents a techno-economic study of the production of formic acid and methanol promoted by ionic liquid at a commercial scale. To that aim, Aspen Plus® V10 was employed to build a simulation that included the solubilization of CO_2 in 1-ethyl-2,3-dimethylimidazolium nitrite ([Edmim][NO_2]) ionic liquid (IL), synthesis of the CO_2-[Edmim][NO_2] adduct with hydrogen, product separation, and recycling of the IL. The CO_2 conversion (87 %) resulted in ~83 % and ~14 % yield of formic acid and methanol, respectively. This result is an improvement in previous conducted findings. Furthermore, it was discovered that a discount rate between 4-5 % (@ 0.78 USD/kg of formic acid) or 0.93-1 USD/kg (@ 10% discount rate) would make the project profitable.

Keywords: Carbon dioxide Conversion; Formic acid; Ionic Liquid; Methanol.

1. Introduction

Carbon dioxide (CO_2) utilization and conversion in the production of fuels, chemicals, and materials are potentially promising CO_2 abatement alternatives by lowering CO_2 emissions, reducing fossil fuel usage (Pérez-Fortes and Tzimas, 2016), and also providing a chemical storage alternative for intermittent renewable electricity (Schlögl, 2013). This approach can significantly contribute to the decarbonization of the energy system (Olah et al., 2009). Formic acid (FA) and methanol (MeOH) are typical examples of chemicals and liquid energy carriers. However, the hydrogenation of CO_2 to formic acid is endergonic in the gas phase (ΔG^{o}_{298} = +33 kJ/mol), hence, thermodynamically unfavorable (Wang & Himeda, 2012; Leitner, 1995). The thermodynamic limitation can be overcome by perturbing the reacting system with a secondary reaction or molecular interaction. One of the available strategies is the neutralization of the reaction with a weak base (tertiary amines or alkali/alkaline earth bicarbonates) to yield formamides (Xu et al., 2011; Jessop et al., 1999). However, there are concerns about the post-treatment of intermediates to get a pure formic acid. (Leitner, 1995; Su et al., 2015). Ionic liquids (ILs) play an essential role in solving these two problems due to their solvating and low volatility property (Zeng et al., 2017). In addition, ILs can fine-tune the properties of the solvent by altering the structure, catalyst immobilization (Ghavre et al., 2011; Kokorin, 2012; MacFarlane et al., 2017), and CO_2 activation (Wang et al., 2015). Hence, in this work, the economic implications of deploying a process plant for the hydrogenation of CO_2 to formic acid and methanol using IL ([Edmim][NO_2]) as the reaction media was examined. The evaluation to retrieve technical and process significant parameters was carried out with the Aspen Plus V10 process simulation software.

2. Process Description.

The process flow diagram of the CO_2 hydrogenation to FA acid and MeOH was developed and shown in Figure 1. The synthesis method is a two-step process comprising CO_2 solubilization and conversion in a column and reactor, respectively. The plant capacity was set at 33,000 t/y of FA and MeOH with a purity of 97.7 % and 99.99 %, respectively. The feedstocks for this process are CO_2 and hydrogen (H_2) and the ionic liquid, which serves as the reaction media. CO_2 and H_2) were assumed to be free from impurity. The [Edmim][NO_2] was initially heated to a temperature of 40 ºC before entering the solubilization section together with compressed CO_2 at 80 bar. The CO_2 dissolves in the IL forming a CO_2-[Edmim][NO_2] adduct. The resulting adduct leaves the solubilization unit and enters the synthesis section together with a stream of compressed hydrogen gas. The reactor operates at 20 ºC and 17 bar for CO_2 conversion. FA, MeOH, and water are the resulting products of the reaction, which, together with unreacted CO_2 and H_2, are sent to a separator to remove and recycle the [Edmim][NO_2], while the remaining compounds leave the column as vapor products. The products are cooled, and unreacted H_2 and CO_2 are separated using black-box separator units and recycled to the reactor and solubilization column, respectively. The stream of

formic acid, methanol, and water are sent to the separation unit, where two distillation column units are employed. Methanol is separated at the first distillation column. The bottom product from the first column, an azeotropic mixture of FA and water, enters an extractive distillation using [Edmim][NO_2] as solvent. FA with 97.7 % w/w is recovered as the top product, while [Edmim][NO_2] and water are separated by a simple flash separation process. Table 1 presents the operating conditions for the main process equipment.

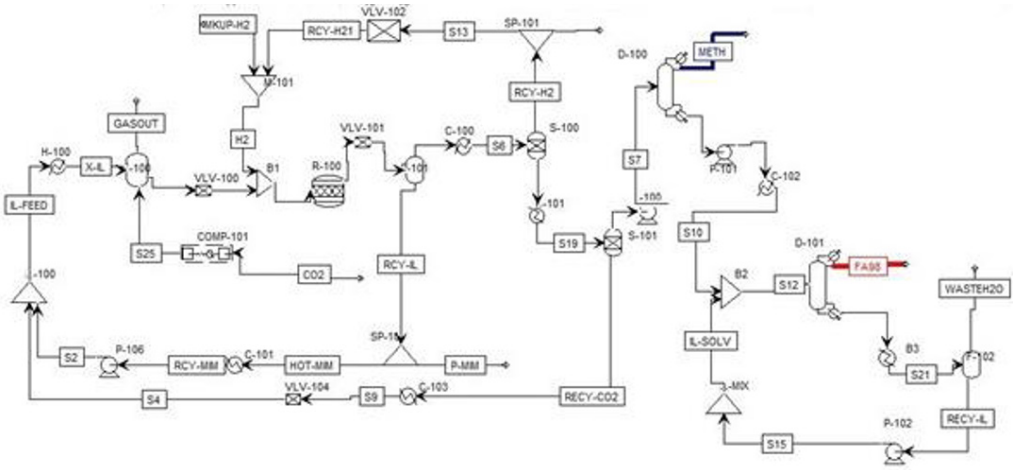


Figure 1: Proposed process flowsheet of CO_2 hydrogenation with [Edmim][NO_2] as reaction media

Table 1: Operating conditions for the main process equipment

Units		Operating Conditions
Compression	CMP-101	Pexit = 80 bar, Number of stages = 3
Solubilization Column	F-100	T = 20 ºC; P = 80 bar
Separator	F-101	T = 150 ºC; P = 0.1 bar
Reactor	R-100	T = 20 ºC; P = 17 bar
Distillation Column	D-100	P = 1 bar; Stages = 22; Feed stage = 11; Reflux ratio = 4; Condenser: Full.
Extractive Distillation Column	D-101	P = 1 bar; Stages = 23; Feed stage = 2; Reflux ratio = 0.01; Condenser: full; Distillate to feed ratio = 0.69

3. Process Simulation.

The thermodynamic models for the CO_2 solubilization and synthesis sections are the conductor-like screening model for segment activity coefficient (COSMO-SAC) with Peng Robinson-Wong Sandler equation of state (ESPRWS). Due to the unavailability of experimental data of [Edmim][NO2], its thermodynamic properties were estimated by Conductor Like Screening Model for real solvents (COSMO-RS) as described in previous works (Bello et al., 2021a, 2021b). The reactor was modeled using RYield with two independent reactions (FA and MeOH formation) (Bello et al., 2021b). The solubilization column was modelled with a two-outlet flash using rigorous vapor-liquid equilibrium.. The distillation columns (D-100 and D-101) were modelled with a rigorous RADFRAC model in equilibrium mode. All the property methods were selected following the guidelines of Towler and Sinnott (2013) and taking into account the reaction system's temperature, pressure, and volatility. Multistage compressors

were selected and modelled as isentropic with a fixed discharge pressure from the last stage. Heat exchangers were modelled by the shortcut method.

4. Techno-Economic Assessment

In any chemical project, estimating capital (CAPEX) and operational (OPEX) costs are critical components in determining the long-term viability of any chemical process. The CAPEX comprises costs such as equipment, land, and installation. Raw materials (CO_2 and H_2), reaction media [Edmim][NO_2], and utilities are all included in the OPEX. The equipment purchase and utility costs were estimated using the inbuilt Aspen Process Economic Evaluation (APEA). Aspen software's cost basis calculation is based on the first quarter of 2016. When compared to other cost correlations, this method can provide reasonably accurate cost estimates during the conceptual phase (Towler and Sinnott, 2012). The installation costs of the sized equipment were then calculated. After that, the total capital investment was determined utilizing several factors linked to the total installation costs. Revenues were calculated by multiplying each product's annual production by its market value. A discounted cash flow analysis was performed assuming a 15-year plant lifespan. The projected interest rate was 10%, the income tax rate was 45%, and depreciation was calculated using the straight-line technique for project years. The impact of the product price and discount rate on the project's Net Present Value (NPV) were evaluated.

5. Results and Discussion

5.1. Process Simulation Results

As seen in Table 2, the technical indicators presented are the per pass and overall CO_2 conversions, as defined by Eq.(1) and Eq.(2), and utility requirements. As depicted in Table 1, per pass CO_2 conversion of 86% was achieved in the presence of the [Edmim][NO_2] as reaction media. The unreacted CO_2-[Edmim][NO_2] adduct was recycled back to the reacting system, which allows nearly 100 % CO_2 conversion.

Table 2. Technical indicators of the CO_2 hydrogenation to formic acid and methanol process.

Indicators	Values	Units
Overall CO_2 conversion	100	%
Per Pass Conversion	87.5	%
Conversion factor (FA)	1.17	tCO2/t FA
Conversion factor (MeOH)	6.87	tCO2/t MeOH
MeOH Produced	0.46	t/h
FA Produced	2.68	t/h
Hot utility	1.35	MWh/t MeOH +FA
Cold utility	3.63	MWh/t MeOH +FA
Electricity	1.97	MWh/t MeOH +FA

$$CO_2ConvR = \left(\frac{CO_2in - CO_2out}{CO_2in}\right)_{Reactor} \quad (1)$$

$$CO_2ConvP = \left(\frac{CO_2in - CO_2out}{CO_2in}\right)_{Process} \quad (2)$$

5.2. Economic Result

Table 3 summarizes the economic breakdown of plant investment and operation cost. The raw material and utility constitute the larger shares of the OPEX. The utility cost is majorly influenced by the compression of H_2 and CO_2, which is required to fulfill the solubilization and synthesis requirements for CO_2 and H_2, respectively. The net present values at different discount rates and formic acid prices are presented in Figures 2 and 3, respectively. At a discount rate of 10%, the project is not economically viable. Hence, a sensitivity analysis of discount rate from 4% to 10% was carried out to determine the discounted cash-flow rate of return (DCFROR, when NPV =0). From the result, a discount rate between 4-5% makes the project profitable. At this discount rate, a free cost of CO_2 would improve the NPV as only H_2 is the major contributor to the raw material cost since the ionic liquid cost is estimated on a biannual basis (low volatility). In figure 3, the price of formic acid was varied to observe the behavior of the NPV at a 10% discount rate. At NPV =0, the selling cost of formic acid is 0.935 USD/kg, which makes it the minimum selling point for the project to be viable at a 10 % discount rate.

Table 3: Estimated CAPEX, OPEX and revenues of simulated process

CAPEX	**USD**	**OPEX**	**USD**	**Revenue**	**USD/YR**
Purchase Equipment Cost	11,775,700	Raw Material Cost	4,089,956	Formic Acid @ 0.78	16,715,161
ISBL	15,308,410	Utilities	4,157,864	Methanol @ 0.5	1,827,864
OSBL	1,837,009	Operating Labour Cost	1,483,442		
Indirect costs (IC)	15,259,423	Other Manufacturing Cost	3,051,827		
Project Contingency	3,240,484				
Process Contingency	1,620,242				
Fixed Capital Investment (FCI)	37,265,569				
Working Capital (WC)	4,471,868				
Cost of Land	2,000,000				
Total Capital Investment (TCI)	43,737,436				

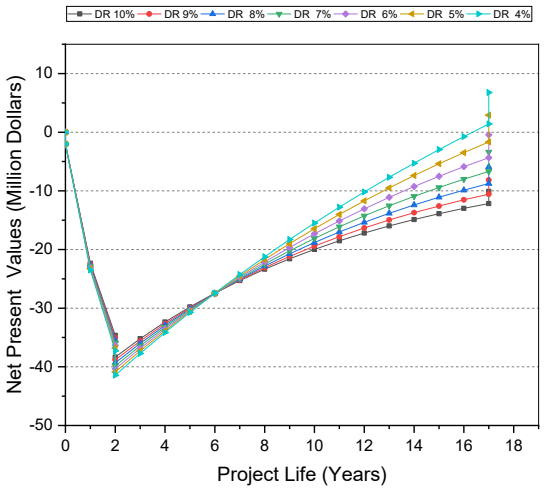


Figure 2: Cash flow diagram at different discount rates

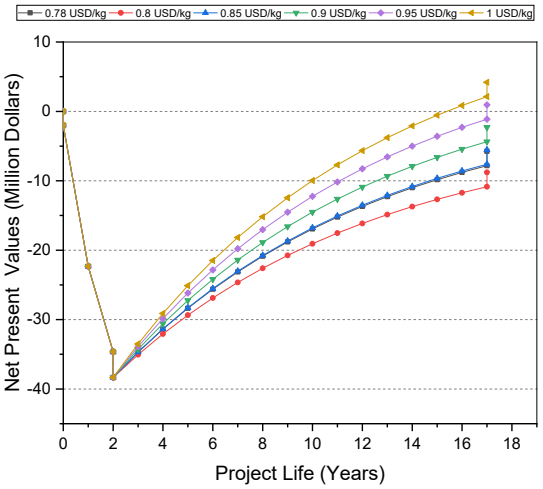


Figure 3: Cash flow diagram at different selling prices of formic acid

6. Conclusions

The techno-economic study based on process simulation has proven the economic feasibility of the hydrogenation of CO_2 promoted by [Edmim][NO_2] at a commercial scale. The results showed that the CAPEX and OPEX required are 43.9 MUSD and 12.7 MUSD, respectively. To ensure economic profitability, the calculated minimum selling cost of formic acid was 0.935-1 USD/kg. In addition, at the current 10% discount rate, the project is profitable with a carbon credit tax of 66 USD/tCO_2. The project's minimum payback time was calculated to be four years.

Acknowledgments

The authors gratefully acknowledge the support of the RCGI – Research Centre for Gas Innovation, hosted by the University of São Paulo (USP) and sponsored by FAPESP – São Paulo Research Foundation (2014/50279-4) and Shell Brazil. This study was financed in part by the Personnel Coordination of Improvement of Higher Level - Brazil (CAPES) - Finance Code 001

References

T.O. Bello, A.E. Bresciani, R.M.B. Alves, C.A.O. Nascimento, 2021a. Systematic Screening of Ionic Liquids for Hydrogenation of Carbon dioxide to Formic Acid and Methanol. Ind. Eng. Chem. Res.

T.O. Bello, A.E. Bresciani, C.A.O. Nascimento, R.M.B. Alves, 2021b. Thermodynamic analysis of carbon dioxide hydrogenation to formic acid and methanol. Chem. Eng. Sci. 242, 116731. https://doi.org/10.1016/j.ces.2021.116731

M. Ghavre, S. Morrissey, N. Gathergoo, 2011. Hydrogenation in Ionic Liquids. Ion. Liq. Appl. Perspect. https://doi.org/10.5772/14315

P. G. Jessop, T. Ikariya, R. Noyori, 1999. Homogeneous catalysis in supercritical fluids. Chem. Rev. 99, 475–494.

Kokorin, A., 2012. Ionic Liquids: Applications and Perspectives, Ionic Liquids: Applications and Perspectives. https://doi.org/10.5772/1782

W. Leitner, 1995. Carbon Dioxide as a Raw Material: The Synthesis of Formic Acid and Its Derivatives from CO<inf>2</inf>. Angew. Chemie Int. Ed. English 34, 2207–2221. https://doi.org/10.1002/anie.199522071

P.D.R. MacFarlane, D.M. Kar, D.J.M. Pringle, 2017. Fundamentals of Ionic Liquids Ionic Liquids in Biotransformations and Electrodeposition from Ionic Handbook of Green Chemistry – Green Solvents Electrochemical Aspects of Ionic Liquids , 2nd edition Nanocatalysis in Ionic Liquids.

G.A. Olah, A. Goeppert, G.K.S. Prakash, 2009. Chemical Recycling of Carbon Dioxide to Methanol and Dimethyl Ether: From Greenhouse Gas to Renewable, Environmentally Carbon Neutral Fuels and Synthetic Hydrocarbons. J. Org. Chem. 74, 487–498. https://doi.org/10.1021/jo801260f

M. Pérez-Fortes, E. Tzimas, 2016. Techno-economic and environmental evaluation of CO2 utilisation for fuel production. Synthesis of methanol and formic acid, Scientific and Technical Research Series. Luxembourg. https://doi.org/10.2790/89238

R. Schlögl, 2013. The Solar Refinery. In Chemical Energy Storage. Walter de Gruyter GmbH: Berlin, Germany, Boston, MA, USA.

J. Su, L.Yang, M. Lu, H. Lin, 2015. Highly efficient hydrogen storage system based on ammonium bicarbonate/formate redox equilibrium over palladium nanocatalysts. ChemSusChem 8, 813–816. https://doi.org/10.1002/cssc.201403251

G. Towler, R. Sinnott, 2013. Chemical Engineering Design Principles, Practice and Economics of Plant and Process Design, Second Edi. ed. Elsevier, Oxford. https://doi.org/10.1016/B978-0-08-096659-5.00022-5

G. Towler, R.K. Sinnott, 2012. Chemical engineering design: principles, practice and economics of plant and process design. Elsevier.

W. Wang, Y. Himeda, 2012. Recent Advances in Transition Metal-Catalysed Homogeneous Hydrogenation of Carbon Dioxide in Aqueous Media. Sch. Enviromental Sci. 250–264. https://doi.org/10.5772/48658

Y. Wang, M. Hatakeyama, K. Ogata, M. Wakabayashi, F. Jin, S. Nakamura, 2015. Activation of CO2 by ionic liquid EMIM-BF4 in the electrochemical system: a theoretical study. Phys. Chem. Chem. Phys. 17, 23521–23531. https://doi.org/10.1039/c5cp02008e

W. Xu, L. Ma, B. Huang, X. Cui, X. Niu, H. Zhang, 2011. Thermodynamic analysis of formic acid synthesis from CO2hydrogenation. ICMREE 2011 - Proc. 2011 Int. Conf. Mater. Renew. Energy Environ. 2, 1473–1477. https://doi.org/10.1109/ICMREE.2011.5930612

S. Zeng, X. Zhang, L. Bai, X. Zhang, H. Wang, J. Wang, D. Bao, M. Li, X. Liu, S. Zhang, 2017. Ionic-Liquid-Based CO<inf>2</inf> Capture Systems: Structure, Interaction and Process. Chem. Rev. 117, 9625–9673. https://doi.org/10.1021/acs.chemrev.7b00072

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50028-2

Synthesis of Distillation Sequence with Thermally Coupled Configurations Using Reinforcement Learning

Jaehyun Shim[a] and Jong Min Lee[a*]

[a]*School of Chemical and Biological Engineering, Institute of Chemical Processes, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea*
buzzinga@snu.ac.kr
**Corresponding Author's E-mail: jongmin@snu.ac.kr*

Abstract

Distillation column is a representative chemical process unit, which is the most popular choice to separate a multicomponent mixture into pure substances. Since a typical industrial process involves multiple distillation columns, it is important to obtain an optimal sequence to optimize energy consumption and separation performance. For this a large number of candidates have to be investigated in the optimization problem, while the number of possible sequences becomes larger when thermally coupled configuration is considered. In this study, reinforcement learning algorithm is applied to find an optimal sequence to avoid the computational burden of exhaustive in solving such large scale problems. Reinforcement learning searches for a solution in an evolutionary fashion via value function approximation in a limited region of the solution space. Case studies demonstrate the efficacy of reinforcement learning to find a nearly optimal solution for distillation sequence synthesis problems. The objective of the case studies is to derive distillation sequence which minimizes the total annual cost for separating five component mixtures. The result is that total annual cost of the configurations of distillation sequence designed using reinforcement learning were only about 2.5% larger than the optimal result obtained from mixed-integer nonlinear programming. This shows that reinforcement learning can find a nearly-optimal structure without exhaustive search.

Keywords: design, distillation column, thermally coupled, reinforcement learning, optimization

1. Introduction

Distillation column is an essential unit operation for multicomponent separation and the efficiency of separating multicomponent depends on the configuration of distillation sequences. Therefore, it is important to design distillation sequences to obtain an optimal sequence with high efficiency and a large number of candidates have to be investigated to find an optimal distillation sequence. However, the size of the search space increases rapidly with the number of components to be separated and grows exponentially when thermally coupled (TC) configurations are considered (Shah, V. H. et al., 2010). Moreover, continuous variables such as liquid and vapor flowrate for mass balance in the distillation column should be determined while the configuration that is a discrete decision is chosen simultaneously, which means synthesis of distillation column is a mixed integer problem (Gooty, R. T. et al., 2019). Thus, the exhaustive search approach is not effective for large size multicomponent separation problems and reinforcement

learning (RL) is proposed as an alternative framework to find an optimal distillation sequence in this study. RL approximates a value function of state via learning based on trial and error. Value function indicates how optimal the decision is and RL optimizes the objective function by outputs from value function. RL has two types of algorithm: value-based method and policy gradient method. Q-learning, SARSA, and deep Q network are well-known value-based methods and REINFORCE, actor critic, and deep deterministic policy gradient are representative policy gradient methods (Nian, R. et al., 2020). In this study, actor-critic algorithm is used to synthesize distillation sequences and case studies are implemented for 5 components separation problem including thermally coupled configurations. The objective of the problems is to minimize total annual cost of the distillation sequence. Finally, the results from RL are compared with those from mixed-integer nonlinear programming (MINLP) in order to analyze the ability of RL to optimize the distillation sequence.

2. Distillation sequence

If there exists difference between relative volatilities, a mixture having more than three components is separated through a train of several distillation columns. When it comes to mixture separation, types of split in a distillation column can be categorized into sharp and non-sharp splits. There are substances of which relative volatility is between that of light key (LK) component and heavy key (HK) component in a case of non-sharp split, whereas sharp split does not have such substances. For simplicity and clear presentation of the proposed concept, this study considers the sharp split only. TC configuration is also introduced to the optimization problem.

2.1. Thermally coupled configuration

Distillation process using conventional columns which includes heat exchangers such as condenser and reboiler shows an inherent inefficiency due to remixing an intermediate component which should be re-purified in the next column. Introducing TC configurations by removing heat exchangers, the inefficiency of the conventional columns can be improved with a side stream because it prevents remixing (Hernández, S. et al., 2003).

2.2. Fenske-Underwood-Gilliland method

Once a structure of distillation sequence is decided, the corresponding variables such as reflux ratio, column diameter, and flow rate of distillate and bottom stream are calculated via distillation system dynamics. In this study, Fenske-Underwood-Gilliland (FUG) method was used for calculating the variables instead of the rigorous method such as Aspen simulator. FUG method is based on the assumption that the relative volatility of the component is constant along the column and the molar overflow of the component is constant along the column. FUG method consists of Equations (1)-(4).

$$\sum_{i=1}^{N} \frac{\alpha_i f_i}{\alpha_i - \varphi} = F(1-q) \tag{1}$$

$$\sum_{i=1}^{N} \frac{\alpha_i \xi_i f_i}{\alpha_i - \varphi} = D(R_{min}+1) \tag{2}$$

$$N_{min} = \frac{\log\left[\left(\frac{\xi_{LK}}{1-\xi_{LK}}\right)\left(\frac{1-\xi_{HK}}{\xi_{HK}}\right)\right]}{\log(\alpha_i/\alpha_i)} \tag{3}$$

$$\frac{N_t - N_{min}}{N_{min}+1} = 0.75\left[1-\left(\frac{R-R_{min}}{R+1}\right)^{0.5668}\right] \tag{4}$$

where i indicates the component, α_i is the relative volatility, f_i is the feed flow rate, φ is the root of Underwood equation, F is the total feed flow rate of the column, q is the quality of the feed, ξ_i is the recovery fraction, D is the total distillate flow rate, R_{min} is the minimum reflux ratio, N_{min} is the minimum number of theoretical stages, N_t is the number of theoretical stages, and R is the actual reflux ratio. Eqs. (1) and (2) are Underwood equations, Eq. (3) is Fenske equation, and Eq. (4) is Gilliland equation (Fenske, 1932; Underwood, 1949; Gilliland, 1940).

2.3. Total annual cost

The objective of this study is to find the most economical distillation sequence and total annual cost (TAC) is used as a criterion for evaluating the economics. Therefore, the optimal distillation sequence has the minimum TAC. TAC consists of the capital cost and sum of the operation cost of distillation sequences. Operation cost includes column equipment investment, condenser equipment investment, and reboiler equipment investment. Column equipment investment is a function of D, R, and N_t, and condenser and reboiler equipment investment is a function of R and ξ_i. Therefore, TAC of a distillation sequence can be estimated with these parameters obtained from FUG method and its calculation formulas were referenced in Zhang, S. et al. (2018).

3. Reinforcement learning

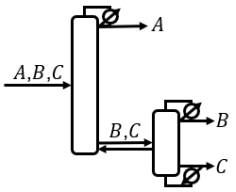
RL refers to a family of algorithms that learns the optimal value function that satisfies the optimality equation of dynamic program using either simulation or operational data. A decision-making entity called an agent takes an action based on the current state, the environment is changed by the action, and it gives the agent a reward as a feedback of the action. As a result, the agent learns the value function and policy in the state space, and its corresponding control policy maps the current state to a nearly-optimal action. Among various RL algorithms, this study employs the actor-critic algorithm since it can make a discrete decision and learn the policy directly with policy gradient method. Given a current state, the actor calculates an action using the learned policy function and the critic evaluates how beneficial the action is. The actor learns policy based on the evaluation from the critic and critic updates evaluations by rewards from environment (Konda, V. R. et al., 2000). The main challenge in this approach is to formulate the problem and define state, action, and reward.

3.1. State

Separation matrix representation (SMR) (Shah, V. H. et al., 2010) was employed in order to convert the topology of distillation sequence into a mathematical form which can be used in the RL algorithm. SMR is an upper triangular matrix as shown in Figure.1 and each element of the matrix means the stream in the sequence correspond to feed, distillate, or bottom flow of each column. 1 is assigned to each element if there exists a reboiler or

$$\begin{bmatrix} ABC & AB & A \\ 0 & BC & B \\ 0 & 0 & C \end{bmatrix}$$

Figure 1. Separation matrix representation for separating 3 components mixture.



(a)

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

(b)

Figure 2. an example of separation matrix representation for separating 3 components mixture.

condenser in the stream, while 2 is assigned if TC configuration exists. For instance, the sequence shown in Figure.2 (a) is represented as the matrix in Figure.2 (b). Additionally, temperature and flow rate of each component in the stream are converted into the matrix form based on the SMR indicating the sequence. As a result, all matrices are stacked and used in the current state.

3.2. Action

A distillation sequence is determined by the choice of which components are separated at which column and where TC configurations are located. For each column, what material is separated and whether there is a TC structure become actions in each column. One of the actions is choosing a HK substance, and the other is deciding if there is a TC structure. Accordingly, a stage is defined as deciding a HK substance and TC structure of each column and a stage-wise reward is described in section.3.3.

3.3. Reward

Since the objective function is TAC of a distillation sequence, the return in RL formulation is also TAC of a sequence, i.e. sum of TAC of all distillation columns. Therefore, TAC of each distillation column in a sequence is set to a reward because sum of the all rewards equals to the return as definition. In addition, negative value of TAC is used as the reward in algorithm so as to minimize total TAC because RL algorithms basically learn in the direction of maximizing the reward.

4. Results and discussion

A case study was implemented for confirming the ability of RL to design optimal distillation sequence with an arbitrary multicomponent mixture. Through the case study, the sequence was found to minimize TAC for separating 5 components mixture. For simplifying explanation in the case study, the stream flowing through the distillation

Table 1. Components and feed composition.

Case study 1	
Component	Mole fraction
ethanol (A)	0.25
n-propanol (B)	0.15
i-butanol (C)	0.35
n-butanol (D)	0.10
phenol (E)	0.15
The flowrate of the feed mixture is 500.4 kmol/h	

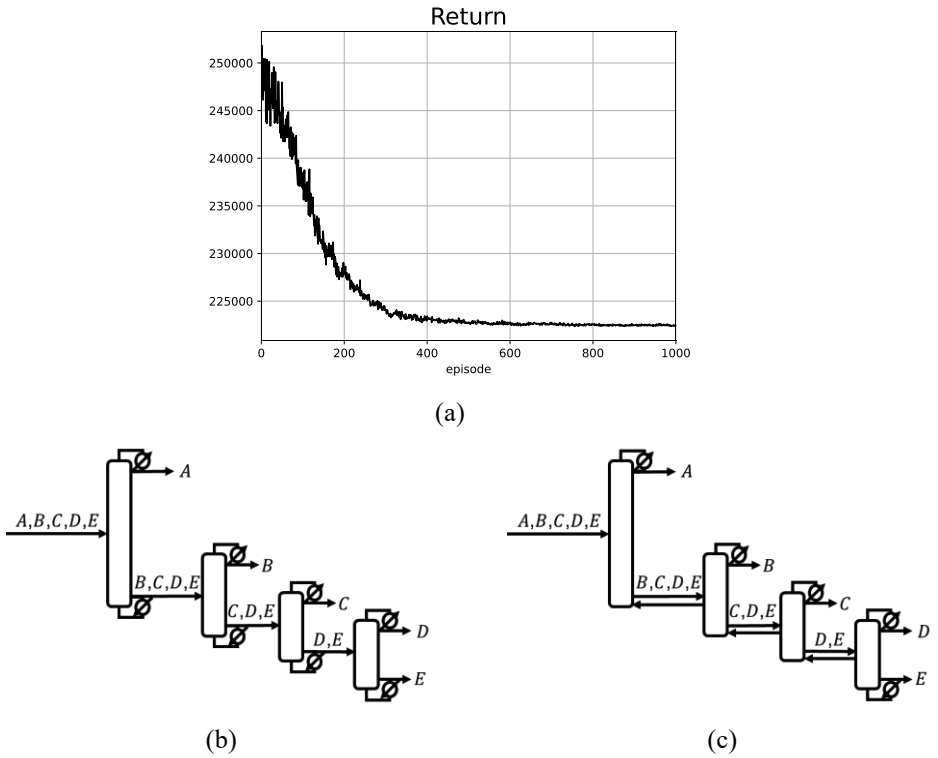


Figure 3. Return of RL and distillation sequences synthesized by RL and MINLP in case studies.

sequence is denoted as ABCDE, where each letter means a component in the stream and is assigned to a nature number to use in mathematical equations: A = 1, B = 2, ..., E = 5. The notation is arranged in order of relative volatility, for example, in a stream ABCDE, A is the most volatile component and E is the least volatile component. RL algorithm was carried out to find the distillation sequence minimizing TAC via trial and error as mentioned above, followed by comparing the result from learned RL and that from MINLP.

4.1. Case study

Components in the inlet mixture are ethanol, i-propanol, n-propanol, i-butanol, and n-butanol, corresponds to A, B, C, D, and E orderly, and their feed composition are

presented in Table.1. Figure.3 (a) shows return, i.e. TAC of a designed distillation sequence, decreased as episode progresses and converged to a certain level after about a thousand iteration. A distillation sequence was determined using the learned RL agent as shown in Figure.3 (b) and its TAC is 222,258.15 $/y, whereas Figure.3 (c) is the optimal sequence found by MINLP and has TAC of 216,747.68 $/y. The separation order of the two structures is same, so the difference in TAC between the sequences comes from TC configuration. TC configuration should be adopted for efficient separation of mixture according to MINLP, but RL was learned in a direction that does not consider the TC configuration. Nonetheless, TAC of the sequence from RL is only about 2.5% higher than that from MINLP, which means RL found the near-optimal structure. Moreover, any superstructure is not required when solving the problem of finding an optimal distillation sequence using RL.

5. Conclusion

Through this study, it was demonstrated that a near-optimal structure can be determined without a superstructure or any prior knowledge except for the material properties of the desired mixture. This is more beneficial when solving a large-scale optimization problem since full search becomes near impossible and time consuming as the scale increases. It can be proved by applying RL algorithm to a larger design problem, for instance, a separation problem for more than 5 components mixture.

Acknowledgements

This study was supported by Hanhwa Solutions Corporation.

References

Fenske, M., 1932, Fractionation of straight-run Pennsylvania gasoline., Industrial & Engineering Chemistry, 24, 5, 482-485.

Gilliland, E. R., 1940, Multicomponent rectification., Industrial & Engineering Chemistry, 32, 8, 1101-1106.

Gooty, R. T., Agrawal, R., & Tawarmalani, M., 2019, An MINLP formulation for the optimization of multicomponent distillation configurations., Computers & Chemical Engineering, 125, 13-30.

Hernández, S., Pereira-Pech, S., Jiménez, A., & Rico-Ramírez, V., 2003, Energy efficiency of an indirect thermally coupled distillation sequence., The Canadian Journal of Chemical Engineering, 81, 5, 1087-1091.

Konda, V. R., & Tsitsiklis, J. N., 2000,. Actor-critic algorithms. In Advances in neural information processing systems, 1008-1014.

Nian, R., Liu, J., & Huang, B., 2020, A review on reinforcement learning: Introduction and applications in industrial process control. Computers & Chemical Engineering, 139, 106886.

Shah, V. H., & Agrawal, R., A matrix method for multicomponent distillation sequences, AIChE journal, 56, 7, 1789-1775.

Underwood, A. J. V. , 1949, Fractional distillation of multicomponent mixtures., Industrial & Engineering Chemistry, 41, 12, 2844-2847.

Zhang, S., Luo, Y., Ma, Y., & Yuan, X., 2018, Simultaneous optimization of nonsharp distillation sequences and heat integration networks by simulated annealing algorithm., Energy, 162, 1139-1157.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50029-4

Optimal Design of Heat Integrated Reduced Vapor Transfer Dividing Wall Columns

Fanyi Duanmu and Eva Sorensen*

Department of Chemical Engineering, University College London (UCL), Torrington Place, London WC1E 7JE, United Kingdom
** Corresponding author: e.sorensen@ucl.ac.uk*

Abstract

A dividing wall column (DWC) is capable of saving capital costs and improving energy efficiency for ternary liquid separations. Alternative DWC structures have been proposed, termed Reduced Vapor Transfer DWC (RVT-DWC) in this work, which involves less difficult-to-control vapor transfer streams. The most interesting RVT-DWC structure, the LL structure, which has a dividing wall extending throughout the column and has no interconnected vapor transfer streams, is studied in this work. Three heat integrated designs of the LL structure, the LL structure with combined condenser and reboiler (LL-CCR), vapor recompression assisted LL structure (VR-LL), and vapor recompression assisted LL structure with combined condenser and reboiler (VR-LL-CCR), are introduced and compared to the standard DWC, standard LL structure, and vapor recompression assisted DWC (VR-DWC) designs, respectively. Although the LL-CCR structure shows only minor improvement in total annualized costs (TAC) when compared to the LL structure, its vapor recompression assisted design (VR-LL-CCR) has the lowest TAC among all the structures studied (17 % lower than LL-CCR, 4 % lower than VR-DWC, and 10 % lower than VR-LL). Moreover, the vapor recompression assisted structures have lower TAC than their corresponding base structures.

Keywords: Distillation, Dividing Wall Column, Optimization, Heat integration, Vapor recompression

1. Introduction

Process Intensification (PI) has received significant interest in recent years as a mean of achieving more energy efficient chemical processes. A prime example of PI is a dividing wall column (DWC) for the separation of ternary mixtures. Agrawal (2000) proposed several alternatives to a standard DWC, denoted as Reduced Vapor Transfer DWCs (RVT-DWCs) in this work, which involve less difficult-to-control vapor transfer streams, thus reducing the complexity of the unit. These RVT-DWCs have been claimed to be superior to a standard DWC in terms of controllability (Cui et al., 2020), and to have a similar economic performance (Agrawal, 2000) and almost identical energy demands (Waltermann et al., 2019), thus the RVT-DWCs can be considered as competitive alternatives to the standard DWC. Out of all the RVT-DWC structures, the LL structure (liquid-liquid structure, both thermal coupling streams replaced by liquid sidedraw streams, the dividing wall extended throughout the column, utilizing two condenser and two reboilers) has the most interesting structure, and is the structure that

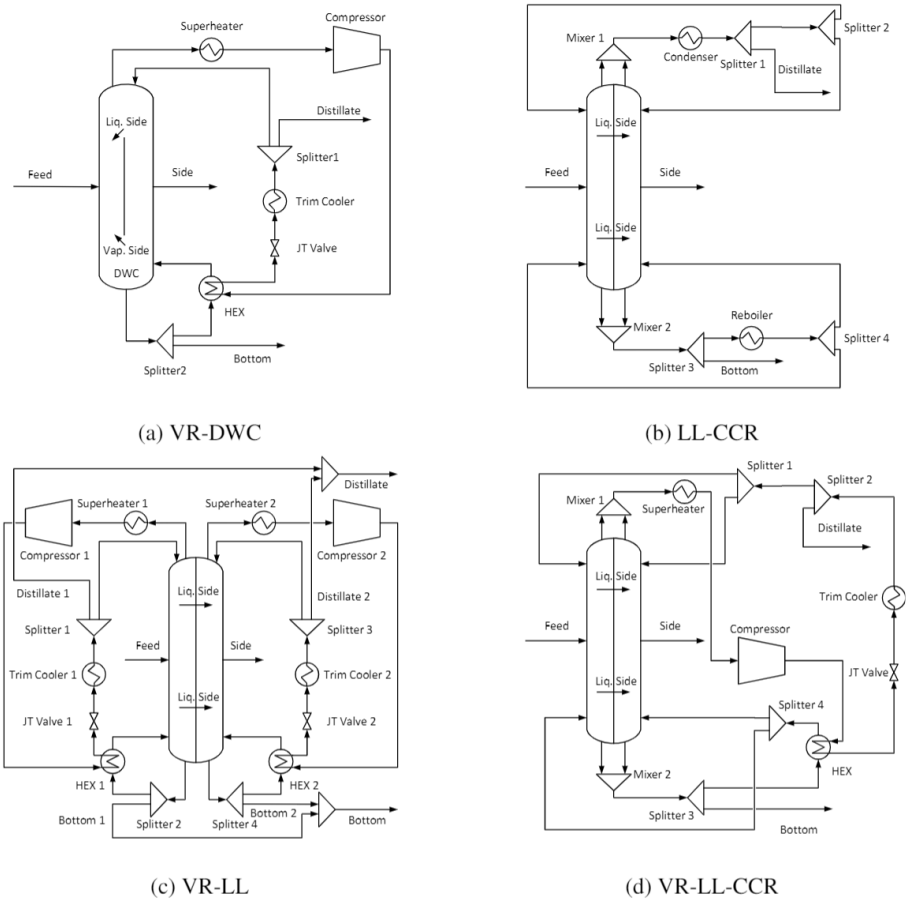


Figure 1: Flowsheets of (a) Vapor recompression assisted dividing wall column (VR-DWC), (b) LL structure with combined condenser and reboiler (LL-CCR), (c) vapor recompression assisted LL structure (VR-LL), and (d) vapor recompression assisted LL structure with combined condenser and reboiler (VR-LL-CCR).

will potentially provide the most improvement with heat integration. Heat integration by combining the condensers and reboilers in the LL structure may improve the economic performance (Ramapriya et al., 2014). Moreover, heat integration by vapor recompression can save both energy and cost, and its effectiveness for a standard DWC was considered by Xu et al. (2017), but its potential effectiveness for a RVT-DWCs has not yet been studied. This work therefore aims to investigate the economic performance of the LL structure by considering heat integration based on vapor recompression.

2. Methodology

In this work, six different structures are designed, optimized, and compared, which are the standard dividing wall column (DWC, not shown), vapor recompression assisted dividing wall columns (VR-DWC, Figure 1a), LL structure (LL, not shown), LL structure with combined condenser and reboiler (LL-CCR, Figure 1b), vapor

recompression assisted LL structure (VR-LL, Figure 1c), and vapor recompression assisted LL structure with combined condenser and reboiler (VR-LL-CCR, Figure 1d). It should be noted that, unlike the structures proposed by Agrawal (2000), in all the LL structures studied in this work, the product streams (distillate and bottom streams) are combined into single product streams, which also reduces the number of product constraints for optimization. For all vapor recompression assisted structures, the same structure is utilized including a superheater installed just after the top vapor stream from the column to improve the efficiency of the heat integration design (Yang et al., 2019). The superheated vapor is compressed in an isentropic compressor with 85 % efficiency. Then, a counter-current shell and tube heat exchanger (HEX) is used to vaporize the boilup flow using the recompressed vapor stream. After that, a Joule-Thompson valve (JT valve) and a trim cooler are used to lower the stream pressure back to the column pressure and to condense the stream, respectively. Finally, a splitter is used to control the product and reflux flowrates. For LL-CCR and VR-LL-CCR, two additional splitters are used to control the flowrate (split ratio) of reflux and boilup streams into the prefractionator (left)/main column (right).

All rigorous simulations (equilibrium based) are performed in gPROMS ProcessBuilder version 1.4 (Process Systems Enterprise, 2020). As there is no built-in column libraries for DWC and LL structures, their corresponding Petlyuk designs are used instead. All designs are optimized using both stand-alone particle swarm optimization (PSO) and a combined stochastic/deterministic optimization method similar to the one proposed by Chia et al. (2021), but using PSO instead of Genetic Algorithm (GA). PSO is coded in MATLAB (The MathWorks Inc., 2019) while OAERAP is built-in within gPROMS ProcessBuilder, with gO:MATALAB (Process Systems Enterprise, 2019) used to transfer data between MATLAB and gPROMS ProcessBuilder. The objective function is the total annualized cost (TAC) based on the summation of annualized capital cost of all equipment and of the operating cost. The operating hour is set as 8400 $h\ y^{-1}$ and the payback period is 8 y. The sizing equations of the column are from Seider et al. (2016) and cost equations and installation factors are from Sinnott and Towler (2020). High pressure steam is used as the heating utility with a cost of 24 $€\ t^{-1}$ and for electricity a cost of 23.5×10^{-6} $€\ kJ^{-1}$. All prices are converted to US dollars at the end for comparison. The design with the lowest TAC from both optimization methods is taken as the final results to ensure a good optimal design and fair comparison. For all the designs, the column pressure is maintained at 1 *bar* and not optimized. For the optimization task, all design and operating variables are optimized simultaneously including the number of stages, feed/sidedraw locations, reflux/boilup ratios, distillate/bottom flowrate, splitter ratio (used in LL-CCR and VR-LL-CCR), temperature in the superheater, and outlet pressure in the compressor. In terms of optimization constraints, other than the three product specifications, the number of stages on both sides of the wall is considered the same for the LL structures (although does not have to be), the vapor fraction of the stream from the compressor should be equal to one, and the heat exchanger inlet temperature difference should be greater or equal to the minimum temperature approach.

3. Case Study

The comparison of the economic performance of the various structures are based on the separation of an equi-molar benzene/toluene/o-xylene (0.33/0.34/0.33) mixture, with

UNIQUAC as the thermodynamic model. The feed is supplied at 1000 *kmol* h^{-1} as a saturated liquid at 1 *bar*, which is the same as the operating pressure in the column with no pressure drop. Calculations are performed using an AMD Ryzen 9 3900X CPU with 3.79 GHz and 64 GB memory. For PSO, the parallel computing function in MATLAB was activated with 18 workers to speed up the optimization, and it takes about 1 to 3 hours to perform the optimization depending on the complexity of the model.

The key design and operating variables for all structures are shown in Table 1. In general, all structures have similar total number of stages (48-52). For the main column, the LL structures have lower reflux ratio compared with DWC structures as the majority of the light and heavy components are removed from the system from the prefractionator, which makes the separation in the main column easier. By comparing the vapor recompression assisted structures with their corresponding base designs, no significant changes of design and operating variables are found, which indicates that the retrofit of the base designs can easily be achieved without changing the column structures. Considering the energy usage of designs without vapor recompression, they have similar total reboiler/superheater and condenser/cooler duties. The LL-CCR structure does not improve the energy efficiency (LL duties similar to LL-CCR duties). The vapor recompression assisted designs require about 60 % less total energy (steam plus electricity), and there are significant savings in the reboiler/superheater and condenser/cooler duties (about 80 % and 70 %, respectively) when compared to the base designs. Out of the three vapor recompression assisted structures, VR-LL and VR-LL-CCR have similar energy consumption, slightly lower than the energy required by VR-DWC. It is worth noting that there are two vapor recompression systems in the VR-LL structure (one at each side of the wall, Figure 1c), but the equipment duties in VR-LL is very close to the VR-LL-CCR which has one vapor recompression system (Figure 1d). Breaking down the equipment duties in VR-LL (not shown), the equipment at the prefractionator side requires more energy (e.g., 1.75 *MW* for compressor on the prefractionator side vs 0.94 *MW* on the main column side) due to the removal of the majority of the light and heavy components in the prefractionator.

The cost information of all the designs is shown in Table 1. Compared to DWC and VR-DWC, all variations of LL structures have lower capital costs (CAPEX) for distillation columns (inclusive of column shell and trays) as the removal of products (distillate and bottom streams) from the prefractionator leads to a smaller column diameter. Considering the total CAPEX of each design, all vapor recompression assisted designs have significantly larger cost (e.g., VR-LL CAPEX is 2.61 times of LL CAPEX) due to the high compressor cost (e.g., in VR-LL structure, the compressors contributes 66 % of total CAPEX). It should be noted that, although VR-LL and VR-LL-CCR require similar total compressor duty, VR-LL uses two separate vapor recompressor systems (i.e., two compressors), thus the compressor CAPEX is higher (about 44 %). By considering the operating cost (OPEX) of each design, the OPEX for the vapor recompression assisted structures are about 40 % lower than their corresponding base structures. The comparison of the total annualized cost (TAC) shows that the standard DWC and standard LL structure have similar TACs (DWC 1 % more expensive). Compared with LL, LL-CCR (LL with combined condenser and reboiler) shows only very minor improvement (1 % lower). More importantly, all vapor recompression assisted structures achieved significant saving in TAC when compared with their corresponding base structures (15 % savings in VR-DWC, 8 % savings in

Table 1: Key design and operating parameters of all the structures

Items	DWC	VR-DWC	LL	LL-CCR	VR-LL	VR-LL-CCR
Prefractionator						
Total stages	26	30	52	49	52	48
Feed stage	14	14	29	24	28	27
Liq. sidedraw stages	-	-	11/38	9/36	10/40	9/37
Liq. side ($kmol\ h^{-1}$)	-	-	283/283	266/311	275/313	266/304
Distillate ($kmol\ h^{-1}$)	-	-	211	-	204	-
Molar reflux ratio	-	-	2.78	-	2.72	-
Main Column						
Total stages	48	48	52	49	52	48
Feed stages	8/35	7/38	9/40	8/36	8/38	8/35
Side prod. stage	21	18	25	20	21	19
Side prod. ($kmol\ h^{-1}$)	337	338	337	337	338	337
Liq. side ($kmol\ h^{-1}$)	231	239	-	-	-	-
Vap. side ($kmol\ h^{-1}$)	690	692	-	-	-	-
Distillate ($kmol\ h^{-1}$)	333	331	122	332	127	332
Molar reflux ratio	2.58	2.66	2.16	-	2.31	-
Mass Split Ratio to Main Column						
Reflux flowrate	-	-	-	0.36	-	0.35
Boilup flowrate	-	-	-	0.35	-	0.34
Vapor Recompression System - Pre/Main						
Superheater temp. (K)	-	-/409	-	-	413/404	-/410
Compressor pres. (bar)	-	-/5.22	-	-	5.21/5.12	-/5.16
Total Duty/Power (MW)						
Reboiler/Superheater	10.93	2.02	10.84	10.86	1.87	1.98
Condenser/Cooler	10.28	3.73	10.19	10.2	3.61	3.64
Compressor	-	2.78	-	-	2.69	2.71
CAPEX (M \$)						
Column	5.155	5.2007	4.7378	4.4081	4.7484	4.3047
Reboiler/Heater	1.4325	0.3363	1.6428	1.4245	0.4379	0.3318
Condenser/Cooler	1.5797	0.5602	1.5273	1.5669	0.6244	0.5468
Compressor	-	9.4839	-	-	13.5107	9.3629
Heat Exchanger	-	1.2427	-	-	1.2968	1.324
Total CAPEX	8.1672	16.8238	7.9079	7.3995	20.6182	15.8702
OPEX (M \$ y^{-1})						
Steam	5.242	0.9696	5.1995	5.2061	0.9487	0.9526
Electricity	-	2.2337	-	-	2.165	2.1746
Total OPEX	5.242	3.2033	5.1995	5.2061	3.1137	3.1272
TAC (M \$ y^{-1}) *	6.2629	5.3063	6.1879	6.131	5.691	5.111

* Operating hours = 8400 $h\ y^{-1}$; Payback period = 8 y

VR-LL, and 17 % savings in VR-LL-CCR). The comparison between all vapor recompression assisted structures shows that the VR-LL-CCR design is the best (4 % and 10 % lower when compared to VR-DWC and VR-LL, respectively). It should be noted that the improvement of the vapor recompressor assisted structures is indeed a trade-off between the increased capital cost of compressor and energy saved, thus the improvement is dependent on the payback period (here assumed to be 8 years).

4. Conclusions

This work introduces three types of heat integrated reduced vapor transfer dividing wall designs (LL-CCR, VR-LL, and VR-LL-CCR) and compares them with the standard DWC, standard LL, and VR-DWC. It was found that the LL-CCR shows only minor improvement of Total Annualized Cost (TAC) compared with LL but its vapor recompression assisted design (VR-LL-CCR) has the lowest TAC (17 % lower than LL-CCR, 4 % lower than VR-DWC, and 10 % lower than VR-LL). Moreover, all vapor recompression assisted designs have lower TAC compared to their corresponding base designs. The improvement is, however, dependent on the payback period as it is a trade-off between the more expensive compressor and the energy saved from the vapor recompression system. It should be noted that the steam and electricity prices will affect the economic performances of the VR designs. A more comprehensive study which includes the sensitivity analysis of utility prices will be carried out in the future. Besides, the VR designs are more complex than the standard structures, which may be more difficult for operation. Thus, the controllability study will be performed as the future work.

References

Agrawal, R. (2000). Thermally coupled distillation with reduced number of intercolumn vapor transfers. *AIChE Journal*, 46(11):2198–2210.

Chia, D. N., Duanmu, F., and Sorensen, E. (2021). Optimal Design of Distillation Columns Using a Combined Optimisation Approach. In Turkay, M. and Gani, R., editors, *31st European Symposium on Computer Aided Process Engineering*, pages 153–158. Elsevier B.V.

Cui, C., Zhang, Q., Zhang, X., and Sun, J. (2020). Eliminating the vapor split in dividing wall columns through controllable double liquid-only side-stream distillation configuration. *Separation and Purification Technology*, 242(December 2019):116837.

Process Systems Enterprise (2019). gO:MATLAB.

Process Systems Enterprise (2020). gPROMS ProcessBuilder version 1.4.

Ramapriya, G. M., Tawarmalani, M., and Agrawal, R. (2014). Thermal coupling links to liquid-only transfer streams: A path for new dividing wall columns. *AIChE Journal*, 60(8):2949–2961.

Seider, W. D., Lewin, D. R., Seader, J. D., Widagdo, S., Gani, R., and Ng, K. M. (2016). *Product and Process Design Principles: Synthesis, Analysis and Evaluation*. Wiley, 4 edition.

Sinnott, R. and Towler, G. (2020). *Chemical Engineering Design*. Elsevier, 6 edition.

The MathWorks Inc. (2019). MATLAB R2019b version 9.7.

Waltermann, T., Sibbing, S., and Skiborowski, M. (2019). Optimization-based design of dividing wall columns with extended and multiple dividing walls for three- and four-product separations. *Chemical Engineering and Processing - Process Intensification*, 146(August):107688.

Xu, L., Li, M., Yin, X., and Yuan, X. (2017). New Intensified Heat Integration of Vapor Recompression Assisted Dividing Wall Column. *Industrial & Engineering Chemistry Research*, 56(8):2188– 2196.

Yang, A., Jin, S., Shen, W., Cui, P., Chien, I.-L., and Ren, J. (2019). Investigation of energy-saving azeotropic dividing wall column to achieve cleaner production via heat exchanger network and heat pump technique. *Journal of Cleaner Production*, 234:410–422.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50030-0

A Model-Data Driven Chemical Analysis System for Products and Associated Processes

Sultana Razia Syeda[a]*, Easir A Khan[a], Nichakorn Kuprasertwong[b], Orakotch Padungwatanaroj[b], and Rafiqul Gani[b,c]

[a]*Department of Chemical Engineering, Bangladesh University of Engineering and Technology,(BUET), Dhaka-1000*
[b]*PSE for SPEED Company, 294/65 RK Office Park, Romklao Rd., Bangkok, Thailand*
[c]*PSE for SPEED Company, Ordrup Jagtvej 42D, DK-2920 Charlottenlund, Denmark*
**syedasrazia@che.buet.ac.bd*

Abstract

Currently, more than one million chemicals can be found on planet earth and thousands of new chemicals-based products are entering the global market every year. Many of the chemicals used in these products serve specific functions and are therefore included in the synthesized product. Some of these chemicals, however, could have harmful hazardous effects, while for others, better functioning alternatives may be available. Therefore, an analysis-based method to identify and substitute chemicals that are classified as hazardous or may have lower economic potential is needed. This paper presents a model-data based chemical analysis method for chemicals-based products and their associated processes. A large database of chemicals has been developed to identify hazardous chemicals. A link to a library of property models has been established to fill out gaps in measured property data. A flexible work-flow for analysis has been developed to identify and substitute chemicals within the product and/or its associated process. The main concepts and tools are highlighted through two case studies.

Keywords: Chemical products; Hazardous chemicals; Chemical substitution; Health hazards, Environmental hazards, Physical hazards

1. Introduction

Today, we are living with chemicals that are in our food, clothes, furniture, appliances, toys, cosmetics and medicines. Society, for its existence anywhere on earth, needs to use a variety of products and/or means that are directly or indirectly connected to chemicals. For example, from the time one wakes up in the morning to the time one goes to sleep, one may use products that are directly connected to chemicals, such as tooth-paste, soap, drugs, preserved milk or juice, perfume, creams for skin-care, and many more. Other chemicals-based products (to be called chemical products) indirectly influence our activities, such as cooking oil, paint, gasoline, fuel for cooking, electricity, etc., while others affect our survival, such as, air we breathe, water we drink, water we use for cleaning, soil we use for various purposes, to name a few.

Over 95 percent of all manufactured goods rely on some form of industrial chemical processes (ICCA, 2019). Chemicals are also a significant contributor to our economies. World chemicals sales were valued at €3,669 billion in 2019. Nearly 500,000 chemical substances from CAS REGISTRY® cover areas of community interest and an estimated 40,000 to 60,000 industrial chemicals are found in commerce globally. As the number of

chemicals grows rapidly, understanding their implications on human health and environment is increasingly becoming a problem. An important and urgent challenge is not only to identify the chemicals, which may have harmful effects but also to substitute and/or control their use. What is needed is an intelligent chemical analysis-substitution system. Note, however, while reducing the use of hazardous chemicals is a primary goal of this system, replacing substances without proper assessment of the alternatives can lead to regrettable substitutions (Hogue, 2013). Regulated chemicals subject to phasing out should be urgent candidates for substitution. For example, Per- and Polyfluoroalkyl Substances (PFAS) are restricted by EU POPs regulation due to their bio persistence and bio accumulative nature (Cousins *et al*., 2019). In addition to regulations, consumer comfort, awareness or economics may also act as driving factors for the substitution of chemicals in many products. Every product or chemical is related to its manufacturing process and life cycle assessment could point to the need for substitution of hazardous chemicals also used in processing. In order to substitute hazardous chemicals in a process, the main functional role of the chemical in the process must be specified along with a complete process description and process operating parameters (Jhamb *et al.,*2018).

Demand for safer alternatives in products is increasing continuously and regulatory authorities, such as EU REACH (EU, 2021), US EPA (Harten, 2014) and Occupational Safety and Health Administration (OSHA, 2021) have taken up substitution of chemicals that are harmful to human health and environment as one of the central elements of their policies. Avoidance of hazardous chemicals in the processes is recommended by Control of Major Accident Hazards Regulations (HSE, 2015). Different international organizations, research institutes and state/provincial governments have proposed frameworks for the assessment of alternative substitute candidates in a product and/or process. These frameworks are mostly case specific and limited to prescribed classes of chemicals. Even though several databases are available, most of them are incomplete and have information gaps that need to be filled. An overview of assessment frameworks, methods and associated tools (such as databases) are discussed in Syeda *et al.* (2021). An important unresolved issue is the accessibility of the data needed to perform analysis for different types of chemical products.

In this paper, aspects of a systematic analysis method for chemical substitution related to chemical products and their associated processes are presented. The method is based on collected data stored in structured databases, a suite of property model libraries, and a work-flow applicable for a wide range of chemical substitution problems. Selected features of the chemical analysis method are highlighted through two case studies.

2. Database and Methodology

The developed (ChemSub) database consists of 3 knowledge sections (chemical product classes, chemical properties, hazard data). For each product class, examples of known products in terms of product classification (single molecule, blend-formulation, functional and device) along with chemical identity, compositions, chemical functions, etc., are also stored. Within each class, the products are further divided into sub-classes. Data of 189 product sub-classes, such as refrigerants, dyeing additives, solvents, adsorbents, etc., have been collected. In the chemical properties section, collected data is divided and stored in terms of pure component property, functional pure component property, functional mixture property and phase equilibrium related properties. For the hazard data section, chemicals are classified according to three types of hazardous effects. A simplified version of the ontology, implemented for knowledge representation in the created library of databases is highlighted in Figure 1, where the chemical identity is the

link to each section of the database. This knowledge representation allows easy retrieval of all information (properties, hazardous effects, product classes) of an identified chemical. A reverse search, that finds, for example, chemicals according to their product class and then determines, which of them have unacceptable properties and/or hazardous effects is also allowed. Currently, as Fig 1 indicates, there are 919823 chemicals in all the databases, there are 3 hazardous effects as defined by Globally Harmonized System (GHS) hazard classes and associated hazard category, under which, there are 10 physical, 2 environmental and 9 health hazardous effects. Within pure component properties, 17 properties are covered (in addition to functional and mixture properties).

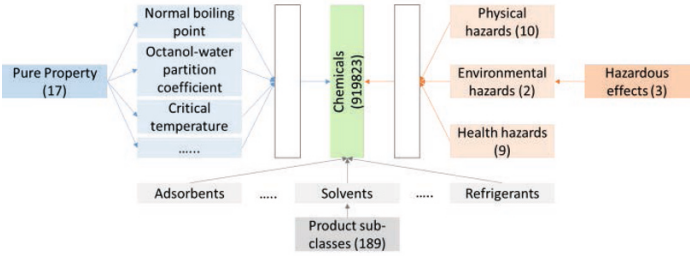


Figure 1: The ontology for knowledge representation in the database

The main steps of the work-flow for chemical substitution, developed from the experience of solving numerous substitution problems, are highlighted in Figure 2. The starting point (step-1) is problem definition, where details of the product are given in terms of product classification, product functions, the identity of chemicals involved, their compositions and many more. In step-2, the identified product is analysed in terms of important properties and their values. Here in-house property estimation models available in the ProCAPE toolbox (ChemSub, 2021) is used. In step-3, the chemical species are further analysed in terms of hazardous properties and/or effects, by retrieving the relevant hazardous effects data for each chemical through the ChemSub software (ChemSub, 2021). In steps 4-5, substitutes that match the desired target values for substitution are identified through the ProCAMD toolbox (Kalakul *et al.*, 2018). That is, chemicals that have the desired product function properties and do not have the undesired hazardous effects are identified. In step-6, the top ranked substitution candidates are listed for further verification of their properties and functions with experiments. An option for the design of experiments for selected products is also available. Note that each of the 6 steps requires its own specific computational methods and tools as highlighted in Fig 2.

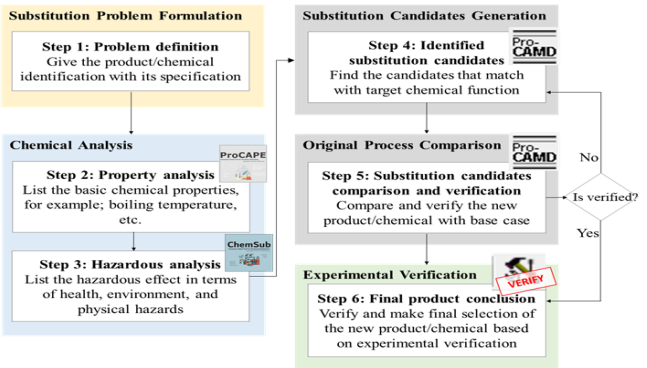


Figure 2: Work-flow for analysis based chemical substitution

3. Case studies

Results from two selected case studies are given below to highlight the main concepts of chemical substitution. These two case studies are selected because they exhibit features that are common to many chemical products. Note that the work-flow outlined in Fig 2 has been found to be applicable for a large number of products that have been analysed.

3.1. Mosquito Repellent

The mosquito repellent product is used to highlight the substitution of the active ingredient (N, N-Diethyl-meta-toluamide or diethyltoluamide, known as DEET), which is widely used in different versions of this product. DEET repels mosquitos by blocking the neuron to smell humans. Currently, insect repellents based on DEET are available in different product types; liquid formulations, lotions or sprays. Table 1 gives an example of the DEET based mosquito repellent product. It can be noted that in addition to DEET, a solvent is used to dissolve DEET, which is solid at standard conditions. Also, additives such as acidic acid (for pH adjustment) and Linalool (for fragrance) are used. The hazardous effects of each compound in the product are also given in Table 1.

Table 1: Typical formulation of a DEET based mosquito repellent product

Chemical	% Weight	Health hazard (Category)	Environmental hazard (Category)	Physical Hazards (Category)
DEET	10	Skin irritant (2), Eye irritant (2) Acute Toxicity (4)	Aquatic Chronic (3)	-
Acetic Acid	0.11	Skin Corrosive (1), Eye Damage (1)	-	-
Iso-propanol	41.8	Skin irritant (2), Specific Target Organ Toxicity STOT(SE) (3)	-	Flammable liquid (2)
Linalool	0.10	Skin irritant (2), Eye irritant (2), Skin Sensitizer (1)	-	-

Note: The category effect numbers in parenthesis range from 1 (very serious) to 4 (least serious)

From the analysis (step 3), it is found that DEET may cause irritation of the skin and eye. Search of the ChemSub database identifies picaridin, methyl nonyl ketone and ethyl 3-(N-butylacetamido) propionate as alternatives, which do not have any restriction from EU or US EPA. Natural compounds like Citronella oil and Eucalyptus oil extracted from plants are also found to be alternatives for DEET. The ChemSub database also helps to check the hazardous effects of the alternative active ingredients, which are listed in Table 2. From Table 2, Picaridin appears to be a safe alternative and is known to be very effective against mosquitoes by forming a vapor barrier on the skin surface leading to difficult landing. The formulation of picaridin based mosquito repellents contains 20% by weight of picaridin with the remaining being solvents and additives (Conte et al., 2011).

3.2. Textile finishing agent

Formaldehyde resins are used in the textile industry, as a finishing agent, to stiffen clothes and make fabrics, such as wrinkle-free cotton, rayon and corduroy. Different hazardous effects of formaldehyde retrieved from the ChemSub database are listed in Table 3. Due to the off-gassing and emission issues of formaldehyde from clothing, replacements are desired. Alternatives that are showing some efficacy in textile finishing include glyoxal, butane tetracarboxylic acid (BTCA) and citric acid. An important issue in the handling of these chemicals is related to their processing requirements along with their functionality in the product. Table 4 lists the limitations, hazardous effects, and processing requirements of some alternatives found in the ChemSub database. The most effective

finishing agent among the three is found to be BTCA, which is less hazardous than formaldehyde but is more expensive as larger amounts are required. On the other hand, citric acid is a cost-effective and environment friendly alternative, but it impacts the fabric adversely. More candidates need to be investigated and needed data need to be measured.

Table 2: Hazardous effects of alternatives to DEET in mosquito repellents

Chemical	**Health hazard (Category)**	**Environmental hazard (Category)**	**Physical Hazards (Category)**
Picaridin	Acute Toxicity (4)	Slight aquatic toxicity (4)	Flammable liquid (3)
Methyl Nonyl Ketone	Dermal toxicity; eye and dermal irritation (3)	Aquatic Chronic (1)	Not classified
Ethyl-P	Eye irritant (2)	-	-
Citronella Oil	Skin irritant (2); Eye damage (1); Skin sensitizer (1)	Aquatic Chronic (2)	Flammable liquid (2)
Eucalyptus oil	Skin (2) & eye irritant (4); Aspiration (1); Skin sensitizer (1); Acute toxicity (4);	-	Flammable liquid (3)

Note: Ethyl-P is short for Ethyl 3-(N-butyl acetamido) propionate

Table 3 Hazardous effects Formaldehyde found in textile finishing agent

Chemical	**Health hazard (Category)**	**Environmental hazard (Category)**	**Physical Hazards (Category)**
Formaldehyde	Oral toxicity (3), Dermal toxicity (3), Skin Sensitization (1), Eye irritant (1), Inhalation toxicity (1), Respiratory Sensitization (1), Organ toxicity (3), Germ cell mutagenicity (2), Carcinogenicity (1), Reproductive toxicity (1)	Flammable substance (1), Compressed gas	Acute aquatic environment hazard (2), Long-term aquatic environment hazard (3)

Table 4: Alternative anti-crease finishing agents

Alternatives	**Limitations/ disadvantages**	**Hazards (Category)**	**Processing requirements**	**Remarks**
Glyoxal	Low mechanical strength; fabric discolouring	Skin irritant (2), Eye irritant (2), Skin sensitization (1), Germ cell mutagenicity (2), Respiratory tract irritation (3)	Aluminium sulphate is used as a catalyst and glycols are used as additives	Processing is expensive and may have corrosion issues
Butane tetracarboxylic acid (BTCA)	Low mechanical strength	Acute Toxicity (4), Skin irritant (2), Eye irritant (2)	Sodium hypophosphite (NaH_2PO_2) is used as additive	High anti-crease capacity, tensile strength; imparts satisfactory whiteness, high cost
Citric acid	Fabric discolouring, low resilience	Eye irritant (2)	Nitrogenous additives	Cost-effective; environmentally friendly

4. Conclusions

The main components of a data-model based analysis method for chemical substitution related to chemical products and their associated processes have been presented. Two examples are given as representatives of many similar problems studied to establish the method. For example, replacement of solvents in products as well as processes,

refrigerants, additives as stabilizer, and many more. The implemented ontology in the ChemSub database, the property model library and the suite of algorithms for different steps define the application range, which is being extended continuously. Even though over 910000 chemicals are listed in the ChemSub database, actual measured property data exist for less than 10% of the chemicals. There is, therefore, a need to establish integrated and accessible databases based on reliable experimental data and accompanying property models to fill the gaps. Also, although the situation is improving, barriers to the implementation of green chemistry principles still exist (Matus *et al.*, 2012). An opportunity exists for the development of machine learning based property models with a wide application range as well as a new class of integrated computer-aided methods and associated tools (Pistikopoulos *et al*., 2021). That is, an opportunity exists for the development of more intelligent search methods to identify better and safer alternatives. Product oriented knowledge representation in the ChemSub database and a framework for chemical substitution are proposed to overcome the case specific nature of existing tools. In case of substitution in processes, computer aided tools for chemical substitution need to be integrated with process simulators to generate necessary information on process parameters. Finally, disclosure of product components by manufacturers and traders is a precondition for successful chemical substitutions

References

ChemSub, 2021, Computer-aided tool for chemical substitution, PSE for SPEED Company Ltd., Bangkok, Thailand.

E. Conte, R. Gani, K. M. Ng, 2011, Design of formulated products: a systematic methodology, AIChE Journal, 57, 2431-2449.

I. T. Cousins, G. Goldenman, D. Herzke, R. Lohmann, M. Miller, C. A. Ng, S. Patton, M. Scheringer, X. Trier, L. Vierke, Z. Wang, J. C. DeWitt, 2019, The concept of essential use for determining when uses of PFASs can be phased out. Environmental Science: Processes & Impacts, 21, 1803-1815.

EU, 2021, European Chemicals Agency (https://echa.europa.eu/).

P. F. Harten, 2014, Program for Assisting the Replacement of Industrial Solvents (PARIS III), 18th Annual Green Chemistry & Engineering Conference, North Bethesda MD, June 17-19.

HSE, 2015, (https://www.hse.gov.uk/comah/background/comah15.htm).

C. Hogue, 2013, Assessing Alternatives to toxic chemicals: governments, businesses, and now the National Academy of Sciences consider how to avoid “regrettable substitution”, Chemical Engineering News, 91, 19–20.

ICCA, 2019, (https://icca-chem.org/wp-content/uploads/2020/10/Catalyzing-Growth-and-Addressing-Our-Worlds-Sustainability-Challenges-Report.pdf).

S. Jhamb, X. Liang, R. Gani, G. M. Kontogeorgis, 2019, Systematic Model-Based Methodology for Substitution of Hazardous Chemicals, ACS Sustainable Chemical Engineering, 7, 7652–7666.

S. Kalakul, L. Zhang, Z. Fang, H. A. Choudhury, S. Intikhab, N. Elbashir, M. R. Eden, R. Gani, 2018, Computer aided chemical product design – ProCAPD & tailor-made blended products, Computers & Chemical Engineering, 116, 37-55.

K. J. M. Matus, W. C. Clark, P. T. Anastas, J. B. Zimmerman, 2012, Barriers to the Implementation of Green Chemistry in the United States, Environ. Sci. Technol. 46, 10892–10899.

OSHA, 2021, (https://www.osha.gov/safer-chemicals/why-transition).

E. N. Pistikopoulos, A. Barbosa-Povoa, J. H. Lee, R. Misener, A. Mitsos, G. V. Reklaitis, V. Venkatasubramanian, F. You, R. Gani, 2021, Process Systems Engineering–The Generation Next? Computers & Chemical Engineering, 147, 107252.

S. R. Syeda, E. A. Khan, O. Padungwatanaroj, N. Kuprasertwong, A. K. Tula, 2022, A perspective on hazardous chemical substitution in consumer products; Current Opinion in Chemical Engineering, 36, 100748.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50031-2

Construction of Database and Data-driven Statistical Models for the Solubility of Nanomaterials in Organic Solvents

Junqing Xia* and Yoshiyuki Yamashita

Department of Chemical Engineering, Tokyo University of Agriculture and Technology, Naka-cho, Koganei City, Tokyo, 184-8588, JAPAN
xiajunqing@m2.tuat.ac.jp

Abstract

Nanomaterials have been put into practical use in many fields. Therefore, the ability to predict the properties of nanomaterials has gained utmost significance. In this study, we have presented a database of solubility of organically-modified and non-modified nanomaterials in organic solvents. Furthermore, we attempted to model the solubility of one collected nanoparticle using various data-driven statistical modeling techniques. The solubility prediction using data-driven models is more accurate than that of the conventional solute-solvent similarity method based on Hansen solubility parameters (HSPs). In addition, the modeling results exhibit that certain solvent features, such as dielectric constant and molar logP, also have a significant influence on the solubility of nanomaterials.

Keywords: process science; machine learning; organically-modified nanomaterial; nanomaterial database; nanomaterial solubility.

1. Introduction

A nanoparticle has been widely accepted as a particle of any shape with dimensions in the range of 1 and 100 nm (Vert et al., 2012). These nanometer-sized materials have been applied in various fields such as medicine, electronics, materials, and chemical engineering. To retain their unique physical and chemical properties, the nanoscale structures of individual nanoparticles must be protected from undesirable interactions with other substances, such as solvents or polymers. One of the many effective protection strategies involves modifying the surface of nanoparticles with organic substances (Tomai et al., 2021), which further introduces unique solubility and cohesiveness behaviors. Therefore, from an industry perspective, the ability to predict the physical and chemical properties of modified and non-modified nanoparticles is becoming increasing vital, as it significantly impacts the design, construction, and operations of the manufacturing processes.

Recently, machine learning techniques have been introduced in the research of molecular and material sciences, for designing new compounds and synthesis routes, as well as revealing new principles hiding behind phenomenon (Butler et al., 2018). To construct and evaluate such machine learning models, training data must be collected first. We believe that to estimate the solubility or cohesiveness of nanoparticles, features such as Hansen solubility parameters (HSPs), UV shift, and Z potential are equally relevant. However, such data have not been previously collected or organized.

In this study, we introduced a new database for the solubility of nanomaterials in various solvents, which is currently at its early development stage with limited data points. In addition, we applied several data-driven statistical techniques to model the relationship between solubility and solvent feature parameters for one nanoparticle.

2. Solubility database for nanomaterials

With the growing numbers of practical use of nanomaterials, various databases have been established to collect the unique properties of these materials, such as caNanoLab, Dortmund data bank (DDB) nanofluids, eNanomapper, NANoREG, NanoDatabank, NanoMILE, and PubVINUS. For example, DDB nanofluids contain experimental data on the thermophysical properties of nanofluids (Mondejar et al., 2021). eNanoMapper serves as a data infrastructure for managing the toxicity of engineered nanomaterials (Jeliazkova et al., 2015). PubVINUS provides physicochemical properties and bioactivities as well as descriptors of various nanomaterials (Yan et al., 2020). However, most of the existing databases are focused on the bioactivities and the environmental impact of nanomaterials; thus, only the solubility in water is included. To establish the correlations among the structure, functionality, and properties of nanomaterials related to process manufacturing (Figure 1), we have proposed a new database as its foundation for future development. Our ultimate goal includes establishment of correlation models among feature measurements/parameters (e.g. molar weight, size distribution), transformed features (e.g. solubility parameter, surface potential), and process properties of nanomaterials (e.g. solubility, cohesiveness), as shown in Figure 2.

2.1. Data collection

Currently, this database is focused on the solubility of nanomaterials. Research papers and reviews were screened for collecting data points. The keywords used for searching include, but are not limited to: colloids, dispersibility, HSPs, nanomaterial, nanoparticle, solubility, stability, and organic solvent. Nanomaterials such as 2D-nanomaterials and nanotubes were also collected, as particles with dimensions smaller than 500 nm, as well as tubular or fiber-like structures with two dimensions below 100 nm are also qualified as “nanoparticles” (Vert et al., 2012).

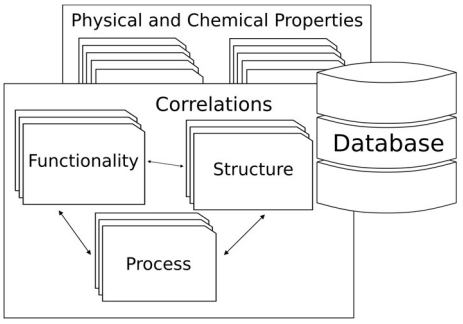


Figure 1 Layout of the database

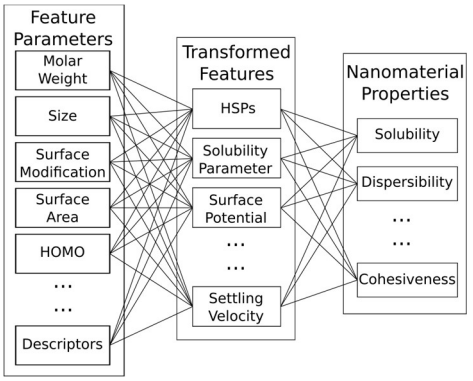


Figure 2 Illustration of correlation models

2.2. Features of the database

The current database contains approximately 1,200 solubility data points from 74 types of organically-modified and non-modified nanomaterials, including six types of nanotubes, 44 types of 2D-nanomaterials, and 38 types of nanoparticles. The core compositions of the collected nanomaterials and their percentages in this database are shown in Table 1. Furthermore, parameters related to solubility, such as the estimated HSPs of nanomaterials, sonification conditions (time, initial concentration, sonification power, temperature), and centrifugation conditions (revolutions per minute, time), were collected as well, depending on the applicability.

Approximately 250 pure, and 30 mixed solvents were collected. Among them, there were 23 frequently-utilized solvents, ranging from low polarity (such as hexane) to high polarity (such as dimethyl sulfoxide). In addition, approximately 53 % of all data points were originated from organically-modified nanomaterials. Decanoic acid, oleic acid, and acetic acid are currently the most commonly used organic modifiers in this database.

Solvents and organic modifiers were reorganized into their own auxiliary databases to enrich the information contained in the database, for modeling purposes. For each substance in its corresponding auxiliary database, more than 230 additional feature parameters were gathered from various sources: for example, surface tension and dielectric constants were collected from handbooks (Haynes et al., 2017), quantum chemical properties such as highest occupied molecular orbital were calculated using the Python package Psi4 (Psi4 project team, 2021), and molecular descriptors were acquired using the Python package RDKit (G. Landrum, 2021).

2.3. Data quality and discussions

The greatest challenge we experienced during data collection was the different formats of solubility in different reports. Only 28.9 % of the solubility data were reported as concentration measurements with the units of mg/mL or mg/mg, and 18.1 % of the data were expressed in terms of light absorption. Nearly 49.5 % of the data points were reported using descriptions such as “good” or “bad,” and the rest can only be evaluated indirectly from photos. In addition, data that related to the preparation of nanomaterial solutions, such as sonification time, were often either unavailable or incomplete.

The quality of the data will certainly affect the quality of the prediction models. It also limits the direct comparisons among different nanomaterials. Therefore, it is crucial to form a standard format for recording experimental data in the future.

Currently, this database is at its early stages. More solubility data, features and properties of nanomaterials will be collected and reorganized in the future.

Table 1 Core composition of nanomaterials and their percentages in the database

ZrO_2	18.0 %	Al_2O_3	4.4 %	WS_2	3.7 %	$MoSe_2$	2.8 %
RGO[1]	11.7 %	ZnO	4.4 %	C_3N_4	3.2 %	MoS_2	2.5 %
C_{60}	7.8 %	Graphene	4.2 %	Germanane	3.2 %	$MoTe_2$	1.9 %
CNT[2]	6.0 %	GO[3]	4.0 %	CeO_2	2.9 %	Others	19.3 %

[1] RGO: reduced graphene oxide; [2] CNT: carbon nanotube; [3] GO: graphene oxide.

3. Selection of solvent features and data-driven solubility models

In previous studies, HSPs have often been used to predict the solubility of nanomaterials (Tomai et al., 2021). This prediction is based on the similarity between the solvent and solute, by calculating $R_s = (4\Delta\delta_D^2 + \Delta\delta_P^2 + \Delta\delta_H^2)^{0.5}$, where $\Delta\delta = \delta_{solvent} - \delta_{solute}$, and δ_D, δ_P, and δ_H are different HSPs that contribute to solubility (Hansen, 2007). When R_s is smaller than an experimental-determined threshold, it indicates that the solvent and solute are adequately similar, so that a stable solution can be formed. This method functions well for bulk materials. However, to reiterate the demonstration by various data points in our database, R_s does not predict well for nanomaterials, regardless of their surfaces being modified or not.

3.1. Feature selection

In this study, we attempted to obtain important solvent feature parameters and their correlations with the solubility of nanomaterials using data-driven statistical models. However, the quality of the data prevented us from creating a universal solubility model. Thus, the CeO_2 nanoparticle modified with decanoic acid (Tomai et al., 2021) was chosen as the modeling nanoparticle because of its typical diameter (5.2 nm) and tight size distribution.

To construct reliable data-driven statistical models, important solvent feature parameters governing the solubility of nanomaterials must be identified first before modeling. Instead of using feature extraction methods (for example, principle component analysis), feature selection methods are used to eliminate irrelevant solvent feature parameters. Feature selection methods are commonly categorized into three types: the filters rely on the correlation of candidate variables with the variable to predict regardless of models, the wrappers select best performed subsets of variables depending on chosen models, and the embedded methods select variable subsets and tune the model parameters simultaneously. In this study, recursive feature elimination (RFE) was used as the wrapper algorithm for several linear and nonlinear data-driven models listed in Table 2. Least absolute shrinkage and selection operator for logistic regression (LASSO-LR) was chosen as the embedded method. Both feature selection and modeling were conducted under the R-4.1.0 environment using “caret” and “glmnet” packages.

Table 2 Selected solvent features using RFE

Logistic Regression	BCUT2D_MWHI, dielectric constant, dipole moment, HOMO, δ_P, FpDensityMorgan1, LUMO, MaxEStateIndex, MinAbsEStateIndex, MinEStateIndex, molar volume, molar weight, QED, VSA_EState1
Random Forest, SVM[1] (linear), SVM (polynomial)	BCUT2D_MWHI, dielectric constant, dipole moment, EState_VSA1, FpDensityMorgan1, δ_P, MinAbsPartialCharge, molar logP, molar refractivity
SVM (radial), LDA[2], kNN[3]	dielectric constant, δ_P, δ_H, MaxPartialCharge, MinAbsPartialCharge, MinPartialCharge, molar logP, PEOE_VSA1, SlogP_VSA2

[1] SVM: support vector machine; [2] LDA: linear discriminant analysis; [3] kNN: k-nearest neighbors.

Twenty-five solubility data points of the modified CeO_2 nanoparticles were divided into an 18-sample training dataset and a seven-sample test dataset. All test samples have missing values for solvent surface tension. Because no detailed measurement of solubility was reported, the nanoparticle solubilities in all solvents were categorized into two classes: samples with solubility larger than 0.01 wt% was labeled as "soluble," while the rest were labeled as "non-soluble." The results of the RFE from 230 solvent features using the training dataset are listed in Table 2. It can be observed that only two HSPs, δ_P and δ_H, were able to be selected as significant features. In addition, it is clear that other features might also play important roles in solubility prediction, such as the dielectric constant, dipole moment, and molar logP.

3.2. Data-driven modeling for solubility of nanomaterial

The selected features using RFE and LASSO were reorganized into combinations of feature data as the inputs for each corresponding modeling method. Any combination containing correlated features was removed. Owing to the small sample size of the training dataset, the leave-one-out cross-validation strategy was used for model training. The prediction performance of the models was tested using the test dataset. A few of the best-performed feature combinations for each model are listed in Table 3.

It can be observed that the solubility prediction performances using the data-driven models, especially the radial-kernel SVM, were much better as compared to using the similarity-based R_s. In addition, solvent features, such as dielectric constant, dipole moment, and molar logP, performed as good as, if not better than HSPs δ_P and δ_H. This coincides with the results obtained from the previous feature selections. However, it needs to be emphasized that these predictions are limited to only one nanoparticle, and further evaluation of the models is constrained by the small size of available data. Therefore, more research will be conducted on how to construct better solubility prediction models using various yet limited information from the entire database.

Table 3 Best-performed feature combinations for each data-driven model

Data-driven Model	Modeling Variables	Accuracy
Logistic Regression	dielectric constant, EState_VSA1, LUMO, MinEStateIndex	85.7 %
Random Forest	dielectric constant, EState_VSA1, MinAbsPartialCharge	85.7 %
SVM (linear)	dipole moment	71.5 %
SVM (polynomial)	dipole moment, EState_VSA1, molar logP	85.7 %
SVM (radial)	dielectric constant, EState_VSA1, MaxPartialCharge, SlogP_VSA2	100 %
LDA	dipole moment, FpDensityMorgan2, δ_H	85.7 %
kNN	dipole moment, molar logP, VSA_EState1	85.7 %
LASSO-LR	dipole moment, δ_P, NumAliphaticCarbocycles, QED, SlogP_VSA4	71.5 %
R_s	δ_D, δ_P, δ_H	42.9 %

4. Conclusions

In this study, we have presented a database for the solubility of organically-modified and non-modified nanomaterials in organic solvents. This database contains approximately 1,200 data points. Furthermore, we attempted to construct data-driven statistical models for solubility using the collected data of one nanoparticle. The data-driven solubility prediction models perform better than the R_s method based on the similarity of solute-solvent HSPs. These models demonstrated that in addition to HSPs δ_P and δ_H, solvent feature parameters, such as dielectric constant and molar logP, should also be viewed as important factors in predicting the solubility of nanomaterials.

Acknowledgement

This study is supported by the grants from Materials Processing Science project ("Materealize") of Ministry of Education, Culture, Sports, Science and Technology, Japan, Grant Number JPMXP0219192801.

References

K.T. Butler, D.W. Davies, H. Cartwright, O. Isayev, and A. Walsh, 2018, Machine Learning for Molecular and Materials Science, Nature, 559, 547–555

C.M. Hansen, 2007, Hansen Solubility Parameters: A User's Handbook, 2nd Edition, CRC Press, LLC, Boca Raton, USA

W.M. Haynes (Editor-in-Chief), 2017, CRC Handbook of Chemistry and Physics, 97th edition, CRC Press, LLC, Boca Raton, USA

N. Jeliazkova, C. Chomenidis, P. Doganis, B. Fadeel, R. Grafström, B. Hardy, J. Hastings, M. Hegi, V. Jeliazkov, N. Kochev, P. Kohonen, C.R. Munteanu, H. Sarimveis, B. Smeets, P. Sopasakis, G. Tsiliki, D. Vorgrimmler and E. Willighagen, 2015, The eNanoMapper database for nanomaterial safety information, Beilstein Journal of Nanotechnology, 6, 1609–1634

G. Landrum, 2021, RDKit: Open-source cheminformatics, http://www.rdkit.org

M.E. Mondejar, M. Regidor, J. Krafczyk, C. Ihmels, B. Schmid, G.M. Kontogeorgis, F. Haglind, 2021, An open-access database of the thermophysical properties of nanofluids, Journal of Molecular Liquids, 333, 115140

Psi4 project team, 2021, http://psicode.org

T. Tomai, N. Tajima, M. Kimura, A. Yoko, G. Seong, T. Adschiri, 2021, Solvent accommodation effect on dispersibility of metal oxide nanoparticle with chemisorbed organic shell, Journal of Colloid and Interface Science, 587, 574–580

X. Yan, A. Sedykh, W. Wang, B. Yan, H. Zhu, 2020, Construction of a web-based nanomaterial database by big data curation and modeling friendly nanostructure annotations, Nature Communications., 11, 2519

M. Vert, Y. Doi, K-H Hellwich, M. Hess, P. Hodge, P. Kubisa, M. Rinaudo, and F. Schué, 2012, Terminology for biorelated polymers and applications (IUPAC Recommendations 2012), Pure and Applied Chemistry, 84, 377–410

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50032-4

Fast, efficient and reliable problem solution through a new class of systematic and integrated computer-aided tools

Orakotch Padungwatanaroj[a*], Nichakorn Kuprasertwong[a], Jakkraphat Kogncharoenkitkul[a], Kornkanok Udomwong[a], Anjan Tula[b], Rafiqul Gani[a,c]

[a]*PSE for SPEED Company Limited, 294/65 RK Office Park, Romklao Rd., Bangkok, 10520, Thailand*
[b]*College of Control Science and Eng, Zhejiang University, Hangzhou 310027, China*
[c]*PSE for SPEED Company, Skyttemosen 6, DK-3450 Allerød, Denmark*
**orakotch@pseforspeed.com*

Abstract

Process systems engineering involves the use of systematic methods and tools to solve a wide range of problems. In this paper, the integration of a new class of computer-aided methods and their associated tools in a flexible software architecture is presented. Two problem specific software architectures are highlighted, one for sustainable process design and another for integrated product design and analysis. In each case, the core software components, which are extended versions of ProCAFD (software for sustainable process design) and ProCAPD (software for product design and analysis) are highlighted together with linked software-components, and, solutions from different case studies.

Keywords: Sustainable process design, Chemical product design, integrated software architecture, computer-aided methods

1. Introduction

To tackle the grand challenges of energy, water, environment, food and health that modern society is currently facing, opportunities exist for development and use of new methods and associated computer-aided tools that are not available in the currently available software tools, such as the well-known process simulators (Pistikopoulos et al., 2021). Synthesis and design of new, innovative and significantly more sustainable processes require implementation of hybrid methods that employ multiple models, data sources and solution algorithms. For example, validation and testing of novel hybrid energy efficient separation techniques require the integration of data, models, simulation and analysis at various levels of complexity and integration of operational tasks (Tula et al., 2020). For generation and verification of new intensified process options, models that can simulate new intensified operations are needed as well as methods for design of chemicals and materials, as in membrane-based hybrid separation schemes (O'Connell et al., 2019). Similarly, synthesis, design, analysis of chemical products needs large databases to identify and/or substitute potentially harmful or hazardous chemicals within the product being designed (Syeda et al., 2022). The verification of new products needs validated models of processes, which are normally not available in commercial software tools, while, implementation of their design methods requires new and innovative unit operations and/or materials, as in membrane-based separation, that may also not be known. In many of the product and/or process synthesis and/or design problems, thermodynamic properties play a more significant role than their conventional use in mass and energy balance-based process models. For example, synthesis and design of energy efficient and environmentally acceptable solvent-based separations or fluids for

refrigeration, need integrated solution strategies accounting for product and process design issues. Solution of these problems contribute to tacking the challenges facing modern society.

Based on the above, any computer-aided software tool needs a number of options, such as, an option for synthesis of processing routes as well as chemical products; an option for design and analysis of chemical processes and products so that targets for sustainable improvements can be identified; and, an option for innovation so that novel and more sustainable alternatives can be generated and verified. To provide the above options, in addition to the model-based process simulation option, a number of computer-aided methods and their associated software tools (to be called linked components) are needed: a versatile database of chemicals with respect to their properties, their use, their environmental impacts etc.; a versatile modelling toolbox, including options for generating new models through data analysis, machine learning as well as first principles, when the needed model is not available in the model library; a versatile toolbox of property estimation options; a toolbox for process synthesis and another for product synthesis based on different methods of solution; a toolbox for design and analysis including economics, sustainability, environmental impacts, etc., and many more.

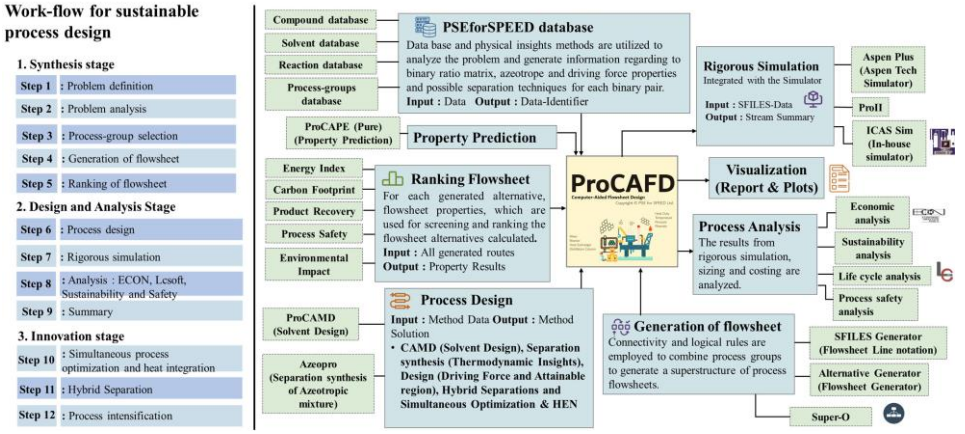
The objective of this paper is to highlight two integrated software tools that allow the needs to be matched through flexible and integrated software architectures. Two specific software tools, ProCAFD (Tula et al. 2019) and ProCAPD (Kalakul et al., 2018), for sustainable process design and chemical product design, respectively, are presented in their extended versions. The extensions are mainly with respect to how different linked components are integrated to the original, thereby enabling them to solve a wider range of problems. The additional software components are, PSE for SPEED database with data on more than 75000 chemicals; HI-Opt toolbox for simultaneous heat integration and process optimization, ProREFD (Udomwong et al., 2020) for refrigerant design and verification; ChemSub for substitution of hazardous and less efficient chemicals, to name a few. Note that external tools like process simulators or numerical solvers needed for specific steps in different work-flows are available. Results from case studies highlighting selected features of the integrated software tools are also presented.

2. Methodology & Associated Tools

2.1 ProCAFD-Sustainable process design: The work-flow for sustainable process design implemented in ProCAFD consists of 12 steps within the 3-stages methodology of Babi et al., (2015). Figure 1a highlights the 12 steps of the work-flow. Figure 1b highlights the tools needed for the different steps of the work-flow. Starting at the centre, where the ProCAFD tool is located, the next level (shaded in blue) shows the component functions and the next outer level (shaded in light green) shows the actual linked-components available for the steps of the work-flow. This component-based architecture provides the flexibility to configure and generate customized integrated software tools tuned for specific application areas and/or problem needs. That is, it removes the options not needed for the problems of interest and allows to add and/or remove linked-components as and when necessary. Figure 1b also highlights that different tool-components are available different stages, for example, Super-O for process synthesis stage, ECON, LCSoft for design-analysis stage and HI-Opt for process innovation stage.

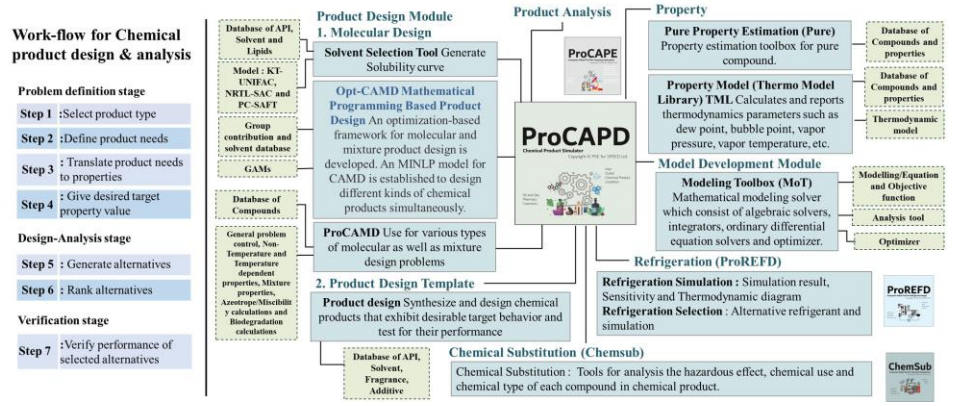
2.2 ProCAPD-Chemical product design: ProCAPD is a collection of linked components that are needed for various stages of the work-flow for design and/or analysis of different

types chemical products. Figure 2a shows the work-flow steps of the design methodology, while Figure 2b shows the architecture of ProCAPD.



(1a) (1b)

Figure 1: Work-flow (1a) and architecture of ProCAFD (1b)



(2a) (2b)

Figure 2: Work-flow (2a) and architecture of ProCAPD (2b)

Note that design of chemical products, which are classified as single molecules, liquid blends, formulated liquid blends, formulated functional products and devices, need for each product type and within it, each product sub-type, different sets of data, models and computational algorithms, even though the steps of the work-flow of the methodology are the same. For example, design of single molecules involving small molecules (as in refrigerant design) or larger organic chemicals (as in organic solvent design) or complex molecules with ions and organic fragments (as in ionic liquids), may require similar properties but different data-sets and/or property models. On the other hand, liquid blends are mixtures of liquid compounds while formulated liquid blends have active ingredients that are usually solids to which solvents and additives are added to obtain a final liquid blended product. A large database of 24000 chemicals classified in terms of lipids, solvents, ionic liquids, active ingredients, etc., is available for selection of a product component. A database of 810000 chemicals help to identify potentially hazardous

chemicals. The link to ProCAPE helps to analyse product and/or process functional properties. Since less than 10% of the chemicals in the different databases have measured property data, property prediction options in ProCAPE for pure component, mixture and phase equilibrium properties help to quickly fill-out the gaps in the required properties. Finally, different problem solution algorithms are available in ProCAPD for, selection of chemicals based on database search, computer-aided molecular design (ProCAMD), mathematical programming based molecular design (OptCAMD), and hybrid solution approaches. ProCAPD is also linked to ProREFD for refrigerant design and ChemSub for chemical product analysis and identification of hazardous chemicals in products.

3. Case Studies

3.1 Sustainable process design: The production of Cumene, an important intermediate substance for phenol and acetone production, is considered to highlight selected work-flow steps of the 3-stages methodology for sustainable process design. All steps and tools of ProCAFD shown in Figure 1 are used. Propylene reacts with benzene to produce cumene. Propylene also reacts with cumene to produce p-diisopropylbenzene as a by-product. Unconverted benzene is reacted with p-diisopropylbenzene in a second reactor to produce additional cumene. The synthesis toolbox generates a total of 8160 processing routes considering two reaction steps and all available process group options. Removing some of the membrane-based separation process groups, 2622 processing routes are obtained. Selecting only vapor-liquid distillation/flash process groups, 20 processing routes are obtained and one of these (Maity et al., 2013) is selected as the base case.

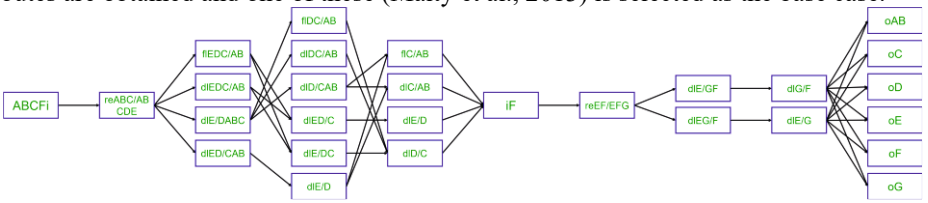
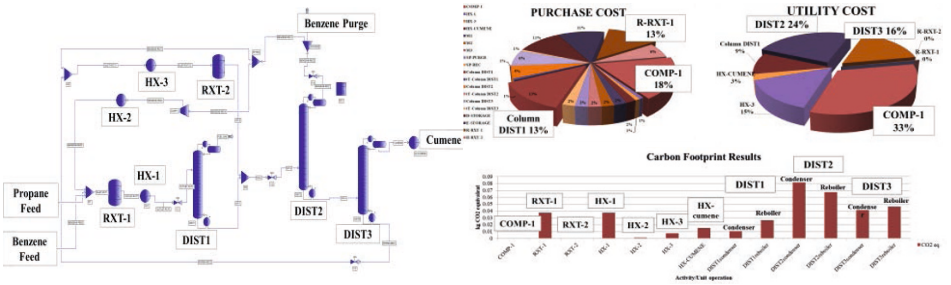


Figure 3: Reduced superstructure of processing routes for cumene production.

The starting point for the design-analysis stage is process simulation (results highlighted in Figure 4a). Driving force-based reverse design is performed (Tula et al. 2019) to obtain the design parameters of unit operations (such as distillation column, reactors, heat exchangers, etc.) that are needed for process simulation. AVEVA PROII (https://www.aveva.com/en/products/pro-ii-simulation/), which is linked to ProCAFD is used to perform steady state simulation and analysis tools such as ECON (cost estimation), LCSoft (life cycle assessment), Safety (inherent safety analysis) are used to identify the process "hotspots". As shown in Fig 4b, distillation columns 2 and 3 have the largest utility costs. This is also confirmed by the plots of the carbon footprints. The environmental impacts (not shown in Fig 4) indicate that the release of benzene, a carcinogenic chemical (also confirmed by ChemSub) leaves the process with the purge stream. Therefore, unlike the published design, the benzene feed is adjusted so that there is zero benzene purge, indicating that a small amount of by-product that has a lower price than the cumene will be produced as part of the more sustainable design. In stage 3, the targeted improvement in terms of decrease of energy demand is realized through simultaneous heat integration and process optimization (Duran and Grossmann, 1986), by adding an extended transhipment model to the optimizer in AVEVA PROII. Compared to the base case design, the profit is increased by 13.57%, the carbon footprint is reduced 78.65 % for the reboiler in DIST2. The ecological footprint is calculated to be 1.219. Detailed results for this sustainable process design can be obtained from the authors.



(4a) (4b)

Figure 4: Flowsheet for cumene production (4a) and selected analysis results (4b)

3.2 Chemical Product Design – Substitution Issues: Two chemical substitution problems related to chemical product design are highlighted. The first problem involves the replacement of refrigerants used in a two-cycle refrigeration process (Biegler et al., 1997) with alternatives that are more efficient and checked for safety issues through the hazards database. The refrigerants to be replaced are propylene (R290) and ethylene (R1150). ProREFD is used to find and verify the alternative refrigerants. One of the best alternatives for cycles 1 & 2 are listed in Table 1 together with refrigeration process performance results from ProREFD. In this example, ProCAPE, ProREFD and ChemSub have been used within ProCAPD. More solution details can be obtained from the authors.

Table 1 Performances of the best alternatives for cycles 1 & 2 & the reference

Cycle 1				**Cycle 2**				**Overall**
Comp	Qc kJ/hr	Qe kJ/hr	W kJ/hr	Comp	Qc kJ/hr	Qe kJ/hr	W kJ/hr	COP
R290	*67820*	*44867*	*22954*	*R1150*	*86098*	*54245*	*31853*	*0.99*
R1270/R1216 (38/62)	43279	23270	20009	Chemicals with similar boiling point could not be found.				

Note: The numbers in italic in the second row give the values of the reference refrigerants

The second example involves the improvement of a formulated liquid product (an insect repellent) by substituting its active ingredient (DEET, found in many commercial products), by another (Picaridin, found in some commercial products). The analysis shows that Picaridin has better functional properties against mosquitos, while Deet is more effective against other insects. Table 2 gives lists their known hazardous properties. Note that if the active ingredient is changed, the formula for the formulated liquid product also needs to be verified. The formulated liquid insect repellent with Deet has the following contents (in terms of weight percent): Deet = 10%; isopropanol = 41.8 %; water = 44.15%; and additives (acetic acid for pH control and linalool for fragrance) = 4.05%. For insect repellent based on Picaridin the formula is the following: Picaridin = 9.7%; 2-propanol = 42.3%; water = 43.95%; additives (acetic acid for pH control and linalool for fragrance) = 4.05%, which has been verified through experiments (Conte et al., 2012). In this example, Product template, ProCAPE, and ChemSub have been used within ProCAPD. More solution details can be obtained from the authors.

Table 2 Example of active ingredient used in mosquito repellent

Chemical	**State**	**Health hazard**	**Environmental hazard**	**Physical hazard**
Deet	Solid	Irritant-, skin (2), eye (2); Acute toxicity (4)	Aquatic chronic (3)	May damage plastic and leather
Picaridin	Solid	Acute toxicity (4)	Slight aquatic toxicity (4)	-

Note: The numbers in parenthesis indicate assigned category, where 5 indicates low hazard level and 1 indicates high hazard level.

4. Conclusions

A flexible software architecture that is capable to meeting the needs of sustainable process design and chemical product design and analysis has been presented. The architecture allows the linking of specific computer-aided methods and associated software tools as linked components for specific integrated software tools. Two integrated software tools one configured for sustainable process design and another for chemical product design and analysis have been presented. Through case studies, some of the features of the integrated software tools have been highlighted. The architecture of the integrated software tools allows linked components to be added or removed according to specific problem requirements. In this way, the changing specifications of methods and associated computer-aided tools needed to efficiently and reliably solve problems tackling the challenges of energy, water, environment, food and health can be matched. Current and future work is developing more case studies and testing the plug and play issues of linked components within problem specific software architectures to identify the desired more sustainable product and process alternatives. Flexible component-tools based architecture of software tools could be the answer for the need for computer-aided tools that are required to tackle the challenges of a changing earth.

References

D. K. Babi, J. Holtbruegge, P. Lutze, A. Gorak, J.M. Woodley, R. Gani, 2015, Sustainable Process Synthesis-Intensification, Computers & Chemical Engineering, 81, 218-44

E. Conte, R. Gani, Y. S. Cheng, K. M. Ng, 2012, Design of formulated products: experimental component, AIChE Journal 58 (1), 173-189

M. A. Duran, I. E. Grossmann, Simultaneous optimization and heat integration of chemical processes, 1986, AIChE J, 32, 123-138

S. Kalakul, L. Zhang, Z. Fang, H.A. Choudhury, S. Intikhab, N. Elbashir, M.R. Eden, R. Gani, 2018, Computer aided chemical product design – ProCAPD & tailor-made blended products, Computers & Chemical Engineering, 116, 37-55

D. Maity, R. Jagtap, N. Kaistha, 2013, Systematic top-down economic plantwide control of the cumene process, Journal of Process Control, 23, 1426-1440

J. P O'Connell, M. R. Eden, A. K. Tula, R. Gani, 2019, Retrofitting Distillation Columns with Membranes, *Chemical Engineering Progress,* 115 (12), 41-49

E. N. Pistikopoulos, A. Barbosa-Povoa, J. H. Lee, R. Misener, A. Mitsos, G. V. Reklaitis, V. Venkatasubramanian, F. You, R. Gani, 2021, Process Systems Engineering–The Generation Next? Computers & Chemical Engineering, 147, 107252

S. R. Syeda, E. A. Khan, O. Padungwatanaroj, N. Kuprasertwong, A. K. Tula, 2022, A perspective on hazardous chemical substitution in consumer products, Current Opinion in Chemical Engineering, 36, 100748

A. K. Tula, M.R. Eden, R.Gani, 2020, Computer-aided process intensification: Challenges, trends and opportunities, AIChE Journal, 66, e16819

A. K. Tula, M.R. Eden, R. Gani, 2019, Hybrid method and associated tools for synthesis of sustainable process flowsheets, Computers & Chemical Engineering, 131:2019, 106572

K. Udomwong, A. Robin, N. Kuprasertwong, O. Padungwatanaroj, A.K. Tula, L. Zhu, L. Zhou, B. Wang, S. Wang, R. Gani, 2020, ProREFD: Tool for Automated Computer-Aided Refrigerant Design, Analysis, and Verification, Computer Aided Chemical Engineering, 50, 457-462

A.W. Westerberg, I. E. Grossmann, L. Biegler, 1997, Chapters 4&5, Systematic Methods of Chemical Process Design, Wiley, USA

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50033-6

Design of Bio-Oil Solvents using Multi-Stage Computer-Aided Molecular Design Tools

Jia Wen Chong[a], Suchithra Thangalazhy-Gopakumar[a], Kasturi Muthoosamy[b], Nishanth G. Chemmangattuvalappil[a*]

[a]*Department of Chemical and Environmental Engineering, University of Nottingham Malaysia, Selangor 43500, Malaysia*
[b]*Nanotechnology Research Group, Centre of Nanotechnology and Advanced Materials, University of Nottingham Malaysia, Selangor 43500, Malaysia*
Nishanth.C@nottingham.edu.my

Abstract

Direct application of fast-pyrolysis bio-oil as biofuel is limited due to its undesirable attributes like low heating value, high viscosity, and storage instability. Solvent addition is a simple and practical method in upgrading pyrolysis bio-oil. In this work, a computer-aided molecular design (CAMD) tool was developed to generate the molecular structure of the solvent with desirable properties. Molecular signature descriptors were employed to represent property prediction models that comprise different classes of topological indices. Because of the differences in the structural details involved in different property prediction models, signatures of different heights were needed in formulating the design problem. However, the complexity of a CAMD problem increases with the height of signatures, due to the combinatorial nature of higher-order signatures. Thus, a multi-stage framework was developed by introducing a novel set of consistency rules that restrict the number of higher-order signatures. With the developed consistency rules, only relevant and consistent signatures were generated to keep the CAMD problem in a manageable size. Phase stability analysis was conducted after solvent candidates were identified to evaluate the stability and miscibility of the solvent-oil blend. As a result, a feasible solvent that fulfils the target properties with low environmental impact was identified.

Keywords: Computer-aided molecular design, bio-oil solvent, molecular signature descriptor.

1. Introduction

Biomass has received increased attention as a potential alternative fuel by converting into bio-oil via various conversion processes. Among the available biomass conversion processes, pyrolysis has the advantage of being a relatively simple and inexpensive technology. However, poor fuel properties of bio-oil from pyrolysis such as corrosiveness, high viscosity and low heating value limit its application as a biofuel. Solvent addition is one of the most popular bio-oil upgrading methods as it is relatively simple and economically viable. Conventionally, the design of solvents involves a trial-and-error process within a large set of candidates which is tedious, time-consuming, and costly. Unlike traditional search and optimisation techniques, a more efficient solvent design can be carried out by utilising CAMD (Computer-Aided Molecular Design) tools

where molecules possessing desired properties are identified based on the pre-determined product requirements. CAMD is a reverse engineering approach in which the optimal molecules can be identified from a given set of molecular building blocks and a specified set of targeted properties. In the past, CAMD has been widely incorporated in designing solvents for biofuel additives. Previous research focused only on the functionality of the solvent itself that can be predicted using GC prediction models with 1st order group contributions. However, it is also important to include the environmental aspects into the design of bio-oil solvent to minimise the environmental impact. Moreover, incorporation of contributions from higher-order molecular groups in CAMD is crucial to account for the interactive effects of molecular groups (Marrero & Gani, 2001). In addition, the selected GC model may not have all the model parameters required for the estimation of property of a specific chemical (Hukkerikar et al., 2012). For this reason, TI (Topological Index) approaches can be applied as they are a function of the entire molecular graph, which reflect the entire nature of the molecular structure (Austin et al., 2016).

Different types of property prediction models can be modelled using either GC or TI approach. Different properties may be expressed with different TI as well. However, different TIs exhibit different mathematical expression, which pose challenges in combining and solving it simultaneously on a common platform (Chemmangattuvalappil & Eden, 2013). To overcome this issue, molecular signature descriptor was introduced, where various GC models and TIs can be expressed on a common platform (Visco et al., 2002). Molecular signature descriptor is one of the 2D fragment-based TI that systematically captures the structural information of a 2D structural formula. It describes the molecular atoms in terms of extended valencies up to a predefined height (Faulon et al., 2003). Owing to the fact that molecular signature descriptor is known as the canonical representation of a molecule, all other 2D classes of descriptors can be represented in terms of molecular signature (Visco & Chen, 2016).

Signatures of higher height were required for the coverage of TIs and higher-order GCs. Despite the high accuracy of estimation, the complexity of CAMD increases due to the combinatorial nature of higher-order signatures. Hence, the height of signatures must be lowered to be used in a CAMD formulation. However, not all the signatures considered in the CAMD problem are consistent with each other to form a feasible molecule. Thus, a consistency rule was developed in this work to reduce the size of CAMD problem by excluding irrelevant molecular signature at a lower height from the building block sets (Chong et al., 2021). Infeasible signatures (signature that do not fulfil the consistency rules) are systematically eliminated at different levels and this can help to keep a manageable problem size. After determining all the possible additives, the accuracy of the estimated higher heating value of solvent candidates were verified through a database search. Other than the thermodynamic properties, Gibbs free energy of mixing was estimated to evaluate the miscibility of solvent-oil blend. With the developed approach, an optimum solvent that improves the solvent-oil blend properties and stability was generated.

2. Methodology

This work presents a novel multi-stage solvent design methodology with consistency rule incorporated to reduce the size and complexity of the CAMD problem. The developed framework can be divided into 4 main stages as shown in Figure 1.

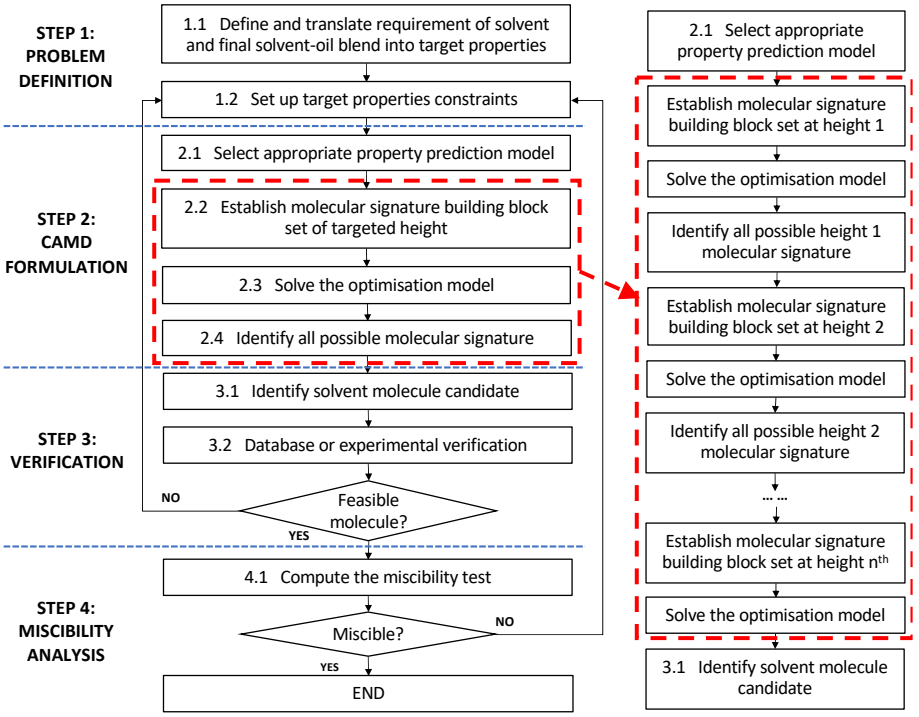


Figure 1 Framework for the development of CAMD model for the design of solvent

2.1. Problem Definition

Firstly, the problem definition was formulated, where the product needs were determined based on the requirements from regulations and specifications. In addition, environmental properties were considered to ensure that the generated solvent molecules have low environmental impact. The identified product requirements were then translated into measurable quantitative target properties. This is then followed by selection of suitable property prediction models to estimate the target properties of the solvent. In this work, property prediction models in terms of GC method and TI were considered and expressed as a function of the molecular signature descriptor. As shown in Eq. 1, the molecular signature descriptor of molecule *G*, *TI(G)* can be expressed as a dot product between two vectors, ${}^{h}\alpha_g$, the vector of occurrence number of atomic signatures of height *h*, and $TI\left(root\left({}^{h}\sum\right)\right)$, the vector of predicted values from the model computed for each of the atomic signatures.

$$TI(G) = k^{h}\alpha_g \cdot TI\left(root\left({}^{h}\sum\right)\right) \tag{1}$$

2.2. Consistency Rules

In the developed approach, signatures of height h were generated based on the collection of height $h-1$ signatures identified from the CAMD problem. The first layer of signature generated must contain one of the height $h-1$ signatures from the previous result. For example, assuming the signatures C1(C), C2(CC), C2(CO) and C3(CCO) were identified as the promising height 1 signature from the CAMD problem, the generated height 2 signatures based on C1(C) are shown as below:

1. C1(C2(CC))
2. C1(C2(CO))
3. C1(C3(CCO))

With this approach, the total number of generated height 2 signatures was reduced from 13 signatures to 3 signatures. In another example, taking the collection of height 2 signatures, the following set is obtained:

1. C1(C3(CCO)
2. C1(C2(CC))
3. C2(C1(C)C2(CC))
4. C2(C2(CC)C2(CC))
5. C2(C2(CC)C3(CCO))
6. C3(C1(C)C2(CC)O1(C))
7. O1(C3(CCO))

In this case, height 3 signatures generated based on the signature (3), C2(C1(C)C2(CC)) are listed as:

1. C2(C1(C2(CC))C2(C1(C)C2(CC)))
2. C2(C1(C2(CC))C2(C2(CC)C2(CC)))
3. C2(C1(C2(CC))C2(C2(CC)C3(CCO)))

2.3. Verification and Miscibility Analysis

To ensure that the molecules generated from previous steps are feasible and practical, verification step was conducted through database search from various platforms. On the other hand, phase stability test was conducted by computing the Gibbs tangent plane distance to avoid phase separation in the final solvent-oil blend.

3. Case Study

3.1. Problem Definition

The main objective of the designed solvent is to improve the physical properties of the bio-oil. Greater higher heating value (HHV) is preferable for better fuel combustion. Thus, the HHV of the designed solvent was maximized, which serves as the objective function. Table 1 summarized the constraints for each target properties identified. In this study, bio-oil derived via palm kernel shells (PKS), with moisture content of 16 wt.% and HHV of 19 MJ/kg, was used as the basis. The main components of the pyrolysis bio-oil include phenol, 2,6-dimethoxyphenol, 2-methoxyphenol, furfural, acetic acid and 1,2-benzenediol.

Based on the target properties identified, suitable property prediction models in terms of GC method and connectivity index were selected to estimate the properties of the designed solvents. In this case study, maximum signature height required in this problem was set at 4. The atoms that are commonly present in solvents (hydrogen, carbon,

nitrogen, and oxygen) were chosen for the design of bio-oil solvent. The hydrocarbon groups considered in this study were limited to alkanes, alkenes, alcohol, carboxylic acid, ketones, aldehyde, esters, ethers, and nitriles which can be predominately found in solvents. Then, the molecular signatures descriptors up to height 4 were generated based on the selected atoms' type and chemical families, resulting in a total of 10,000 different molecular signature combinations. By applying the consistency rule, the signature set size was reduced to the final 21 height 4 signatures.

Table 1 Target properties and constraint for each identified product requirements.

Requirements/Needs	**Targeted Properties**	**Constraints**
Liquid state at room temperature	Normal boiling point / K	> 400.15
	Normal melting point / K	< 298.15
Combustion quality	Higher heating value	To maximize
Fuel flow consistency	Viscosity / mPa s	$1 > v > 6$
	Density / kg m^{-3}	$800 > \rho > 1000$
Environmental related properties and toxicology	Aquatic acute toxicity, LC50	> 100
	Aquatic acute toxicity, EC50	> 100
	Oral acute toxicity, LD50	> 100
	Bioconcentration factor	< 1000
	Soil-water partition coefficient / L kg^{-1}	< 31622
	Global Warming Potential	< 10
	Photochemical Oxidation Potential	< 10

3.2. Results and Discussion

Based on the database search conducted, the feasible solvent molecules were identified as 2-octanol, 2-heptanol, 2-hexanol and 2-pentanol, respectively. The higher heating value estimated in present work for the abovementioned solvent candidates were close to the actual higher heating value obtained from NIST database, with less than 1% differences. All the resulting molecules possess a higher heating value of at least 37.5 MJ/kg. It can be concluded that 2-octanol is the most suitable solvent candidate with the highest higher heating value at 40.89 MJ/kg. Figure 2 shows the Gibbs energy and tangent plot for 2-octanol and bio-oil blend. From Figure 2, the blend is stable and exhibit homogenous single-phase as the tangent line was plotted below the Gibbs curve.

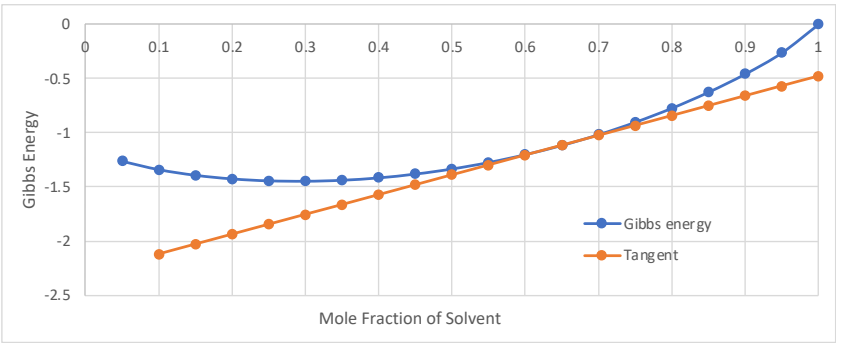


Figure 2 Gibbs energy and tangent plot for 2-octanol and bio-oil.

4. Conclusion

A multi-stage CAMD methodology was developed to design an optimal solvent that can upgrade bio-oil while possessing low environmental impact. The developed multi-stage approach was used to reduce the size of problem due to the combinatorial nature of higher-order signatures. Moreover, consistency rules were applied to ensure only relevant and consistent signatures are generated. The results from the case study shows that solvent generated can achieve good functionality while displaying promising environmental characteristics. To conclude, the developed methodology in this work can be applied in the design of molecules for any application. Further improvements can be made by considering the addition of emulsifiers and/or reactive solvents in the design of additives for bio-oil upgrading purposes.

Acknowledgements

The authors would like to express sincere gratitude to Ministry of Higher Education Malaysia for the realization of this research project under the Grant FRGS/1/2019/TK02/UNIM/02/1.

References

A.S. Hukkerikar, S. Kalakul, B. Sarup, D.M. Young, G. Sin & R. Gani, 2012, Estimation of environment-related properties of chemicals for design of sustainable processes: Development of group-contribution+ (GC +) property models and uncertainty analysis. Journal of Chemical Information and Modeling, 52(11), 2823–2839.

D.P. Visco & J.J. Chen, 2016, The Signature Molecular Descriptor in Molecular Design: Past and Current Applications. Computer Aided Chemical Engineering, 39, 315–343.

D.P. Visco, R.S. Pophale, M.D. Rintoul & J.L. Faulon, 2002, Developing a methodology for an inverse quantitative structure-activity relationship using the signature molecular descriptor. Journal of Molecular Graphics and Modelling, 20(6), 429–438.

J. Marrero & R. Gani, 2001, Group-contribution based estimation of pure component properties. Fluid Phase Equilibria, 183–208.

J.L. Faulon, D.P. Visco & R.S. Pophale, 2003, The Signature Molecular Descriptor. 1. Using Extended Valence Sequences in QSAR and QSPR Studies. Journal of Chemical Information and Computer Sciences, 43(3), 707–720.

J.W. Chong, S. Thangalazhy-Gopakumar, K. Muthoosamy & N.G. Chemmangattuvalappil, 2021, Design of bio-oil additives via molecular signature descriptors using a multi-stage computer-aided molecular design framework. Frontiers of Chemical Science and Engineering, https://doi.org/10.1007/s11705-021-2056-8.

N.D. Austin, N.V. Sahinidis & D.W. Trahan, 2016, Computer-aided molecular design: An introduction and review of tools, applications, and solution techniques. Chemical Engineering Research and Design, 116, 2–26.

N.G. Chemmangattuvalappil, & M.R. Eden, 2013, A Novel Methodology for Property-Based Molecular Design Using Multiple Topological Indices. Industrial & Engineering Chemistry Research, 52(22), 7090–7103.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50034-8

Synthesis of azeotropic distillation processes without using a decanter

J. Rafael Alcántara-Avila*, Maho Okunishi, Shinji Hasebe

Department of Chemical Engineering, Kyoto University, Katsura Campus Nishikyo-ku, Kyoto 615-8510, Japan
jrafael@cheme.kyoto-u.ac.jp

Abstract

This work presents a synthesis method for deriving distillation process structures to separate a ternary mixture with a homogeneous azeotrope. First, the liquid composition space is divided into small subspaces, each of which is assigned to a distillation module. Then, all of the connections among the heating, cooling, feed, product, and distillation modules are contained in a superstructure. The relations expressed in the superstructure are mathematically reformulated as a linear programming problem in which the utility cost is minimized. The optimal solution shows the connection between modules that indicates the best structure. The optimal solution is interpreted and translated to a feasible and realistic distillation process at a post-optimization step.

The proposed method is applied to the separation problem of the mixture of acetone, chloroform, and benzene. The optimization result shows that the direct sequence of two columns is a possible candidate for the optimal structure. Moreover, this structure is derived without assigning any *a priori* knowledge of the process structure.

Keywords: Process Synthesis, Azeotropic Distillation, Process Optimization

1. Introduction

Distillation is the most widely used technique to separate liquid mixtures. However, it uses large amounts of energy because the vaporization of liquid streams is inevitable. Although distillation research is often regarded as a mature area, new design and/or operation discoveries renew the interest in distillation because such improvements can translate into huge economic benefits (Caballero and Grossmann, 2013).

Li, Demirel, and Hasan (2019) proposed a block-based phenomena methodology for the synthesis of distillation processes. The blocks with phase and material assignment assemble to represent various phenomena (e.g., mixing, reaction, phase contact, and phase transitions), which are typical in distillation processes. These blocks can connect to represent various distillation processes, including single columns, dividing wall columns, and reactive distillation columns. Although an MINLP formulation was presented, the proposed methodology can generate intensified distillation processes without any *a priori* enumeration of candidate distillation processes. The separation of a ternary zeotropic mixture was taken up as a case study, and the optimization result showed that the dividing wall column (DWC) was the best process.

Takase and Hasebe (2015) proposed a synthesis method for the separation of a ternary mixture. In their method, the composition space was discretized, and each discretized subspace was assigned to a distillation module. For each module, the liquid and vapor

compositions and their molar enthalpies are uniquely determined. Thus, the synthesis problem was formulated as a linear programming (LP) problem. The vapor and liquid flow rates among modules were treated as optimization variables. The case study problem showed that the DWC was the best process to minimize operating costs.

It is known that even though a distillation boundary exists, in some cases, it is possible to separate a ternary mixture by using simple distillation columns (Biegler, Grossmann and Westerberg, 1997). The method proposed by Takase and Hasebe (2015) was applied to separating a ternary zeotropic mixture. However, this research extends their method because it considers separating a ternary mixture with an azeotrope and a distillation boundary. Thus, the separation is more difficult because a distillation boundary divides the composition space into two distillation regions.

2. Problem Statement

In this work, the synthesis problem for the separation of a ternary mixture containing a homogeneous azeotrope, based on the following assumptions, is formulated:

1. The ternary mixture forms a homogeneous azeotrope.
2. The feed is saturated liquid, and its flow rate and composition are given in advance.
3. All products are withdrawn as saturated liquids, and the purity specification of each product is given in advance as the lower bound of its key component.
4. The operating pressure is given in advance, and the pressure drop is negligible.
5. The process is at a steady-state.
6. The vapor leaving from a distillation stage is in equilibrium with the liquid on that stage.
7. Utility costs per unit amount of heating and cooling are given in advance.

Under these assumptions, the optimal process is derived without any structural assumption. The state-space discretization used in the IDEAS approach is adopted to formulate the material and heat balance equations (Drake and Manousiouthakis, 2002). A stage in any distillation column is treated as a distillation module. All possible flow connections among these modules are considered. Figure 1 represents the superstructure for the connections among modules.

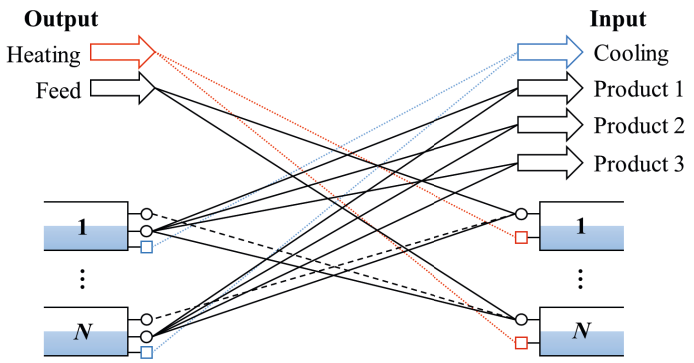


Figure 1. Superstructure representation for the separation of a ternary mixture

The squares numbered from 1 to *N* represent each distillation module. Each distillation module is further divided into input and output parts in which the same number is assigned.

Products 1 to 3 have only input flows, and the feed has only output flows. The heating and cooling utilities are also treated as modules. The solid, dashed, and dotted lines between modules represent feasible paths of liquid flows, vapor flows, and heat, respectively.

The proposed superstructure allows liquid and vapor flows among all distillation modules. The feed can be supplied to any distillation module. Each product module can accept liquid flows from any distillation module. Furthermore, heating and cooling are allowed in any distillation module.

3. Mathematical formulation

Based on the superstructure in Figure 1, the optimization problem which aims to minimize the utility cost (UC) is formulated by Eq. (1) to Eq. (6). Since the utility cost is the most relevant cost in distillation, the UC minimization is a valid assumption. Moreover, this research aims to derive novel distillation structures that do not depend on *any a priori* structural assumption rather than optimizing a predefined distillation structure.

Eq. (1) shows the objective function, Eq. (2) to Eq. (4) represent the material balance at each module, Eq. (5) represents the energy balance at each module, and Eq. (6) represents the purity constraint on each product

$$UC = \sum_{i \in S_M} (c_H Q_i^H + c_C Q_i^C) \tag{1}$$

$$\sum_{i \in S_M} L_{Fi} - L^F = 0 \tag{2}$$

$$\sum_{j \in S_M} \left(L_{ji} + V_{ji}\right) + L_{Fi} - \sum_{j \in S_M} \left(L_{ij} + V_{ij}\right) - \sum_{p \in S_P} L_{ip}^P = 0 \qquad i \in S_M \tag{3}$$

$$\sum_{j \in S_M} \left(L_{ji} x_{jk} + V_{ji} y_{jk}\right) + L_{Fi} x_k^F - \sum_{j \in S_M} \left(L_{ij} x_{ik} + V_{ij} y_{ik}\right) - \sum_{p \in S_P} L_{ip}^P x_{ik} = 0 \qquad i \in S_M, k = 1,2 \tag{4}$$

$$\sum_{j \in S_M} \left(L_{ji} h_j^L + V_{ji} h_j^V\right) + L_{Fi} h^F + Q_i^H - \sum_{j \in S_M} \left(L_{ij} h_i^L + V_{ij} h_i^V\right) - \sum_{p \in S_P} L_{ip}^P h_i^L - Q_i^C = 0 \qquad i \in S_M \tag{5}$$

$$x_k^P \sum_{j \in S_M} L_{jk}^P - \sum_{j \in S_M} L_{jk}^P x_{jk} \geq 0 \qquad k = 1,2,3 \tag{6}$$

where S_{M} and S_{P} are the sets of distillation module numbers and product numbers, respectively. c_{H} and c_{C} are the heating and cooling costs per unit amount of energy. Q_i^{H} and Q_i^{C} are the amounts of heat supply and removal at distillation module i. $L_{\mathrm{F}i}$ is the feed liquid flow rate entering to module i and L^{F} is the feed flow rate. L_{ji} and V_{ji} are the liquid

and vapor flow rates from module i to module j, respectively. L_{ip}^{P} is the liquid flow rate from module i to product p. x_{jk} and y_{jk} are molar fractions of the kth component in the liquid and vapor at module j, respectively. x_k^F is the molar fraction of the kth component in the feed stream. h_j^{L} and h_j^{V} are liquid and vapor molar enthalpy at module j. It is assumed that the purity specification of product k is given as the lower bound of the molar fraction of kth component and is given by x_k^{P}.

Eq. (1) to (6) are linear equations and inequalities for the optimization variables. Therefore, the optimization problem can be formulated as a linear programming (LP) problem. In plotting the result on a ternary distillation diagram, it is possible to recognize how the separation of the mixture with a distillation boundary is performed (Królikowski *et al.*, 2011). The result shown on the ternary distillation diagram is used to find the plausible process structure.

4. Case Study

The separation of acetone, chloroform, and benzene was taken as a case study (Biegler, Grossmann and Westerberg, 1997). The feed composition was 35 mol% of acetone, 25 mol% of chloroform, and 40 mol% of benzene, while the product specifications were 90 mol% for each product. This mixture forms a maximum temperature azeotrope between acetone and chloroform. The ternary diagram is separated into two distillation regions by a distillation boundary. The feed flow rate is 120 kmol/h, and the heating and cooling costs are 10 GJ/$ and 0.5 GJ/$, respectively. The NRTL method in Aspen HYSYS® V9 was used to estimate the physical properties.

5. Results and Discussions

5.1. Reference designs

Figure 2 shows the reference designs that meet the target 90 mol% compositions for all products. The designs were simulated in HYSYS® V9 for the sake of comparison. Design 1 has 50 stages in the first column (C1) and 60 stages in the second column (C2), while Design 2 has 45 stages in C1 and 50 stages in C2. The liquid composition profile of each design is shown in Figure 3. The utility cost for Design 1 was 130.60 $/h, while that for Design 2 was 152.20 $/h.

5.2. Optimization results

The composition space was discretized in increments of 0.025, and each discretized composition represents a subspace assigned to a distillation module. However, near the distillation boundary and above a benzene composition of 0.7, the composition space was discretized with increments of 0.005. As a result, 2,700 distillation modules were placed on the ternary composition diagram. Figure 4 shows the liquid flows between distillation modules obtained by solving the optimization problem. The squares in the figure represent the modules with liquid flows above 5 kmol/h, and the coordinates of each square represent the liquid composition assigned to the module. The solid grey lines represent the selected liquid path. The solid red squares near the benzene product mean that heating is necessary at each of those modules. In contrast, the solid blue squares near

the acetone and chloroform products mean that cooling is needed. The utility cost was 129.3 $/h, which is slightly lower than that of Design 1.

Since the composition space is discretized, the obtained process includes many secondary flows of small flow rate among the modules. However, extracting the dominant flows from the result shown in Figure 4 makes it possible to construct a simple and efficient process structure. In this case, the dominant flow paths in Figure 4 are similar to those in Figure 3. Thus, it can be said that the structure in Figure 2 is close to the optimal structure. As a result, the optimal structure's utility cost is close to that of Design 1. Though a new process structure was not created, the process structure was derived without inputting the structure information.

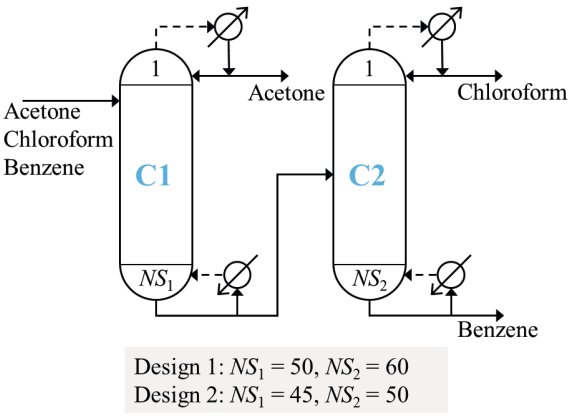


Figure 2. Reference designs

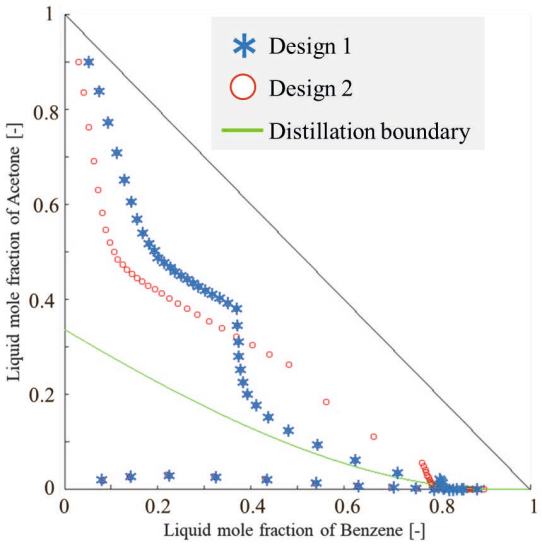


Figure 3. Liquid composition profile for each reference design

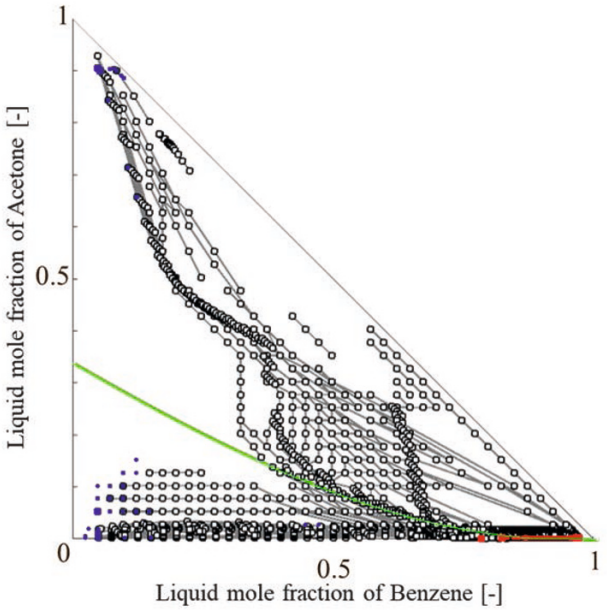


Figure 4. Liquid composition profile of the optimization result

6. Conclusions

A module-based synthesis method using the IDEAS approach was proposed for the separation of homogeneous azeotropic mixtures, which did not require a decanter. The method was applied to the separation problem of acetone, chloroform, and benzene, and the result showed that the optimal structure was close to the structure known as a direct sequence. The proposed method can be used without assuming the process structure in advance. Thus, it has a great potential to generate an original process structure that has not been used in the existing chemical plants.

References

Biegler, L. T., Grossmann, I. E. and Westerberg, A. W. (1997) *Systematic Methods of Chemical Process Design*. Prentice Hall.

Caballero, J. A. and Grossmann, I. E. (2013) 'Synthesis of Complex Thermally Coupled Distillation Systems Including Divided Wall Columns', *AIChE Journal*, 59(4), pp. 1139–1159.

Drake, J. E. and Manousiouthakis, V. (2002) 'IDEAS approach to process network synthesis: minimum utility cost for complex distillation networks', *Chemical Engineering Science*, 57, pp. 3095–3106.

Królikowski, A. R., Królikowski, L. J. and Wasylkiewicz, S. K. (2011) 'Distillation profiles in ternary heterogeneous mixtures with distillation boundaries', *Chemical Engineering Research and Design*, 89(7), pp. 879–893.

Li, J., Demirel, S. E. and Hasan, M. M. F. (2019) 'Systematic Process Intensification involving Zeotropic Distillation', *Computer Aided Chemical Engineering*, 47, pp. 421–426.

Takase, H. and Hasebe, S. (2015) 'Optimal Structure Synthesis of Ternary Distillation System', in *Computer Aided Chemical Engineering*. Elsevier B.V., pp. 1097–1102.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50035-X

Reduce Environmental Impact and Carbon Footprint for Cost Competitive Process Plant Design: Integrating AVEVA™ Process Simulation with modeFRONTIER®

Yutaka Yamada[a*],Simone Genovese[b], Cal Depew[c] , Ralph Cos[d], Hiroshi Kuwahara[e] and Taiga Inoue[f]

[a]*Product Development, AVEVA, Tokyo, 108-0023, JAPAN*
[b]*Project Management, ESTECO, Trieste 99 - 34149, ITALY*
[c]*Director, AVEVA, Lake Forest 92630, USA*
[d]*Principal SW Development, AVEVA, Garching 85748, Germany*
[e,f]*Simulation Technology, IDAJ, Kanagawa 220-8137, JAPAN*
yutaka.yamada@aveva.com

Abstract

There is a significant need in chemical and petrochemical industries for process synthesis tools to identify cost-competitive processes with low greenhouse gas emissions and environmental impact compared to existing designs, based on process-wide multi-objective optimization. Deadlines and human resources usually constrain the design process; however, only single-objective optimization for individual units is usually considered. To extend the optimization scope, AVEVA™ Process Simulation can be coupled with the multi-objective optimization software modeFRONTIER®. The integrated system allows solving multi-objective problems in a short time. This study demonstrates how a given industrial process can be optimized, ensuring stable and robust operation at low total annual cost (TAC) while reducing idle time and outperforming manual work. The proposed methodology is demonstrated using a case study of an ethanol distillation process plant. Designers and consultants can use this methodology to aid decision-makers in the design phase to identify robust, low-cost designs that deliver preferred performance in terms of environmental impact in the future operation.

Keywords: Multi-objective Optimization; Process Design; Process Simulation.

1. Introduction

Globally, governments are starting to impose regulations on greenhouse gas emission in addition to the existing regulations on water, and toxic pollution. Manufacturers are looking to make their operations greener and simulation can assist companies in their pursuit of greener production. Generative design, an iterative design process that uses simulation and machine-learning to mimic nature's evolutionary approach to design, offers the potential to boost sustainability across the manufacturing industry. Until now, generative design has been used primarily to optimize products, such as reducing the mass of products or costs while preserving functionalities. Currently generative design has shown capability and versatility to provide benefits in other applications, succeeding, for example, in construction industry's new challenges. This study will investigate if Generative Design has the potential ability to redesign larger manufacturing systems entirely. It must be emphasized that an ideal plant that is at the

same time efficient, cost-effective, environment-friendly, and risk-free is hard to achieve. There are always some necessary trade-offs to be made to ensure optimal use of energy resources while limiting environmental, health impacts or operability as shown by L. Gerber (L. Gerber et al., 2013) or B. Brent (B. Brent et al., 2020).

When the fluid includes the azeotropic components and recycles, it's hard to determine the optimum design for total number of trays, feed stages, recycle flow rate, reflux ratio, operating pressures and heating/cooling medium. Also the detailed engineering data associated with unit operations are necessary to estimate operating, equipment costs and environmental burden. In this project, AVEVA™ Process Simulation has been used as the first-principles simulation tool that has the flexibility to change various specification and the functionality to calculate the COD (Chemical Oxygen Demand) in the wastewater (environmental load), the Capital Cost(CapEx), Operating Costs (OpEx) and Greenhouse gas (GHG) emissions. To support this design process, to get a better understanding of the model itself, and to reduce the number of evaluations to find the optimal solution, it has been coupled with modeFRONTIER, the desktop solution for process automation and optimization in the engineering design process.

2. Simulation model and Design condition

2.1. Base model for the case study

The separation of ethanol and water is complex because of the existence of azeotrope; several distillation sequences are described in the book of Seader (Seader et al., 2010). The distillation feed mixture is composed of 90% water and 10% ethanol. In this study, following the most common and classic distillation method, three-column sequences using benzene as entrainer have been selected. The first column, Preconcentrator concentrates most of the ethanol from the feed. From the bottom of the second column, Azeotropic column, ethanol with more than 99% purity is produced. The overhead of the azeotropic column is condensed and decanted. The effluent with a high concentration of entrainment and ethanol is recycled to the Azeotropic column with the fresh benzene as reflux from the L1 phase. The water-rich phase is sent to the Entrainer Recovery column. The third column, the Entrainer Recovery column, recovers the entrainer and ethanol from the overhead and recycles them to the Azeotropic column. The distillation processes were simulated in AVEVA Process Simulation software using NRTL physical property model with the default binary parameters.

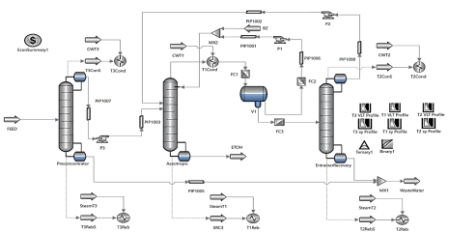


Figure 1: Flowsheet of the entire ethanol distillation process

2.2. Design conditions summary

The column diameter is set to keep the flooding factor = 0.7 and the height for each stage is fixed as 0.59m. The flooding factor is calculated using the Fair correlations. The Sieve tray is selected as the internals of the columns. The column efficiency is set as 0.6. Cooling water return temperature is set as 45C and U value is set as 1,000kcal/h-m2-K for all condenser. Steam outlet vapor fraction is set as 0 and U value is set as 5,000 kcal/h-m2-K for all reboiler. Counterflow heat exchange was applied for all heat exchanger. Pump efficiency is set as 0.7 and the elevation of each reflux drum is set as 10m for all pumps. The pipe length for the suction and discharge of the pump is set as

20m and the velocity is set as 2m/s for the liquid line. The preconcentrator column design condition is separately done from the entire optimization by creating NQ-curve using Python API because there are no recycles to this column. The Azeotropic column and the EntrainerRecovery column conditions are designed using modeFRONTIER as mentioned in section 4.3.

3. Objective function

3.1. Environmental impact

The environmental load is evaluated by the concentration of ethanol in the wastewater. In particular, the measure used to evaluate such concentration is the COD. Many governments impose strict regulations regarding the maximum chemical oxygen demand allowed in wastewater before it can be returned to the environment. For example, in Japan , a maximum oxygen demand of 160mg/l must be reached before wastewater or industrial water can be returned to the environment due to the Water Pollution Prevention Act. COD of the ethanol can be considered as 2.09 g COD/g ethanol from the following reaction: $C_2H_6O + 3O_2 \rightarrow 2CO_2 + 3H_2O$. The equation to calculate COD is implemented as a flowsheet equation.

3.2. Total annual cost (TAC)

AVEVA™ Process Simulation has the functionality to add the cost estimate model to each unit. The operational cost is based on the paper by Ulrich (Ulrich et al., 2006) as shown below. Each cost depends on two utility-specify coefficients (a, b), the current CEPCI, and $C_{s,f}$ (the cost of fuel in $/GJ).

$$C_{S,u} = a(CEPCI) + b(C_{S,f}) \tag{1}$$

The price is based on the 2019 CEPCI (607.5) and the average wholesale price of No. 2 fuel oil in 2019 for the United States ($12.83/GJ). Benzene cost is taken from 2015 ICIS. Operation time is considered as 8000hr/year. Installed cost is calculated based on the below equation from the book of James (James et al., 1988). As reported on Ethanol Producer Magazine (http://www.ethanolproducer.com), ethanol plants require stainless steel (SS) for most of the equipment hence it has been used for the cost calculations.

$$(Installed\ cost, \$) = C\left(\frac{M\&S}{280}\right) H^{n1} D^{n2} A^{n3} F_T \tag{2}$$

M&S index is calculated based on the below equation from the book of Zacharias (Zacharias et al., 2007).

$$M\&S = 1250 + 25(Year - 2005) \tag{3}$$

With SI units, the following equation is used for the cost of the column, internals, and heat exchanger.

$$(Column\ cost, \$) = (957.904)\left(\frac{M\&S}{280}\right) H^{0.802} D^{1.066} (2.18 + F_m F_p) \tag{4}$$

$$(Internal\ cost, \$) = (97.243)\left(\frac{M\&S}{280}\right) H^{1} D^{1.55} (F_s + F_t + F_m) \tag{5}$$

$$(HeatExchanger\ cost, \$) = (474.67)\left(\frac{M\&S}{280}\right)A^{0.65}(2.29 + (F_d + F_p)F_m) \quad (6)$$

The cost for the piping, electrical, buildings, and indirect costs are considered as double of the installed costs based on the rule of thumb. The wastewater treatment is not included in this cost model because the COD in the wastewater will be controlled to be under the regulation. The above equations are all included in the equipment cost submodel. The Total Annualized Cost (TAC) is calculated as a function of an annualized form of total capital cost in three years and the total operating cost. Cost escalation, interest, location factor is not considered here.

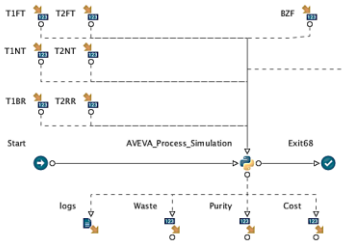
4. Optimization-driven design

4.1. Design exploration

Design Space Exploration is the process of finding a design solution (unique combinations of the settings of the independent variables) or solutions that best meet the desired design requirements from a space of tentative design points. One of the difficulties of the distillation process, especially azeotropic distillation process, is due to its high nonlinearity and number of variables to be considered. It's difficult to find the optimal number of stages, feed stage, reflux/boilup ratio and recycle flow rate as there are no analytical solutions. Also trying to analyse every possible design requires millions of case studies. This is a time-consuming method and due to the shortage of engineer's time, the case studies are usually stopped far before it gets the best solution. Process automation and machine learning algorithms are required to obtain optimal solutions with minimum computation and time. In this context, the engineer's know-how is fundamental while modeFRONTIER advanced data analysis and visualization tools allow to turn data into valuable insights and pick the best design solution.

4.2. Process Automation

To leverage modeFRONTIER's Machine Learning algorithms to optimize the process parameters in order to reduce the waste and the cost, an interface between the process simulation software AVEVA™ Process Simulation and modeFRONTIER® has been developed. The integration uses Python and communicates directly to AVEVA™ Process Simulation thanks to native Python APIs.



Variables	Description	Type
T1FT	Azeotropic column feed stage of recycle stream	Discrete
T1NT	Azeotropic column total number of tray	Discrete
T1BR	Azeotropic column boil up ratio	Continuous
T2FT	EntrainerRecovery column feed stage	Discrete
T2NT	EntrainerRecovery column total number of tray	Discrete
T2RR	EntrainerRecovery column reflux ratio	Continuous
BZF	Entrainer flow rate	Continuous

Figure 2: Workflow created by modeFRONTIER® and the input variables table

The studied model has seven Input variables and eight Output Variables. Increasing the recycle flow rate raises the product purity. On the other hand, it will cause a higher pump power consumption and the need for pipes with bigger diameters so TAC goes to increase. Increasing the number of stages, feed stage, and reflux/boil-up ratio improves

product purity, but the CapEx and OpEx increase. So the trade-off relation between the wastewater, product purity and TAC can be expected by changing the above variables.

5. Case Study

5.1. Preliminary investigation using pilOPT

As a preliminary study, 400 case study using Uniform Latin Hypercube DOE(Design of Experiments) and 100 case study using pilOPT, autonomous optimization alogrithm, is done to explore the relatively wide range of the design space. This work provides some insight into this process optimization like the minimum T1BR, sensitivity of T1FT.

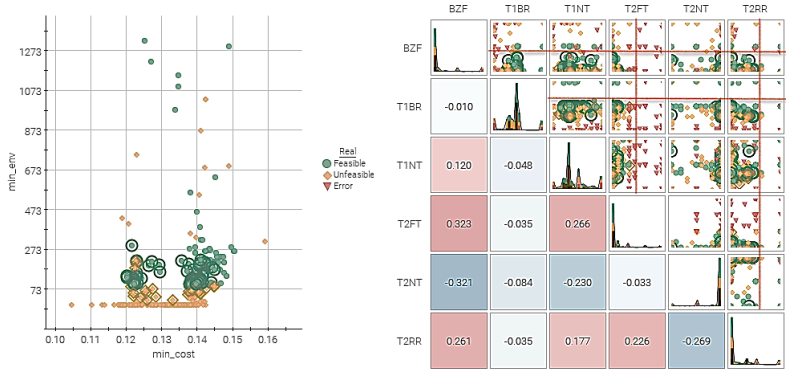
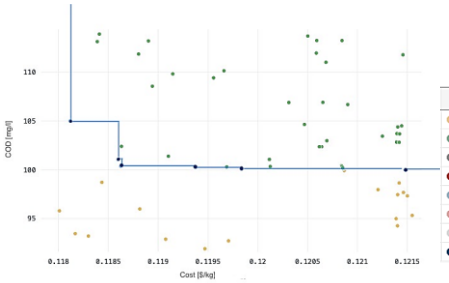


Figure 3: Scatter Matrix chart on Input Variables

5.2. Detailed study based on the above study

In order to reduce the domain of Input Variables, in the output domain, the designs of the most promising region, have been selected and this allowed to highlight their distribution in the space of inputs (Figure.3 Scatter Matrix chart on the right). From this study, the design space is narrowed down, shrinking the Upper Bound of T1BR and T2FT Variable. Few designs from the Pareto Front of this initial study have been selected and used as DOE for a second phase of optimization. Given the high nonlinearity of the model a Genetic Algorithm has been adopted. To speed up the optimization, we leveraged on two AVEVA™ Process Simulation running in parallel under the control of the same MOGA-II genetic algorithm. To run 900 designs it took little more than 25 hours and the resulting Pareto Front is visible at figure 4.



RID_CAT	BZF	T1BR	T1NT	T2FT	T2NT	T2RR	Cost	Purity	Waste
6	0.105301	5.01072	15.0000	3.00000	16.0000	2.85240	0.120857	0.997302	100.101
8	0.106813	5.01772	14.0000	3.00000	13.0000	2.76008	0.119377	0.991313	100.265
9	0.104266	5.01072	15.0000	3.00000	16.0000	2.85240	0.120849	0.997295	100.249
11	0.144902	5.01590	14.0000	3.00000	9.00000	2.77452	0.118632	0.992168	102.398
14	0.144902	5.01590	14.0000	3.00000	9.00000	2.77304	0.118624	0.992170	100.357
16	0.106813	5.01772	14.0000	3.00000	10.0000	2.76008	0.118597	0.991305	101.043
18	0.105321	5.01072	15.0000	3.00000	12.0000	2.85240	0.119840	0.997302	100.134
19	0.112385	5.01758	14.0000	3.00000	6.00000	2.75600	0.118116	0.991347	105.006

Figure 4: MOGA-II Pareto Front

6. Result

The Generative design approach produced a Pareto Front composed of eight designs (Figure.5), with a difference of ~3$/ton of produced ethanol between the minimum and maximum waste options. For each of these nominal designs (Figure. 4), to validate the robustness of each solution, a cloud of "Robust Designs" that follow a Gaussian distribution around the nominal design, has been generated automatically by modeFRONTIER® and validated with AVEVA™ Process Simulation.

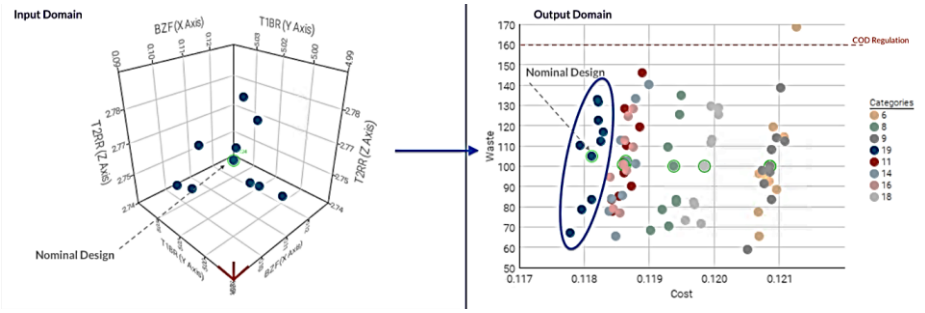


Figure 5: Robust design analysis

The result visible from Figure.5 is that even if very small noise is applied to BZF, T2RR and T1BR, it causes high variation in the Waste, while Cost and Purity are not affected so much. In particular, Design N. 6 has one Robust Design that reaches COD = 168 that is clearly out of regulation. Decision-makers should reject this design and select the N.19 instead that seems less sensitive to noise.

7. Conclusions

We have seen that several charts generated by modeFRONTIER® give some insights about the direction of the design and reduce the execution time of a parametric study thanks to its algorithms. Moreover, modeFRONTIER® identified several good designs that satisfy constraints and reduce environmental impact and total annual cost (TAC). We could reduce the COD by ~56% and the Cost of 21% with respect to the rough baseline design in just a few days. Furthermore, the project demonstrated that among several possible solutions feasible in a deterministic way, some of them may not satisfy requirements when considering noise affecting input variables. A probabilistic analysis helped to reject one of the eight designs present in the Pareto Front.

References

James Merrill Douglas, 1988, Conceptual Design of Chemical Processes, McGraw-Hill, 568-77
Ulrich, G.D.,Vasudevan, P.T., 2006, How to Estimate Utility Costs, Chemical Engineering, 66-69
Zacharias B. Maroulis, George D. Saravacos, 2007, Food Plant Economics, CRC Press, 95-97
L. Gerber, S. Fazlollahi, F. Maréchal, 2013, A systematic methodology for the environomic design and synthesis of energy systems combining process integration, Life Cycle Assessment and industrial ecology, Computers & Chemical Engineering, 59, 2-16
Brent Bishop and Fernando V. Lima, 2020, Modeling, simulation, and operability analysis of a nonisothermal, countercurrent, polymer membrane reactor, Processes, 8(1), 78
J. D. Seader, Ernest J. Henley, D. Keith Roper, 2010, Separation Process Principles with Applications using Process Simulators, Willey, 343-351

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50036-1

Reliability incorporated optimal process pathway selection for sustainable microalgae-based biorefinery system: P-graph approach

Juin Yau Lim[a], Akos Orosz[b], Bing Shen How[c], Ferenc Friedler[d], and ChangKyoo Yoo[a,*]

[a] *Integrated Engineering, Department of Environmental Science and Engineering, College of Engineering, Kyung Hee University, Republic of Korea*
[b] *Department of Computer Science and Systems Technology, Faculty of Information Technology, University of Pannonia, Hungary*
[c] *Biomass Waste-to-Wealth Special Interest Group, Faculty of Engineering, Computing and Science, Swinburne University of Technology, Jalan Simpang Tiga, 93350 Kuching, Sarawak Malaysia*
[d] *Széchenyi István University, H-9026 Győr, Egyetem tér 1, Hungary*
ckyoo@khu.ac.kr

Abstract

Biofuel from microalgae is one of the promising solutions on addressing climate change by its possibility of reducing the fossil fuel dependency. Till-date, the overall competitiveness of microalgae based biorefinery is the major concern due to its unique operational mechanism, especially the biological growth of microalgae that fluctuates towards the surrounding. Therefore, a novel graph-theoretic approach has been proposed to provide an optimization approach for identifying optimal process design with the consideration of three aspects that includes: economic, environmental, and reliability. The optimization is conducted using P-graph (a powerful graph-theoretic tool) which is capable to determine optimal and near-optimal solutions based on three objective functions: (i) minimizing annual operating cost, (ii) minimizing potential environmental impact, and (iii) maximizing reliability of process. The pool of feasible solutions (optimal and near-optimal) is obtained by satisfying the constraints on both greenhouse gas emissions and its respective reliability along. Thereupon, a further analysis was carried out with the aid of TOPSIS considering three of the assessment aspects to identify the optimal microalgae biorefinery configuration

Keywords: Optimization, Reliability, Microalgae biorefinery, Climate Change, P-graph

1. Introduction

Exponential growth of the worldwide population has led to the urge on exploring sufficient energy sources to suppress the effect of energy droughts that would bring negative impact on the evolution of human civilization. Till-date, the main energy supplies are still dominated by fossil-fuel based energy sources which are depleting at an unaccountable pace. One of the promising solutions on addressing such is by increasing the portion of renewable energy shares in the energy supplies. Among the conventional renewable energy sources, solar and wind energy have been recognized as the most promising alternative energy supplies that could be harvested from the nature. However, one of the critical issues reported with the aforementioned conventional renewable energy

sources is the intermittency in providing stable energy as it is highly dependent on the weather condition at the particular location where harvesting facilities are installed. Biofuel, which is another type of renewable energy that has received attention due to its similar calorific value while compared to mineral diesel and is further validated with its compatibility in ignition engine. Common biofuel production is involved with the crops and lignocellulosic biomass in agricultural activities which are categorized as 1st and 2nd generation biofuels, respectively. Yet, both input source of biofuel generation has the known issue of food security (1st generation) and extra processing steps are required for the lignocellulosic biomass (2nd generation).

Microalgae-based biofuel, which is known as the third-generation biofuels, are gaining its popularity especially with its capability in supplying lipids that is essential for producing biofuel *via* microalgae cultivation. A good cultivation condition (i.e., sunlight, nutrient, pH, and salinity) is expected for the microalgae cultivation to maximize the lipid content that is essential in producing biofuel. In norm, the monocultures of microalgae are used in practice where the species with high lipid content is preferable such as the *Chlorella vulgaris* that is commonly used in the industry as it has been reported with 40% to 53% of dry weight lipid content. However, the monocultures of certain microalgae species are often reported with extensive operational cost especially in maintaining the surrounding condition and preventing the cultivation farm from contamination. Therefore, the co-culturing of microalgae is more favourable as it allows the existence of multiple strains at one time by allowing healthy symbiotic relationship among species. Such cultivation strategy has been proven with the capability in enhancing biofuel production as the amount of biomass from cultivation are reported with significant enhancement.

Apart from the microalgae cultivation, a series of biorefinery process is required to upgrade the biomass harvested from the cultivation farm such as pre-treatment, lipid extraction, upgrade, and post-treatment. Due to the intensive energy requirement on producing biofuel, a proper process selection for the microalgae biorefinery is essential to ensure the overall sustainability on producing the biofuel. Herein, an extensive work was proposed to identify the optimal configuration for the biorefinery considering the aspects of economic, environment, and reliability in a single stage optimization framework. An optimization framework is developed to address the aim of this study with P-graph. The graphical aided optimization tool - P-graph is originally developed to solve the problem involving process network synthesis (PNS) alongside on providing a series of near-optimal solutions that is essential in decision making. Till recent, an innovative modification to the algorithm of P-graph was successfully proposed and implemented by Kovacs et al., (2019) which aims to solve the problem of process design with the consideration of its reliability simultaneously. Such reliability incorporated P-graph will be extensively implemented on the process design of microalgae biorefinery considering economic, environment, and reliability.

2. Methodology

A brief superstructure optimization framework targeted on proposing an optimal process configuration for microalgae biorefinery that considers economic, environment, and reliability aspects as shown in Figure 1. The raw material *i* required for certain process technology *j* in process stage *k* along with the final product *m* are identified accordingly. P-graph tool were originally developed to identify economical feasible solution of the problem declared and further providing a ranking of all feasible pool of solutions. Yet, certain modifications could be made to address the desired assessment criteria on the

problem. In such, Lim et al., (2021) has proposed a novel fertilizer selection with the aid of P-graph in consideration of the economic, environmental, and health aspects by aligning certain configuration towards the desired evaluation metrics. Similarly, a configuration was made to assess the economic and environmental aspects of the process configuration proposed alongside with the modification towards the P-graph algorithm by incorporating the reliability assessment which was firstly proposed by Kovacs et al., (2019).

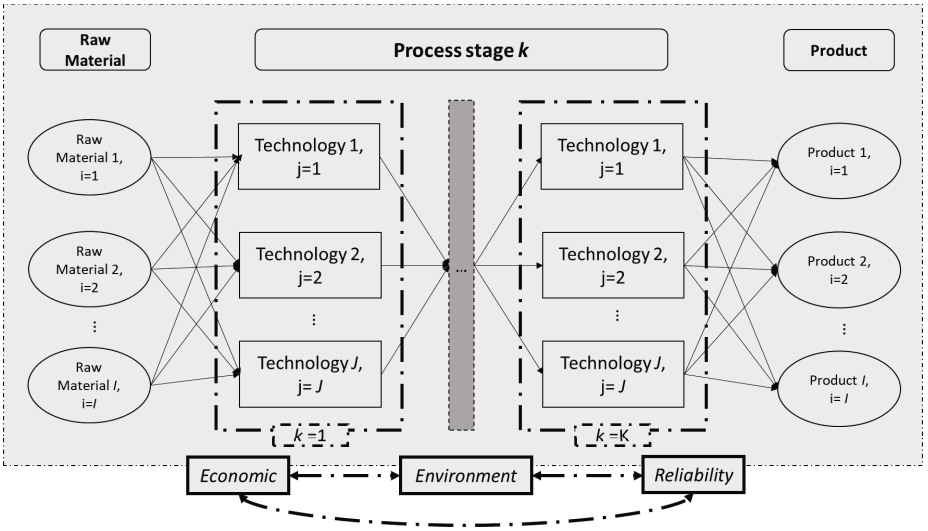


Figure 1. Microalgae process technology selection *via* superstructure optimization considering economic, environmental, and reliability aspects.

The objective function is expressed by minimizing the annual operating cost, *AOC* required for the process (See Eq. (1)) considering the required raw materials (RC_i) and the utility ($U_{n,j,k}$) alongside with its respective cost of materials (RC_i) and utility (UC_n). Whereas the greenhouse gas emissions, *GHG* was assessed similarly based on Eq. (2) in terms of materials and utility used with its respective emission factor of RGF_i, and GFU_n.

$$Min\ TAC = \sum_{i,j,k}(RC_i \times F_{i,j,k}) + \sum_{n,j,k}(UC_n \times U_{n,j,k}) + \sum_{i,j,k}(EC_{i,j,k} \times F_{i,j,k}) \quad (1)$$

$$GHG = \sum_{i,j,k}(RGF_i \times F_{i,j,k}) + \sum_{n,j,k}(GFU_n \times U_{n,j,k}) \quad (2)$$

Reliability analysis of the solutions is incorporated through the P-graph-based method given by Kovacs et al., (2019). The reliability formula, given by Eq (3), determines the $\hat{r}$ reliability from the reliabilities of the operations (p_n), where U represents the operational subnetworks in the solution, and x_n, the state of a unit.

$$\hat{\boldsymbol{r}} = \sum_{(x_1,x_2,\ldots,x_N)\in U} \left(\prod_{n=1}^{N} \boldsymbol{p_n}^{x_n}(\boldsymbol{1}-\boldsymbol{p_n})^{(1-x_n)}\right) \quad (3)$$

3. Development of P-graph model

The configuration of a generic superstructure optimization is demonstrated in Figure 2(a) for selecting feasible processes among two different technology. As previously mentioned, the reliability assessment has been incorporated into the P-graph which a new configuration on such modification is displayed in Figure 2(b). Herein, three different types of process configuration were considered for the specific technology that avail for the selection. Type A: a single equipment with the size x is considered, Type B: three similar size equipment with the halving of size A ($\frac{x}{2}$), and Type C: two similar equipment with size x. The difference in configuration could allow the system to propose an optimal configuration that considers the reliability.

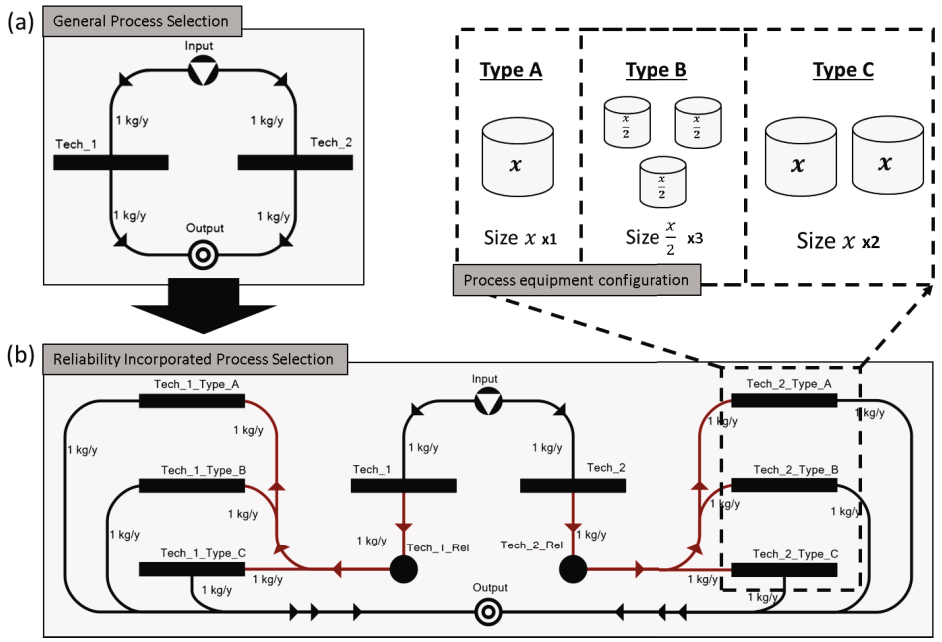


Figure 2. Demonstration of generic process selection superstructure optimization in P-graph: (a) general configuration and (b) configuration after incorporating reliability.

A P-graph model is developed to identify the optimal process in microalgae biorefinery as shown in Figure 3. Such process selection considers the aspects of technical, economic, environment, and reliability which are discussed in Section 2. The P-graph model comprises eight different sections covering: (A) Cultivation, (B) Dewatering and Cell Disruption, (C) Lipids extraction, separation, and recovery, (D) Anaerobic Digestion, (E) Digestate purification, (F) Biogas upgrade, (G) greenhouse gas emissions, and (H) Electricity distribution. Such model developed is solved with the custom solver of combining solution structure generation, linear programming, and reliability analysis. A case study was implemented where an approximation of 6.82 t of dry microalgae are cultivated in an area of 200 ha in Incheon, South Korea. All the process selection parameters considered in this study were based on the one reported by Lim et al., 2020. Due to strategic location of Incheon, various renewable energy sources are avail to be incorporated to fulfil the electricity demand of the microalgae biorefinery.

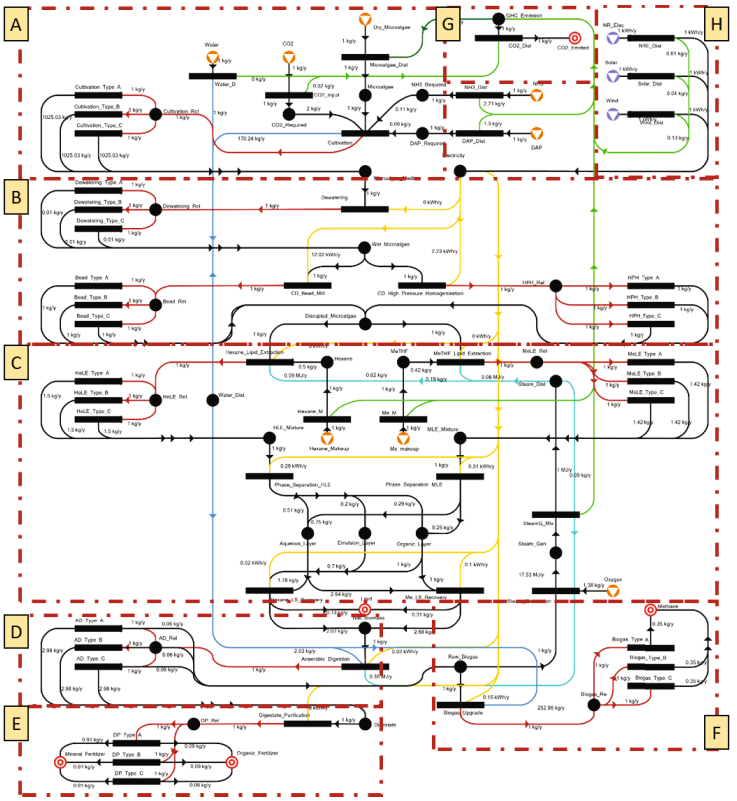


Figure 3. One-stage technical-economical-environmental-reliability P-graph model targeting microalgae biorefinery.

4. Results and discussion

A total of 34,992 solutions were proposed by the P-graph model developed based on the modified solver that incorporated the reliability alongside the calculation (see Figure 4). The total annual cost from the solution pool has reported with the range of 9.04×10^7 to 1.35×10^8 USD/y; whereas the GHG emissions were reported in the range of 1.62×10^5 to 7.24×10^5 kg $CO_{2\text{-eqv}}$/y. As of the incorporated reliability aspects that bounds within 0 (lowest reliability) to 1 (highest reliability), the solution has reported with the reliability range of 0.35 to 0.86. Thereupon, a further analysis was conducted with the aid of *Technique of Preference by Similarity to Ideal Solution* (TOPSIS) that was first proposed by Hwang and Yoon (1981). A detailed calculation steps of TOPSIS can be found similarly in Lim et al., (2021) which a non-bias approach has been implemented where all assessment aspects are considered equally important. Top 100 solutions that are re-ranked according to TOPSIS score are highlighted in Figure 4. The ranked 1 microalgae biorefinery configuration is reported with 9.69×10^7 USD/y, 1.81×10^5 kg $CO_{2\text{-eqv}}$/y, and 0.81 in respect of the total annual cost, GHG emission, and reliability, respectively. The process configuration of the ranked 1 solution is as: cultivation, dewatering, high pressure homogenization, hexane lipid extraction, phase separation, solvent recovery, anaerobic digestion, digestate purification, and biogas upgrade.

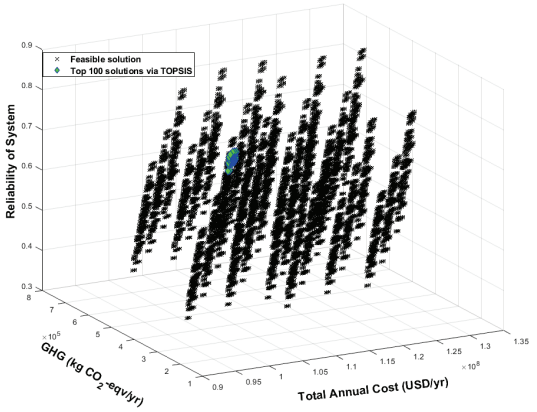


Figure 4. Solution pools generated from P-graph model for the microalgae biorefinery process selection alongside with the top 100 solutions highlighted from TOPSIS.

5. Conclusions

This study has successfully proposed a one stage technical-economical-environmental-reliability P-graph model on the selection of sustainable microalgae biorefinery process configuration. TOPSIS is then implemented to identify the ranked 1 configuration which resulted with 9.69×10^7 USD/y, 1.81×10^5 kg $CO_{2\text{-eqv}}$/y, and 0.81 in respect of the total annual cost, GHG emission, and reliability, respectively.

Acknowledgements
The authors would like to acknowledge financial support from a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIP) (No. 2021R1A2C2007838), and the Korea Ministry of the Environment (MOE) as a Graduate School specializing in Climate Change. The research presented in this paper was funded by the "National Laboratories 2020 Program – Artificial Intelligence Subprogram – Establishment of the National Artificial Intelligence Laboratory (MILAB) at Széchenyi István University (NKFIH-870-21/2020)" project. Project TKP2020-NKA-10 has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the 2020-4.1.1-TKP2020 Thematic Excellence Programme 2020 - National Challenges sub-program funding scheme.

References

C-L. Hwang, K. Yoon, 1981, Methods for multiple attribute decision making. In: Multiple attribute decision making. Springer, pp 58–191

J.Y. Lim, B.S. How, S.Y. Teng, W.D. Leong, J.P. Tang, H.L. Lam, C.K. Yoo, 2021, Multi-objective lifecycle optimization for oil palm fertilizer formulation: A hybrid P-graph and TOPSIS approach, Resour Conserv Recycl, 166, 105357.

J.Y. Lim, K.J. Nam, C.K. Yoo, 2020, Circular economy assessment towards optimal process configuration of microalgae-based bio-refinery system with consideration of reliability assessment: P-graph approach, PRES'20, Xi'an, China.

K.L. Yeh, J.S. Chang, 2012, Effects of cultivation conditions and media composition on cell growth and lipid productivity of indigenous microalga Chlorella vulgaris ESP-31, Bioresour. Technol. 105, 120–127.

Z. Kovacs, A. Orosz, F. Friedler, 2019, Synthesis algorithms for the reliability analysis of processing systems, Cent Eur J Oper Res, 27, 573–595.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50037-3

Framework for Designing Solid Drug Product Manufacturing Processes Based on Economic and Quality Assessment

Kensaku Matsunami[a*], Sara Badr[a], and Hirokazu Sugiyama[a]

[a]*Department of Chemical System Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*
kensaku.matsunami@ugent.be

Abstract

This work presents a design framework for solid drug product manufacturing processes based on economic and quality assessment. Process alternatives were generated using a superstructure, which includes novel continuous manufacturing. Each generated alternative can be assessed considering various uncertainties in the design phase. Economic assessment calculates a net present value from the decision stage to the end of commercial production. Product quality assessment predicts dissolution behavior by surrogate models developed from existing mechanistic models. In this work, the use of the tools was presented in the form of an activity model and demonstrated in a case study. The proposed framework and the assessment tools can assist rational and efficient simulation-based design of solid drug product manufacturing processes.

Keywords: Pharmaceuticals, continuous manufacturing, superstructure, economic assessment, quality.

1. Introduction

Solid drug products, e.g., tablets and capsules, are major products in the pharmaceutical industry. Solid drug product manufacturing consists of powder-based unit operations with the active pharmaceutical ingredient (API) as the initial raw material. Examples of unit operations are mixing, granulation, drying, compression, and coating, which have been traditionally performed batch-wise. Recently, continuous manufacturing, where all unit operations are interconnected, has been developed as a novel technology in pharmaceutical manufacturing. Unlike in the chemical industry, continuous solid drug product manufacturing is generally performed within certain running hours, e.g., 12 h. This characteristic increases flexibility in demand change (Lee et al., 2015), but the start-up operation has high impacts on material losses. The number of potential process alternatives has increased with the emergence of continuous manufacturing, which makes process design more complicated.

Numerous studies have been performed toward rational design of solid drug product manufacturing considering continuous manufacturing. Experimental investigations of continuous manufacturing have been reported to identify critical process parameters (e.g., Liu et al., 2019). Economic assessment to compare between batch and continuous manufacturing has been developed focusing on tablet manufacturing using wet granulation (Matsunami et al., 2018). Regarding product quality assessment, mechanistic modeling has been proposed by many researchers. Van Hauwermeiren et al. (2018) established a population balance model for continuous twin-screw wet granulation; Metta

et al. (2019) proposed a flowsheet model for continuous wet granulation. However, these studies have focused on specific manufacturing processes and/or drug types. A comprehensive design approach applicable for any drug is still needed.

This study proposes a design framework for solid drug product manufacturing along with assessment models. A superstructure for alternative generation is presented with models for economic and product quality assessment. The use of assessment tools is then presented in the form of an activity model. This work describes the comprehensive design strategy, details of individual tools can be found in the article paper (Matsunami et al., 2020) and the book chapter (Matsunami et al., 2021) by the authors' research group.

2. Developed assessment tools

Tools are developed for process synthesis with alternative generation from a superstructure. The tools also include subsequent economic and quality assessments.

2.1. A superstructure for process synthesis

Process alternatives of solid drug product manufacturing are comprehensively generated using a superstructure created based on the unit, port, conditioning stream (UPCS) representation (Wu et al., 2018), as shown in Figure 1. Characters B and C in general units represent batch and continuous mode in the unit operations. Each process alternative is defined as the combination of streams, ports, and units from source (providing raw materials) to sink (collecting final products) units.

The superstructure in Figure 1, which was created through a literature survey as well as the expert knowledge of the pharmaceutical industry, yields 9,452 process alternatives. Among these alternatives, 1,261 alternatives were identified as "continuous technology," where all unit operations were performed in continuous operation mode. This work uses the granulation step, one of the essential production steps, as a case study for applying the assessment tools. Therefore, a higher-resolution of the available options was used compared to the other steps. In general, there are three options within this step: wet

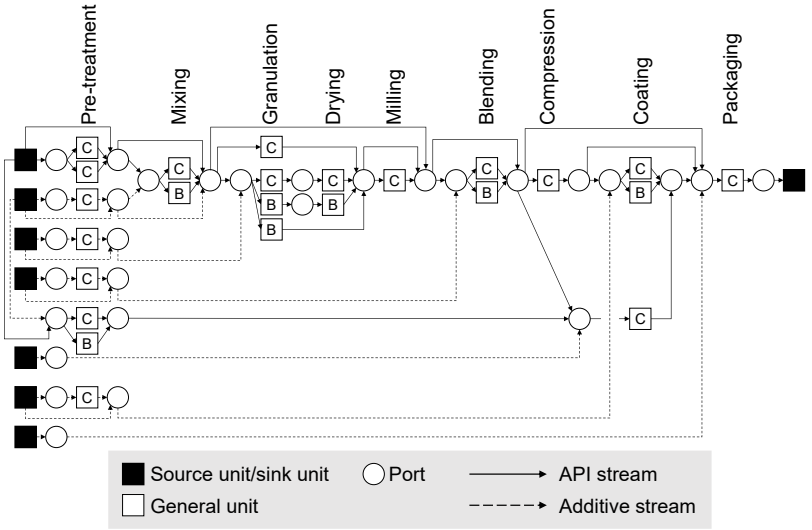


Figure 1. The developed superstructure (Matsunami et al., 2020)

granulation, dry granulation, or direct compression methods. Each option is associated with a range of options for unit operations.

2.2. Economic assessment

The indicator for the economic assessment is the net present value (NPV) to consider the cash flow from a decision stage to the end of the commercial production. Assuming the common decision stage of process alternatives is the beginning of phase II in clinical development, NPV can be defined as shown in Eq.(1):

$$NPV = -\sum_{\tau=0}^{\tau_3} \frac{C_{\mathrm{dev}}(\tau)}{(1+r)^{\tau}} - \sum_{\tau=0}^{\tau_{\mathrm{prod}}} \frac{C_{\mathrm{invest}}(\tau)}{(1+r)^{\tau}} + \sum_{\tau=\tau_{\mathrm{lau}}}^{\tau_{\mathrm{prod}}} \frac{C_{\mathrm{sales}}(\tau) - C_{\mathrm{op}}(\tau)}{(1+r)^{\tau}} \tag{1}$$

where C_{dev} [USD yr^{-1}], C_{invest} [USD yr^{-1}], C_{sales} [USD yr^{-1}], and C_{op} [USD yr^{-1}] represent the development cost, investment cost, sales, and operating cost, respectively. The parameter τ [yr] represents the period from the decision stage to the target phase. The subscripts 3, lau, and prod represent the clinical trials in phase III, the product launch, and the end of the commercial production, respectively. The calculation of NPV can be made by setting the interest rate r. Material losses at the start-up operation are included in C_{op}. Variations in the production scales to accommodate changing demand are reflected in this calculation.

To reflect uncertainty in the decision stage in the economic assessment, the stochastic optimization problem of the expected (E) NPV was determined, as shown in Eq.(2):

$$\begin{aligned} &\max E_{\boldsymbol{\theta}}\big(NPV(l)\big) \\ &\text{s.t.} \\ &E_{\boldsymbol{\theta}}\big(NPV(l)\big) > 0 \\ &\text{(Mass balance constraints)} \\ &\text{(Processing time constraints)} \\ &\text{(Pharma-specific constraints)}\,, \end{aligned} \tag{2}$$

where the design variable and uncertainty parameters are process alternative l and the vector $\boldsymbol{\theta}$, respectively. The developed economic assessment has been implemented as a part of a software tool "SoliDecision" (Matsunami et al., 2020).

2.3. Product quality assessment

As an example of product quality assessment, the dissolution behavior was studied. In this work, dissolution behavior is defined as the profile of mass ratio of API dissolved in water during a dissolution test. It was chosen as an example since it represents a critical quality attribute of solid drug products with a direct impact on drug efficacy. Surrogate modeling was chosen because both applicability to new drugs and calculation speed are important for the assessment in process design. The modeling activities consist of the four steps: (i) flowsheet model development, (ii) input/output data generation, (iii) dissolution behavior fitting, and (iv) random forest regression. Firstly, existing mechanistic models

are integrated to create flowsheet models of solid drug product manufacturing. The flowsheet models were then used for the calculation of dissolution behavior by changing the values of input model parameters. After obtaining a set of output data, mass ratio of API dissolved D [%] were fitted by the Weibull model to transfer it into a lower dimension, as shown in Eq.(3):

$$D(t) = 100 \cdot \left(1 - e^{\{-k(t-t_0)^b\}}\right) \tag{3}$$

where the parameters t [min], k [–], t_0 [min], and b [–] represent time from the start of a dissolution test, reciprocal of the scale parameter, time lag, and the shape parameter, respectively. The Weibull model was chosen because it showed the highest fitting accuracy among seven popular dissolution fitting models, e.g., first-order kinetics. Finally, the relationships between the input model parameters and the Weibull model parameters were trained by random forest regression. The hyper-parameters of random forest models were chosen by maximizing coefficients of determination in five-fold cross-validation.

By integrating all the steps, the surrogate model can be expressed as the combination of the Weibull model W and the random forest regression $\boldsymbol{g}$, as shown in Eq.(4):

$$D(t) = W(\boldsymbol{g}(\boldsymbol{P}), t) \tag{4}$$

where $\boldsymbol{P}$ represents the vector of the input model parameters. The developed surrogate model can calculate expected ranges of dissolution behavior under the uncertainty of the input model parameters.

3. Design framework

The application of the assessment tools in the design activities was described as a design framework by using the type zero method of integration definition for function modeling (IDEF0). The top activity of the framework is “evaluate processes for process design,” where the viewpoint was set as process designers in the pharmaceutical industry (Matsunami et al., 2021). The four sub-activities of the top activity were defined as shown in Figure 2. The sub-activities are controlled by design cases, which are defined by new drug information as well as the pharma-specific constraints, e.g., regulations and clinical trial results. The developed assessment tools are used as a mechanism of the activities. By executing the activities, promising alternatives are determined, which will be further tested in the subsequent design activities.

The proposed framework was demonstrated by setting a scenario where the beginning of phase II in clinical development was set as the decision stage. The design problem was set as “find a process which maximizes NPV and dissolution rate.” In A1, possible process alternatives were chosen based on material properties and design policy. Here, 32 alternatives, including batch/continuous dry granulation and wet granulation methods, were focused on in this demonstration. After defining probability density functions (PDFs) of all input parameters in A2, both economic and product quality assessments were performed. Figure 3 shows violin plots of PDFs of NPV differences between target

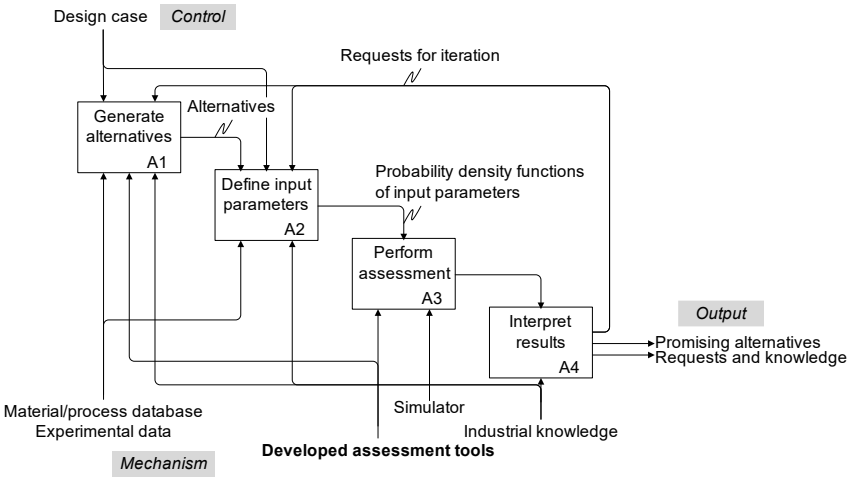


Figure 2. Design framework for solid drug product manufacturing processes

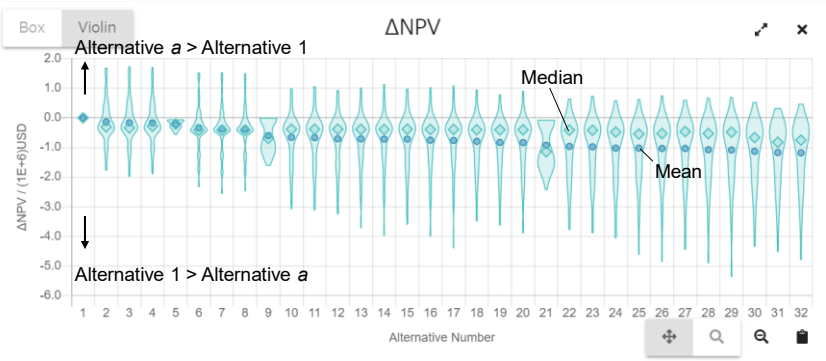


Figure 3. Economic assessment result using SoliDecision.

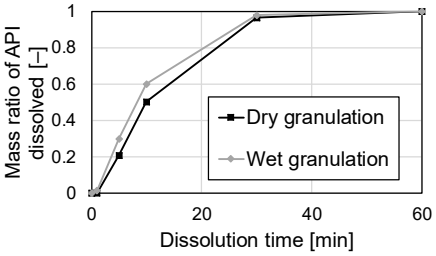


Figure 4. Simulation results of dissolution behavior using average values.

alternatives and the best alternative in terms of $E_{\boldsymbol{\theta}}\big(NPV(l)\big)$ generated by "SoliDecision". Alternative numbers are defined by the order of $E_{\boldsymbol{\theta}}\big(NPV(l)\big)$, where an alternative of continuous dry granulation was the best alternative. In product quality assessment, dissolution behavior in continuous dry granulation was compared with that in batch high-shear wet granulation, which is one of the typical production methods. Dissolution behaviors using average values are presented in Figure 4. Batch high-shear wet granulation made dissolution faster than continuous dry granulation. The results of

economic and product quality assessments were interpreted in A4. At commercial scale, the expected NPV of continuous dry granulation was USD3.37×10^5 higher than that in batch high-shear wet granulation, whereas mass ratio of API dissolved at 10 mins was 9.86% slower. This trade-off between cost and quality should be considered by weighting each indicator to determine the promising alternatives. Continuous dry granulation could be chosen if lowering cost was more important than fast dissolution. Batch wet granulation could be selected if a higher dissolution profile was the most important.

4. Conclusions

A design framework for solid drug product manufacturing processes was presented along with economic and product quality assessment tools as a new mechanism. Process alternatives were comprehensively generated using a superstructure and can be assessed in terms of NPV and dissolution behavior. Both assessments can propagate the uncertainty of the input parameters, which should be high in the design phase, into the results. The application of the tools was described by an activity model and demonstrated in the case study. A part of the tools has been implemented as software, "SoliDecision," which is ready for use in the actual decision-making in the pharmaceutical industry.

Acknowledgment

The authors acknowledge Dr. Hiroshi Nakagawa, Dr. Shuichi Tanabe, and Mr. Keita Yaginuma from Daiichi Sankyo Co., Ltd. for supporting the development of the assessment tools. H. S. is thankful for the financial supports by Grants-in-Aid for Young Scientists (A) [number 17H04964] and Grant-in-Aid for Scientific Research (B) [number 21H01704] from the Japan Society for the Promotion of Science (JSPS). K. M. appreciates the financial support by Research Fellows [number 18 J22793] from JSPS.

References

S. Lee, T. O'Connor, X. Yang, C. Cruz, S. Chatterjee, R. Madurawe, C. Moore, L. Yu, J. Woodcock, 2015. Modernizing pharmaceutical manufacturing: from batch to continuous production, *Int. J. Pharm.*, 10, 191–199.

H. Liu, B. Ricart, C. Stanton, B. Smith-Goettler, L. Verdi, T. O'Connor, S. Lee, S. Yoon, 2019. Design space determination and process optimization in at-scale continuous twin screw wet granulation, *Comput. Chem. Eng.*, 125, 271–286.

K. Matsunami, S. Badr, H. Sugiyama, 2022. Design framework and tools for solid drug product manufacturing processes, *Optimization of Pharmaceutical Processes*, accepted.

K. Matsunami, T. Miyano, H. Arai, H. Nakagawa, M. Hirao, H. Sugiyama, 2018. Decision support method for the choice between batch and continuous technologies in solid drug product manufacturing, *Ind. Eng. Chem. Res.*, 57, 9798–9809.

K. Matsunami, F. Sternal, K. Yaginuma, S. Tanabe, H. Nakagawa, H. Sugiyama, 2020. Superstructure-based process synthesis and economic assessment under uncertainty for solid drug product manufacturing, *BMC Chem. Eng.*, 2, 6.

D. Van Hauwermeiren, M. Verstraeten, P. Doshi, M. am Ende, N. Turnbull, K. Lee, T. De Beer, I. Nopens, 2019. On the modelling of granule size distributions in twin-screw wet granulation: calibration of a novel compartmental population balance model. *Powder Technol.*, 341, 116–125.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50038-5

Marine flue gas desulfurization processes: recent developments, challenges, and perspectives

Gwangsik KIM[a,1], Van Duc Long NGUYEN[a,1], Dongyoung LEE[a,b,1], Yujeong LEE[a], Jonghoon BAEK[a], Wonseok JEONG[a], Myungjin KIM[b], Choongyong KWAG[b], Youngmok LEE[b], Sungwon LEE[b], and Moonyong LEE[a*]

[a]*School of Chemical Engineering, Yeungnam University, Gyeongsan 712-749, South Korea*
[b]*Hanbal Masstech Ltd, holden root complex, Gimhae, South Korea*
shikki47@gmail.com

Abstract

Seawater flue gas desulfurization (SWFGD) is considered to be a viable solution for marine and coastal applications. SWFGD, however, still has a couple of drawbacks that have to be resolved. High pumping costs and poor mass transfer efficiency require large volume and heavy construction. Thus, intensified and advanced process configurations have become necessary in process industries to improve the SWFGD processes performance. This work presents an overview of several advanced SWFGD systems such as swirling gas flow, square-based shapes scrubber, rotating packed bed and addition of a pre-scrubber. This work also presents an overview of industrial uses, challenges, and improved SWFGD processes.

Keywords: seawater flue gas desulfurization (SWFGD); process retrofit; process improvement; Heat integration

1. Introduction

As the energy demand worldwide is overgrowing, many power plant constructions using fossil fuel are increasing [1]. Sulfur oxides (SO_x), mainly consisting of SO_2, are generated and emitted when fossil fuel is combusted, negatively influencing the environment [2]. As a result, many countries have propagated strict SO_x emissions regulations [3].

Furthermore, sea transport, which accounts for more than 90% of international trade [4], emits a considerable amount of SO_x [5]. The International Marine Organization (IMO) recently consented to regulations of sulfur emissions to prevent the negative effect on the environment from its emissions [6].

Many power plants prefer to be built offshore because they require lots of water for cooling. In addition, owing to its high availability and natural alkalinity, seawater has been considered an environmentally and economically reliable candidate for the solvent of flue gas desulfurization (FGD) processes in offshore and maritime applications [7].

However, the seawater FGD (SWFGD) process has a couple of drawbacks that have to be resolved. High pumping costs and poor mass transfer efficiency require large volume and heavy construction [8]. Due to the typical limitations of maritime applications, these drawbacks must be investigated and addressed [6]. Consequently, a lot of research has been performed to improve SWFGD process efficiency and capacity and make this process more compact and lighter due to the constraints of space and weight for maritime

applications. This paper focuses on producing a comprehensive review of the improvement of SWFGD, which has been given significant attention recently.

2. SWFGD systems in a coastal area

The flue gas from power plants' combustion of coal contains approximately between 210 to 1540 ppm_v of SO_2 [7], and typical seawater usually has a pH value of 7.6 to 8.4 at a temperature of 5 to 15 °C The reaction procedures of absorption of SO_2 into seawater are tabulated in Table 1 [2,9].

Table 1. SO_2 absorption reactions into seawater

Procedure	Reaction	
Absorption	$SO_2 + H_2O \leftrightarrow HSO_3^- + H^+$	(1)
Oxidation	$HSO_3^- + 1/2O_2 \leftrightarrow SO_4^{2-} + H^+$	(2)
Neutralization	$HCO_3^- + H^+ \leftrightarrow CO_2 + H_2O$	(3)
	$CO_3^{2-} + 2H^+ \leftrightarrow CO_2 + H_2O$	(4)

Figure 1 shows simple configurations of open-loop, closed-loop, and hybrid mode. In the open-loop mode in Figure 1a, raw seawater is fed into the scrubber as a solvent, absorbs SO_2, and neutralizes using its natural carbonate ions (CO_3^{2-}) and bicarbonate ions (HCO_3^-). As shown in Figure 1b, in closed-loop mode, the seawater or freshwater is not discharged into the sea but is recycled inside the scrubber system. Hence, a wash water unit is necessary before releasing the water. Before recycling the wash water, use a heat exchanger to lower its temperature. To remove the moisture in the gas from the scrubber, it goes through a demister or moisture eliminator before entering the stack in both the open and closed loops.

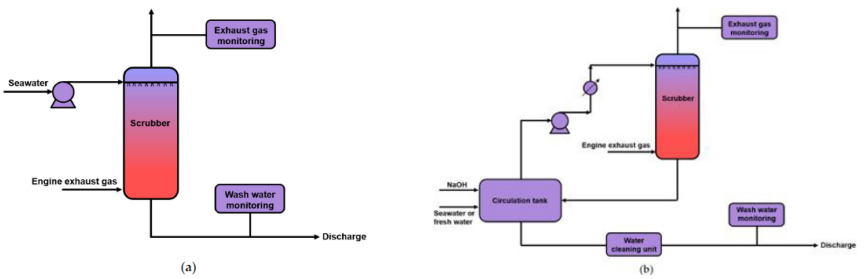


Figure 1. Schematic configurations of the (a) open-loop, (b) closed-loop FGD process mode

3. Maritime SWFGD systems

The fuel gas from the engines of existing ships typically emits SO_2 in concentrations of 80 to 1000 ppm_v [10,11]. Table 1 shows the typical desulfurization process design conditions of the land-based SWFGD system and the SWFGD system used in large ships [9].

Table 2. The typical desulfurization process design conditions of the land-based SWFGD system and the SWFGD system used in large ships

		Land-based	**Large Ships**
Flue gas flow rate	Nm^3/h	600,000 to 4,000,000	23,000 to 540,000

Inlet SO_2 level	ppm_d	100 to 1800	700
Outlet SO_2 level	ppm_d	10 to 220	20
SO_2 removal efficiency	%	75 to 98	97.1 (3.5 %S to 0.1%S): SECAs 85.7 (3.5%S to 0.5%S): global sea areas excluding SECAs
Regulatory items for seawater discharge		pH, dissolved oxygen (DO), temperature, etc	pH, PAH, turbidity, nitrates

For long-distance sea transportation, the SWFGD system with scrubber has many advantages such as simple structure, easy operation, and low initial investment compared to using low-sulfur oil or replacing it with an LNG vessel [12,13]. It usually consists of a spray column using seawater or seawater with NaOH added as a solvent [6], and this column is generally utilized when pressure drop is a critical factor [14] or when high levels of separation are not required [15].

Nevertheless, the scrubber still has some disadvantages that need to be addressed. High pumping costs and poor mass transfer efficiency require large volume and heavy construction [8]. Due to the typical limitations of maritime applications, these drawbacks must be investigated and addressed [6]. So possible developed and integrated process configurations have become necessary in process industries to improve the SWFGD processes performance.

4. Improvement of water and SWFGD systems

If the swirling gas flow concept is used for the flow of gas entering into the spray column, the effect of mass transfer is increased [16]. When this concept is applied, the residence time of the flue gas inside the column becomes longer, resulting in improved mass transfer performance. Figure 2 shows the conceptual diagram of the swirling gas flow of the SOx scrubber [17]. Recently, Schrauwen and Toenes studied the effect of mass transfer after generating a swirling gas flow by injecting flue gas tangentially into the column [18]. The swirling gas flow increases the mass transfer performance, allowing a compact scrubber design and consequently expanding the applicability to marine vessels.

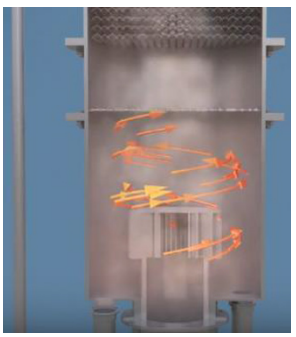


Figure 2. The conceptual diagram of the swirling gas flow of the Sox scrubber.[17]

A square scrubber [19], which is more efficient in terms of volume than a cylindrical scrubber, is frequently applied to the FGD system [9]. It is because the diameter of the cylindrical scrubber is larger than the side of the square scrubber in the same area. However, the square scrubber has difficulty in evenly distributing the flue gas and can only be used when the operating pressure is low [15,20]. PacificGreem Technologies

recently developed a square scrubber that is compact, flexible and does not compromise efficiency [17].

Hansen [8] has developed a method of installing an additional pre-scrubber to effectively remove SO_2 in flue gas from engines of marine vessels. The flue gas temperature is rapidly cooled from about 180-250 °C to 45-60 °C in the pre-scrubber before flowing into the main scrubber. As shown in Figure 3, Alfa Laval installed a jet scrubber that uses water as a coolant to lower the flue gas temperature before the primary scrubber [21]. A venturi scrubber can be used as a pre-scrubber to screen PM [22], but this will increase the pressure drop. The venturi scrubber can be used to cool the hot flue gas (up to 1000 °C) [17].

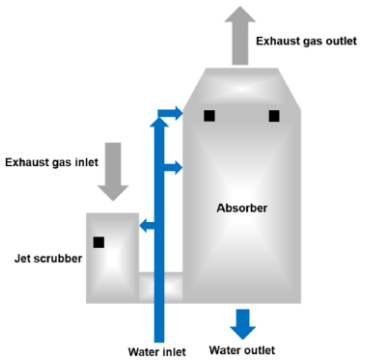


Figure 3. Schematic configuration combining a jet scrubber and an absorber [21].

Rotating packed bed (RPB) generates high acceleration through centrifugal force and forms a thin liquid film and tiny droplets through centrifugal acceleration to improve mass transfer performance [23]. As a result, RPB can improve the removal efficiency of SO_2 [24]. Recently, research has been conducted to enhance the absorption of SO_2 using RPB and ionic liquid [25].

5. Challenges

a. Due to the acidity of the flue gas, the scrubber must consider the effect of corrosion [6]; thus, it is crucial to choose a suitable material that can prevent decay.

b. Since the flue gas pressure from the marine engine is similar to the atmospheric pressure, the pressure drop due to the auxiliary internals such as the demister and gas distributor inside the column should be low [6].

c. Designers must consider essential design factors such as the material of constructions, nozzle droplet size or nozzle type, solvent flow rate, and pressure drop to design an appropriate scrubber. [17].

d. The flue gas from a marine engine contains a complex PM composed of carbon particles [10], especially the soot particles that can cause severe disease [26]. A venturi scrubber, cyclone, dust collector, or electrostatic precipitator can be used to prevent PM from entering the scrubber [27,28].

f. The pH of seawater that absorbs acid gases from marine engines can be reduced to a range of around 2.53 [13]. However, the 2015 IMO guideline regulates the pH of the discharged seawater to be above 6.5. To increase the pH of the discharged seawater, caustic soda can be used as an additive in the SWFGD system [29].

6. Conclusion

Considering the strict IMO environmental regulations worldwide, this paper succeeded in investigating and analyzing SWFGD's current researches and industrial applications. The SWFGD process can be an excellent alternative for marine applications to meet the stringent IMO regulations due to seawater's natural alkalinity and high availability, design simplicity, convenient operation, no chemical solvent requirement, no solid waste, and relatively higher performance of SO_2 removal. Several solutions were assessed, including process modification, integration, and intensification to enhance the SWFGD process effectively. Developed and improved SWFGD systems with compact units, lighter construction, low energy consumption, and reduced seawater flowrate have been commercialized. However, several issues are still to be solved, such as establishing an accurate and broader range of SO_2 solubility, equilibrium data, and developing models describing advanced and improved scrubbers.

7. Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2021R1A2C1092152) and by the Priority Research Centers Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2014R1A6A1031189).

8. References

[1] H.N. Soud, Z. Wu, East Asia-air pollution control and coal-fired power generation, (1998).

[2] K. Oikawa, C. Yongsiri, K. Takeda, T. Harimoto, Seawater flue gas desulfurization: Its technical implications and performance results, Environ. Prog. 22 (2003) 67–73.

[3] C. Zhang, L. Yang, One-dimensional simulation of synergistic desulfurization and denitrification processes for electrostatic precipitators based on a fluid-chemical reaction hybrid model, Energies. 11 (2018) 3249.

[4] R.F. Nielsen, F. Haglind, U. Larsen, Design and modeling of an advanced marine machinery system including waste heat recovery and removal of sulphur oxides, Energy Convers. Manag. 85 (2014) 687–693.

[5] Z. Wan, M. Zhu, S. Chen, D. Sperling, Pollution: three steps to a green shipping industry, Nat. News. 530 (2016) 275.

[6] D. Flagiello, A. Parisi, A. Lancia, C. Carotenuto, A. Erto, F. Di Natale, Seawater desulphurization scrubbing in spray and packed columns for a 4.35 MW marine diesel engine, Chem. Eng. Res. Des. 148 (2019) 56–67.

[7] D. Flagiello, A. Erto, A. Lancia, F. Di Natale, Experimental and modelling analysis of seawater scrubbers for sulphur dioxide removal from flue-gas, Fuel. 214 (2018) 254–263.

[8] J.P. Hansen, Scrubber System and Method Technical Field, WO. 45272 (2013) A1.

[9] R. Sasaki, T. Nagayasu, T. Shingu, Y. Watanabe, T. Mori, H. Sakurai, Practical design of marine SOx scrubber for mega-container ships, Mitsubishi Heavy Ind. Tech. Rev. 56 (2019) 1–8.

[10] F. Di Natale, C. Carotenuto, Particulate matter in marine diesel engines exhausts: Emissions and control strategies, Transp. Res. Part D Transp. Environ. 40 (2015) 166–191.

[11] EPA, Analysis of commercial marine vessels emissions and fuel consumption data, Tech. Rep. (2000).
[12] J. Rodríguez-Sevilla, M. Álvarez, M.C. Díaz, M.C. Marrero, Absorption equilibria of dilute SO_2 in seawater, J. Chem. Eng. Data. 49 (2004) 1710–1716.
[13] A. Andreasen, S. Mayer, Use of seawater scrubbing for SO_2 removal from marine engine exhaust gas, Energy & Fuels. 21 (2007) 3274–3279.
[14] N.K. Yeh, G.T. Rochelle, Liquid-phase mass transfer in spray contactors, AIChE J. 49 (2003) 2363–2373.
[15] A. Cousins, L. Wardhaugh, A. Cottrell, Absorption-Based Post-Combustio n Capture of Carbon Dioxide, Woodhead Publ. (2016) 649–684.
[16] K.H. Javed, T. Mahmud, E. Purba, Enhancement of mass transfer in a spray tower using swirling gas flow, Chem. Eng. Res. Des. 84 (2006) 465–477.
[17] N.V.D. Long, D.Y. Lee, K.M. Jin, K. Choongyong, L.Y. Mok, L.S. Won, M. Lee, Advanced and intensified seawater flue gas desulfurization processes: Recent developments and improvements, Energies. 13 (2020) 5917.
[18] F.J.M. Schrauwen, D. Thoenes, Selective gas absorption in a cyclone spray scrubber, in: Tenth Int. Symp. Chem. React. Eng., Elsevier, 1988: pp. 2189–2194.
[19] N.V.D. Long, D.Y. Lee, M.J. Kim, C. Kwag, Y.M. Lee, K.J. Kang, S.W. Lee, M. Lee, Desulfurization scrubbing in a squared spray column for a 720 kW marine diesel engine: Design, construction, simulation, and experiment, Chem. Eng. Process. Intensif. 161 (2021) 108317.
[20] P.-E. Just, Advances in the development of CO2 capture solvents, Energy Procedia. 37 (2013) 314–324.
[21] MAN, Emission project guide, MAN Energy Solut. (2017) 111.
[22] Wärtsilä, SOx scrubber technology, (2017) 8.
[23] C. Ramshaw, R.H. Mallinson, Mass transfer process, (1981).
[24] L. Zhang, S. Wu, Y. Gao, B. Sun, Y. Luo, H. Zou, G. Chu, J. Chen, Absorption of SO_2 with calcium-based solution in a rotating packed bed, Sep. Purif. Technol. 214 (2019) 148–155.
[25] Y.-Z. Liu, W. Wu, Y. Liu, B.-B. Li, Y. Luo, G.-W. Chu, H.-K. Zou, J.-F. Chen, Desulfurization intensification by ionic liquid in a rotating packed bed, Chem. Eng. Process. Intensif. 148 (2020) 107793.
[26] D.T. Silverman, C.M. Samanic, J.H. Lubin, A.E. Blair, P.A. Stewart, R. Vermeulen, J.B. Coble, N. Rothman, P.L. Schleiff, W.D. Travis, R.G. Ziegler, S. Wacholder, M.D. Attfield, The diesel exhaust in miners study: A nested case-control study of lung cancer and diesel exhaust, J. Natl. Cancer Inst. 104 (2012) 855–868. https://doi.org/10.1093/jnci/djs034.
[27] H.E. Hesketh, M.P. Cal, Air Pollution Control: traditional and hazardous pollutants, J. Environ. Qual. 26 (1997) 1442.
[28] A. Poullikkas, Review of Design, Operating, and financial considerations in flue gas desulfurization systems, Energy Technol. Policy. 2 (2015) 92–103.
[29] S.-K. Back, A.H.M. Mojammal, H.-H. Jo, J.-H. Kim, M.-J. Jeong, Y.-C. Seo, H.-T. Joung, S.-H. Kim, Increasing seawater alkalinity using fly ash to restore the pH and the effect of temperature on seawater flue gas desulfurization, J. Mater. Cycles Waste Manag. 21 (2019) 962–973.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50039-7

A Novel Process Synthesis of a Dehydrating Unit of Domestic Natural Gas Using TEG Contactor and TEG Regenerator

Renanto Renanto[a*], Sony Ardian Affandy[a], Adhi Kurniawan[a], Juwari Juwari[a] and Rendra Panca Anugraha[a]

[a]*Department of Chemical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia*
renanto@chem-eng.its.ac.id

Abstract

The use of natural gas as a source of energy is widely known. Before it is processed to be a sale gas, the water content in the natural gas should be reduced to 200 ppmv in low pressure level, as water is corrosive in the pipeline. This work was carried out after several configurations in the dehydration unit had been simulated to find the best one in terms of the total annual cost. (Affandy et al, 2020; Affandy et al., 2017). The dehydration unit considered consists of a TEG (triethylene glycol) contactor and TEG regenerator. The contactor is used for dehydrating natural gas, where the water is absorbed using TEG solution while in the regenerator the water is desorbed by using stripping gas to increase the purity of TEG so as to absorb water in the TEG contactor. A new configuration was proposed in this paper. The configuration consists of a packed column in the TEG contactor on one hand, and a coldfinger and the recycled flare gas in the TEG regenerator on the other hand. The results showed that the new configuration had a total annual cost of US$ 212.829 x 10^3 per year. This indicated that the total annual cost of this configuration was lower than that of the base case where there were no coldfinger and recycled flash gas used. The reduction of total annual cost from the base case was about 33.3 %.

Keywords: Dehydrating Unit; Natural Gas; Process Synthesis; TEG Contactor; TEG Regenerator.

1. Introduction

The use of natural gas as a source of energy source is widely known. It is naturally containing some impurities such as hydrogen sulphide, carbon dioxide, or nitrogen. It is also typically saturated with water moisture at the wellhead. The water moisture may cause problems like hydrate formation and also potentially leads to corrosion in the pipeline. The water moisture needs to be reduced to below certain level, for example to below 200 ppmv before sold as sale gas. There are few methods to reduce the water moisture from the natural gas, i.e. gas condensation using refrigerant, solid adsorption and using liquid absorption (Carrol, 2014).

One of the most widely solvent used in the liquid absorption water dehydration is the Triethylene Glycol (TEG). There are several published works to improve the performance of the TEG dehydration and regeneration system. Kong et al (2018) elaborated several available methods in the TEG regeneration system, covering from conventional regeneration to the use of several stripping agents such as nitrogen, portion of dehydrated

gas, as well as liquid volatile hydrocarbon (Drizo process). It also covers the use of water exhauster (Coldfinger) as well as the emerging technologies such as membrane and supersonic separation technologies. The use of water exhauster in the regeneration system has been subject of recent research as well. Rahimpour et al (2013) simulated the dehydration unit using the Coldfinger system in a domestic gas plant. They focused on studying the influence of temperature, pressure and flowrate of the Coldfinger unit to the performance of the dehydration system. They also proposed a new mathematical model to approach the Coldfinger unit. Romero et al (2019) proposed two equilibrium stages operated at different temperatures to model the Coldfinger unit. Affandy et al (2020) investigated the performance improvement through the use of flash gas as a stripping gas source to the regenerator. They revealed that there will be some improvements in the TEG purity (up to 98.8%-wt) and the water moisture in the treated gas. However, it will be limited due to the availability of the flash gas, as well as the water content in the flash gas itself. The Total Annual Cost (TAC) of the proposed configuration showed around 20% reduction of TAC compared to the base case.

This work focuses on the synthesis of the regeneration system involving the Coldfinger concept and reusing the flash gas as the stripping gas source. The performance of the unit should satisfy the water moisture content in the treated gas and should provide lower TAC compared to the base case.

2. Process Description

Process configuration in Figure 1 used as the base case in this work was based on the work developed by Affandy et al (2017). The wet natural gas (stream 3507) was brought into contact with a lean Triethylene Glycol (TEG) solution (stream 4) in a trayed TEG Contactor (C-1). The dehydrated gas (stream 3604) is expected to have water moisture content less than 200 ppmv.

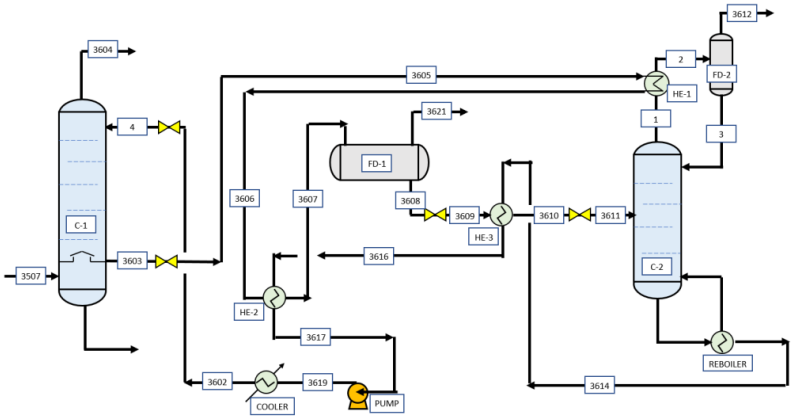


Figure 1 Process configuration of natural gas dehydration unit using TEG: base case

The glycol solution that has absorbed water (i.e. rich glycol solution, stream 3603) was then routed to a regeneration system in which the water content to be reduced to certain level so that it can be used again as lean glycol. The regeneration system consists of a flash drum (FD-1), a series of lean-to-rich heat exchanger (HE-1, HE-2, and HE-3) to conserve the heat, and a regenerator consisting of still column (C-2), condenser and

reboiler. The conventional regeneration typically uses a regenerator operating pressure of near atmospheric to maximize the water vaporization. The operating temperature of reboiler typically is limited to 204 ºC to minimize TEG glycol degradation. The regenerated lean glycol will have a water content of about 1.5 %-wt. The lean glycol (stream 3614) is then cooled to a temperature of around 45-48 ºC (stream 3602) through the lean-to-rich heat exchanger system and further in a glycol cooler, prior to entering the TEG Glycol Contactor.

3. Method

The process simulation model was built in ASPEN Plus V10. The Predictive Redlich-Kwong-Soave (PSRK) was used as the thermodynamic property package used in the simulation model from Affandy et al. (2017). The simulation model was also validated using the actual plant data taken from a domestic gas plant. The base case process configuration was subjected to area optimization of heat exchanger networks as defined in previous work from Affandy et al. (2017).

The proposed process configurations were evaluated to give the lowest Total Annual Cost (TAC). The TAC itself was determined using formulae from Luyben (2011) in which it is the sum of total energy cost and the total capital cost divided by payback period. In this work the payback period was taken as 3 (three) years. The total operating cost consists of steam cost, cooling water, and TEG makeup (Affandy, 2020)

$$TAC = TOC + \frac{TCC}{PB} \quad (1)$$

Where: TAC = Total Annual Cost ($ / year) TCC = Total Capital Cost ($)

TOC = Total Operating Cost ($ / year) PB = Payback period (year)

4. Proposed Configurations

4.1. General

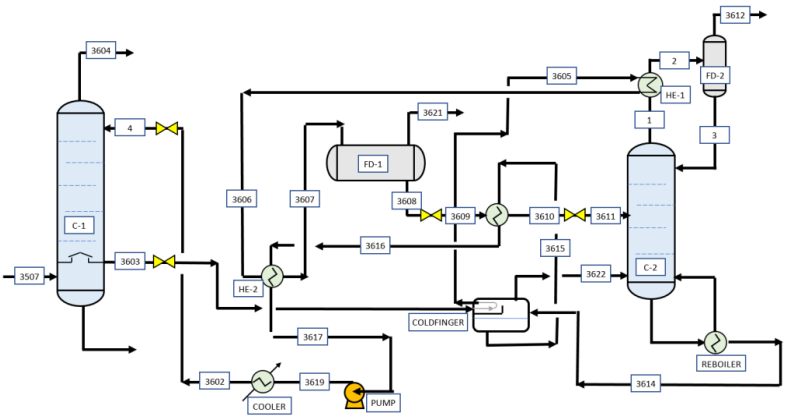


Figure 2 Process configuration of natural gas dehydration unit using TEG: regeneration package using Coldfinger system

The proposed process configurations in this work were based on the additional Coldfinger unit to the regeneration system as depicted in the Figure 2. In this scheme, the lean glycol outlet from reboiler (stream 3614) is routed to the Coldfinger unit that produces a leaner glycol (stream 3615). Furthermore as depicted in the Figure 3, the flash gas (stream 3621) from the Flash Drum was rerouted to the regenerator as the stripping agent (stream 5), following the scheme proposed by Affandy et al (2020).

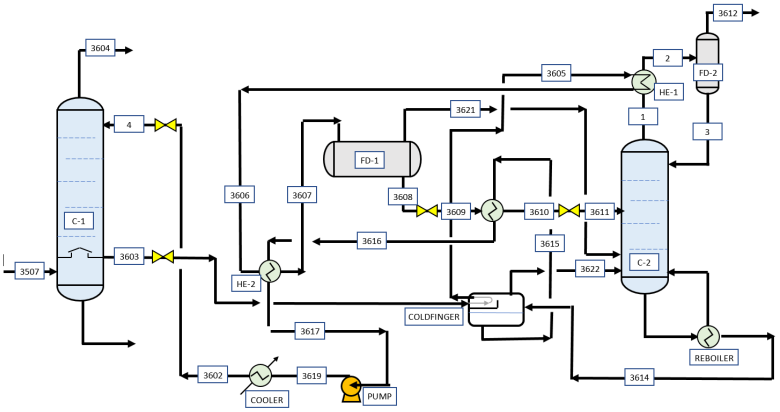


Figure 3 Process configuration of natural gas dehydration unit using TEG: Absorption column using trays column and regeneration package including Coldfinger system and flash gas as stripping gas

4.2. Coldfinger Model

The principle of the Coldfinger process is that the partly regenerated glycol from the distillation column is further concentrated in a two-phase tank with a cold spot (the Coldfinger) and a condensed collector in the gas phase. Figure 4 depicts the model of Coldfinger unit. The condensing of the water results in reduced water pressure in the tank and more water will evaporate from glycol. It shows the coldfinger simulation using two flash units (Affandy, 2020).

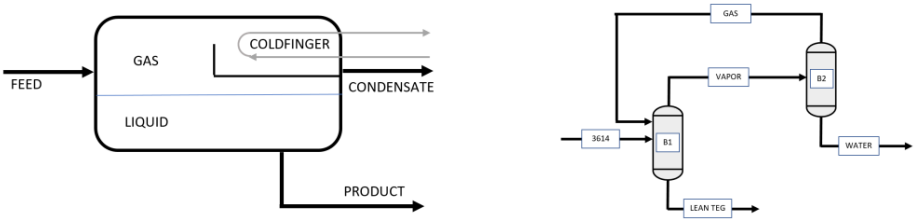


Figure 4 The Coldfinger model

Since the coldfinger unit is difficult to simulate in steady-state simulation program due to the unit is not in equilibrium, the condensing at the cold spot is at a lower temperature than the glycol, therefore coldfinger is simulated using two flash drums similar to the work of Erik and Tyvand (2002). The first flash drum for simulating the equilibrium

between the glycol and the vapor above. The vapor is sent to second flash tank simulating the condensing and removal of the water at a lower temperature.

5. Results and Discussion

Among all configurations, the base case simulation has moisture concentration close to maximum limitation (200 ppmv) due to the availability of data validation to plant data. The proposed configurations were set to achieve 180 ppmv moisture content. The purpose of setting the moisture concentration in dry gas stream below the base case result, is to show that the optimized base case and proposed configuration are better than the base case (Affandy, 2020).

Table 1 Performance comparison between the base case and the proposed configurations

Configuration	Lean TEG flowrate (kmole/h)	Lean TEG purity (%-mole)	Dry Gas Moisture (ppmv)	Solvent Loss (kg/h)
Base case	70.0	90.7	198	0.016
+ Coldfinger	16.5	96.1	181	0.013
+ Coldfinger + Flash gas as stripping	17.5	94.9	180	0.037

The proposed configurations using coldfinger system or flash gas as stripping gas have improved the solvent purity. This result will have big relation with the TEG flow rate that needed to obtain the water moisture concentration target in the dry gas stream. Since both proposed configurations gave higher solvent purity, therefore to achieve the moisture target, the system could be run at much lower solvent flowrate as presented in Table 1. The solvent loss based on the TEG make-up needed for every configuration also be studied here. The reason solvent loss occurs in the natural gas dehydration unit using TEG is the solvent carried over in dry gas increase as well as at TEG purification step. The coldfinger system and flash gas as stripping gas in the regeneration column is extremely high compared to another configuration. The combination between coldfinger system and flash gas as stripping gas likely has effect on the solvent loss (Affandy, 2020)

Table 2 Cost comparison between the base case and the proposed configurations

Configuration	TCC ($)	TCC / PB ($/year)	TOC ($/year)	TAC ($/year)
Base case	362,654	120,884	198,144	319,028
+ Coldfinger	366,301	122,100	87,306	209,406
+ Coldfinger + Flash gas as stripping	369,048	123,016	89,803	212,819

Table 2 shows that the addition of coldfinger system in the regeneration package affects to the total capital cost of the configuration that use coldfinger system. The main variables affecting the TOC reduction in the proposed configurations are TEG flow rate and TEG purity. Lower circulation rate has big impact on the total duty in the regeneration process. The proposed configuration to involve flash gas as stripping agent was found to give slightly higher reboiler loads compared to the configuration of regeneration with

Coldfinger only, hence more operating cost required. However, all proposed configurations have lower TAC compared to base case (Affandy, 2020).

6. Conclusions

This article described process synthesis of new configuration for natural gas dehydration using TEG. The configuration consists of a packed column in TEG Contactor and combined Coldfinger and rerouted vent gas as additional stripping gas in the regeneration section. The results showed that the new configuration is capable of providing similar water moisture content in the treated gas stream while minimizing the utilities requirements. This has led to about 33% reduction of the Total Annual Cost of the new configuration compared to the Base Case. It reduced the TAC to US$ 212,819 per year from US$ 319,028 per year.

Acknowledgements

The authors would like to express their gratitude for the financial aid given by the Ministry of Research and Technology/National Research and Innovation Agency Republic of Indonesia and Institut Teknologi Sepuluh Nopember.

References

Affandy, Sony A., Kurniawan, A., Handogo, R., Sutikno, J. P., & Chien, I. (2020). Technical and economic evaluation of triethylene glycol regeneration process using flash gas as stripping gas in a domestic natural gas dehydration unit. *Engineering Reports*, *August 2019*, 1–15.

Affandy, Sony Ardian. (2020). *DESIGN AND OPTIMIZATION OF NATURAL GAS DEHYDRATION UNIT USING TRIETHYLENE GLYCOL (TEG)*. Institut Teknologi Sepuluh Nopember Surabaya.

Affandy, Sony Ardian, Renanto, Juwari, & Chien, I. L. (2017). Simulation and optimization of structured packing replacement in absorption column of natural gas dehydration unit using triethylene glycol (TEG). *2017 6th International Symposium on Advanced Control of Industrial Processes, AdCONIP 2017*, 275–281.

Carrol, J. (2014). Dehydration of natural gas. In *Natural Gas Hydrates A Guide for Engineers* (pp. 175–195). Gulf Professional Publishing. https://doi.org/10.1016/b978-075067569-7/50007-2

Kong, Z. Y., Mahmoud, A., Liu, S., & Sunarso, J. (2018). Revamping existing glycol technologies in natural gas dehydration to improve the purity and absorption efficiency: Available methods and recent developments. *Journal of Natural Gas Science and Engineering*, *56*(June), 486–503. https://doi.org/10.1016/j.jngse.2018.06.008

Lars, Erik Øi, Tyvand, S. (2002). Process simulation of glycol regeneration. *GPA Europe's Meeting*.

Luyben, W. (2011). Economic Basis. In *Principles and Case Studies of Simultaneous Design* (pp. 1–317). John Wiley & Sons.

Rahimpour, M. R., Jokar, S. M., Feyzi, P., & Asghari, R. (2013). Investigating the performance of dehydration unit with Coldfinger technology in gas processing plant. *Journal of Natural Gas Science and Engineering*, *12*, 1–12. https://doi.org/10.1016/j.jngse.2013.01.001

Romero, I. A., Andreasen, A., Nielsen, R. P., & Maschietti, M. (2019). Modeling of the coldfinger water exhauster for advanced TEG regeneration in natural gas dehydration. *Chemical Engineering Transactions*, *74*(June), 661–666.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50040-3

A new trust-region approach for optimization of multi-period heat exchanger networks with detailed shell-and-tube heat exchanger designs

Saif R. Kazi[b], Ishanki A. De Mel[a], and Michael Short [a*]

[a] *Department of Chemical and Process Engineering, University of Surrey, Guildford GU2 7XH, UK*
[b]*Department of Chemical Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15232, USA*
m.short@surrey.ac.uk

Abstract

Multi-period Heat Exchanger Networks (HENs) are designed as heat recovery energy efficient systems over a set of operating conditions for process streams. The problem becomes more complex when detailed exchanger designs are accounted for in the network synthesis problem. Typically, in mixed-integer nonlinear programming (MINLP) multi-period HEN optimisation, the maximum area heat exchanger across all periods is considered. However, when considering detailed designs, often this exchanger is unsuitable for operation over all periods. In this study, a trust-region algorithm is proposed to incorporate detailed exchanger designs for multi-period operation. The exchanger design is modelled using surrogate models inside a network-level NLP model which is derived from the multi-period MINLP HENS model solution. The method is applied to a case study and the results show the effectiveness of the proposed algorithm.

Keywords: Process Synthesis, Heat Exchanger Network, Optimization

1. Introduction

HENs are common heat recovery systems in process industries that minimize utility costs. These systems exchange heat between hot process streams required to be cooled and cold process streams which need heat to reach higher desired temperatures. It is increasingly common for industrial plants to be operated over multiple, differing operating conditions, especially in batch processing. This requires HEN designs to be robust and feasible for different possible operating conditions, with this design process called multi-period HEN synthesis. Many studies have addressed this variation of the HEN problem, beginning with Aaltola (2002), which solved a simultaneous MINLP problem over different periods. Verheyen and Zhang (2006) solved the multi-period HEN using the largest exchanger area over the periods in the objective function. While most of these approaches focus on the trade-off between capital costs (exchanger area costs) and operating costs (heating and cooling utilities), the underlying detailed designs of the exchangers are not considered in the optimization. With varying conditions and streams with different thermophysical properties, it becomes more important to include the effect of exchanger design (number of shells, baffles, tubes etc.) in the HEN solution performance (Kang and Liu, 2019).

Mizutani et al. (2003) was the first to use Bell-Delaware based MINLP exchanger design models in HEN synthesis. They used integer variables for discrete decisions such as

number of baffles, fluid allocation, tube diameter etc. and nonlinear equations for heat transfer correlations and pressure drop. Due to the nonconvex MINLP nature of the exchanger design model, it is difficult to solve large HEN problems with many exchangers. Short et al. (2016a) developed a two-step hybrid strategy to incorporate the exchanger design based on the Bell-Delaware method, using correction factors in the HEN MINLP model. Short et al. (2016b) used a similar strategy to solve multi-period HEN problems, with feasible detailed exchanger designs over all operational periods, using manual heuristics to design the exchangers at each iteration.

Recently, Goncalves et al. (2019) have used linearization techniques and heuristics to solve the exchanger design model much more efficiently. Kazi et al. (2021a) proposed a discrete differential algebraic equation (DAE) model for detailed exchanger design which requires fewer assumptions than LMTD based methods. They also proposed a trust-region algorithm (Kazi et al., 2021b) to directly incorporate detailed exchanger designs into HEN synthesis, ensuring feasibility and optimality of the solution. In this work, we extend the formulation of Mahmood et al. (2021) to design shell and tube heat exchangers over multiple operating conditions using a discrete first principles model. We formulate a trust-region algorithm which embeds these discrete models inside a network-level HEN NLP model to solve for detailed exchanger designs along with optimal splitting ratios for process streams. To the extent of the authors' knowledge, this is the first study to incorporate multi-period exchanger design directly in HEN synthesis models.

2. Heat Exchanger Model

The heat exchanger design model uses a two-step algorithm, where the first step applies an enumeration-based approach to determine the discrete decision variables (tube diameter, number of baffles etc.). This model is similar to Mahmood et al. (2021) which uses an LMTD-based NLP formulation and smart enumeration to solve multi-period multi-shell heat exchanger designs. We expand this model to reformulate it to include the ability for different heat exchanger shells to have different geometry (tube diameter, tube length and number of baffles), in addition to including multiple tube pass arrangements. This enables the exchanger design to be more robust to varying stream.

The model also has additional degrees of freedom by introducing splitting variables on both tube and shell side for each operational period. This allows the streams to split and mix at the inlet and outlet of each shell respectively. The fraction of splitting provides additional degrees of freedom to find more feasible designs. For simplicity, the number of passes and tube allocation is kept the same across the multiple shells. Splitting and mixing constraints are as follows: For period j and shell k, volumetric flow rate (V_{jk}) is related to total inlet mass flow rate (m_j) using split ratio variable r_{jk}:

$$V_{jk} = (m_j/\rho_j).(1 - r_{jk}) \quad \forall j, 1 \le k \le N \tag{1}$$

Similarly, the mixing occurs at the exit of each shell using the following energy balance:

$$T_{jk}{}^{out}(1 - r_{jk}) + T_{jk}{}^{in} r_{jk} = T_{jk+1}{}^{in} \quad \forall j, 1 \le k \le N-1 \tag{2}$$

The duty for the shells, using the tube and shell side temperatures and split ratios, is:

$$Q = m_j.Cp_j.(1 - r_{jk}).(T_{jk}{}^{out} - T_{jk}{}^{in}) \quad \forall j, 1 \le k \le N \tag{3}$$

LMTD for each shell is approximated using a small positive parameter (ϵ) as shown:

$$LMTD \approx \Delta T_2.\sqrt{(((\Delta T_1/\Delta T_2 - 1)^2 + \epsilon)/(log(\Delta T_1/\Delta T_2)^2 + \epsilon))} \quad (4)$$

The other constraints in the model and the design equations are provided in the supplementary section of Kazi et al. (2021a), extended to multi-period.

2.1. DAE Model

The heat exchanger design based on the LMTD equation has certain assumptions such as constant physical properties and no phase change. Moreover, the LMTD approximation makes the Bell-Delaware model unsuitable for derivative based solvers. In the second step, a more accurate first principles DAE model which was proposed in Kazi et al. (2021a) is used which does not use the LMTD formula, F_t correction and its assumptions. This is particularly useful in situations where thermophysical fluid properties can change with temperature which is commonplace in batch processing.

The DAE model uses coupled ODEs with algebraic design equations to size the exchanger. The ODEs are discretized using the discrete geometric design variables (number of tube passes and baffles) into finite elements. The heat equation for both tube and shell side is discretized and solved and the size of the discretized elements are used to calculate overall exchanger design variables such as number of tubes, shell diameter etc. In the algorithm, values of the discrete variables are obtained from the solution of the first step using the LMTD method.

Similar to the first step, the multi-period DAE model gives for more flexible design by allowing for streams to split and bypass over exchanger shells. The discretized heat equations inside each element are written as:

$$C^h{}_{jk}\,({T_{jk}}^{i+1} - {T_{jk}}^{i})/2 + U_{jk}\Delta A_k({T_{jk}}^{i+1} - {t_{jk}}^{i+1})/3 + U_{jk}\,\Delta A_k({T_{jk}}^{i} - {t_{jk}}^{i})/6 = 0$$

$$C^c{}_{jk}\,({t_{jk}}^{i+1} - {t_{jk}}^{i})/2 - U_{jk}\Delta A_k({T_{jk}}^{i+1} - {t_{jk}}^{i+1})/3 - U_{jk}\,\Delta A_k({T_{jk}}^{i} - {t_{jk}}^{i})/6 = 0 \quad (5)$$

The complete DAE model is described in Kazi et al. (2021a) with details on discretization and solution strategy.

3. Trust Region Algorithm

The multi-period HEN model has inaccurate and insufficient design equations for each exchanger. To obtain accurate detailed designs for each exchanger, we use a trust-region algorithm, similar to Kazi et al. (2021b), that embeds detailed DAE models as black box functions using surrogate models. The network NLP model can be written as Eq.(6) :

$$\min f(x), s.t.\, g(x) \le 0, h(x) = 0, y = d(w) \quad (6)$$

where f, g, and h are twice differentiable functions and d is the black box DAE model, whose equations are unknown to the NLP model. y and w are input and output variables to the black box function, and the other variables in the NLP are denoted by z, such that $x^T = [w^T, y^T, z^T]$. The black box function is replaced by reduced model r(w) and a trust-region constraint is added to ensure that the reduced model is a "*good*" approximation of the original model within the trust-region. The trust region subproblem ($TRSP_k$) is then:

$$min\, f(x), s.t.\, g(x) \le 0, h(x) = 0, y = r_k(w), ||x - x_k|| \le \Delta_k \quad (7)$$

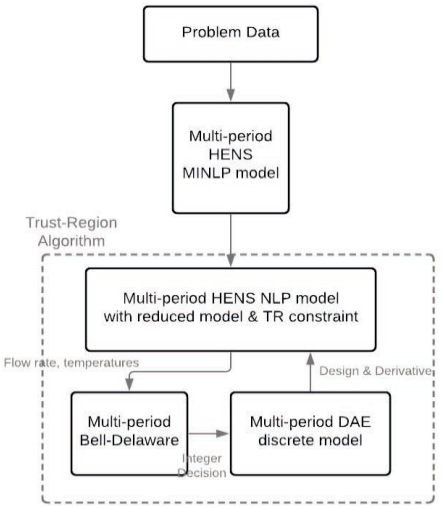


Figure 1: Algorithm illustration

To further simplify and reduce the size of the TRSP NLP problem, the variables can be partitioned into $x^T = [u^T, v^T]$, where u are the degrees of freedom and v are the rest of the variables. The reformulated $TRSP_k$ is written as Eq.(8) follows:

$$\min f(x) + \beta q, s.t. g(x) \leq q \in R^+,$$
$$h(x) = 0, y = r_k(w), \qquad (8)$$
$$||u - u_k|| \leq \Delta_k$$

where β is a penalty parameter and q is a relaxation variable. Eq.(8) ensures that $TRSP_k$ is always feasible and has the same solution as Eq.(7) for sufficiently large value of β.

3.1. Trust-region filter and update

The trust-region filter (TRF) compares the infeasibility (θ=||d(w)-r(w)||) and the objective value ($f(x_k)$) to update the trust-region radius (Δ_k) after each iteration of NLP solve. The reduced model (r_k) is also updated using the correction order formula as:

$$r_k(w) = s(w) + (d(w_k) - s(w_k)) + (\nabla d(w_k) - \nabla s(w_k))^T (w - w_k) \qquad (9)$$

where s is a simple surrogate model (in this case s(w) = 0) and $d(w_k), \nabla \mathbf{d}(\mathbf{w}_k)$ are the solution and sensitivity of the black box function model. The trust-region radius (Δ_k) is updated using the ratio test based on the decrease in infeasibility.

$$\rho = (1 - \theta(w_{k+1})/\theta(w_k), \qquad \Delta_{k+1} = \gamma \Delta_k \; if \; \rho \geq \eta_2, else \; \Delta_{k+1} = \Delta_k/\gamma \qquad (10)$$

4. Case Study

The proposed algorithm is tested with a multi-period HEN case study from Verheyen and Zhang (2006) with 3 hot streams, 4 cold streams, over 3 operational periods. Detailed stream parameters required for detailed design are obtained from Short et al. (2016b).

4.1. Results

The optimal network results are shown in Figure 2, with the detailed heat exchanger designs summarised in Table 1. The MINLP finds the optimal network topology to contain 7 heat exchanger matches and 5 utility exchangers, with an overall total annual cost (TAC) of $3,152,295 p.a. When detailed heat exchanger designs are incorporated via the NLP TRF algorithm, a TAC of $3,342,549 p.a. is found, when using the same objective function as used in Short et al. (2016b). This shows that without the consideration of the detailed heat exchanger designs that the costs may be significantly underestimated, particularly when considering multi-pass heat exchangers that may stray far from ideal counter-current flow and requiring more shells. Interestingly, when the tube-side velocities are allowed to be quite low (lower bound of 0.5 m/s), all the optimal

detailed heat exchangers can carry out the required duty across all periods of operation. However, if a lower bound of 1 m/s is enforced for the tubeside, there are certain periods that require an additional heat exchanger to perform the heat exchange.

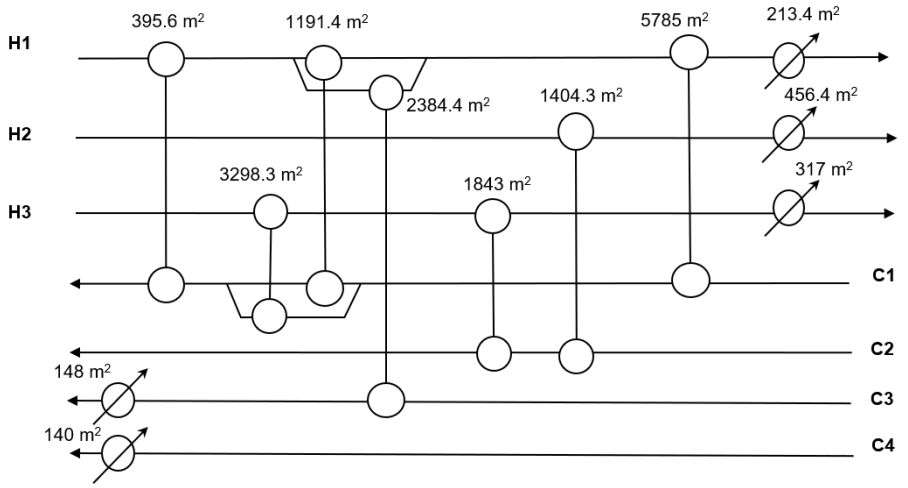


Figure 2: Solution obtained for illustrative Case Study from Verheyen and Zhang (2006)

Table 1: Summarized results for the detailed heat exchanger designs

Exchanger Assignment	Exchanger Area (m^2)	Number of shells	Split-ratio Tubeside (r_p^{tube}) for periods {1, 2, 3}	Split-ratio Shellside (r_p^{shell}) for periods {1, 2, 3}
[H1, C1, 1]	395.6	1	{0.055, 0.147, 0}	{0.263, 0.344, 0.03}
[H1, C1, 2]	1191.4	5	{0.02, 0.038, 0.09}	{0.0154, 0.0267, 0}
[H1, C1, 4]	5785	7	{0.014, 0.0124, 0.01}	{0.051, 0.038, 0.0225}
[H1, C3, 2]	2384.4	4	{0.083, 0.0347, 0}	{0.0334, 0.0203, 0.0083}
[H2, C2, 4]	1404.3	3	{0.023, 0.12, 0.035}	{0, 0, 0.063}
[H3, C1, 2]	3298.3	6	{0.02, 0.132, 0.02}	{0.023, 0.0132, 0.0201}
[H3, C2, 3]	1843.04	4	{0.02, 0.053, 0.073}	{0.003, 0.032, 0.034}

5. Conclusions

In this paper a new approach to the design of multi-period HENs is presented that incorporates detailed heat exchanger designs into the network optimisation via surrogate modelling and a trust-region filter algorithm. This is the first algorithm presented in literature that incorporates detailed shell-and-tube heat exchanger designs in the network synthesis problem for multi-period operation automatically. Detailed shell and tube heat exchangers are designed optimally for the input-output information provided from the NLP network optimisation via a novel hybrid multi-period heat exchanger design algorithm using DAEs to make for a general and rigorous process synthesis

framework. The results obtained show that it is important to consider the detailed heat exchanger designs during network synthesis, as the obtained design and optimal solution were shown to be very different, with the MINLP network synthesis using only the maximum area across the operational periods underestimating the overall costs significantly. By including more information regarding numbers of shells, non-counterflow behaviour in the exchangers, and the potential for stream bypassing and splitting, the new algorithm provides more realistic answers within an optimisation framework. The algorithm presented here may help in finding improved networks in dynamic multi-period industrial environments. The DAE formulation also can consider fluids with physical properties that can change within a heat exchanger, as are common in the food and fast-moving consumer goods industries. The current implementation does not allow for multiple network topologies to be systematically assessed, and is computationally costly, and hence future work will focus on improving these aspects.

References

J. Aaltola, 2002, Simultaneous synthesis of flexible heat exchanger networks, Applied Thermal Engineering, 22(8), 907-918.

Goncalves C.O., Costa A.H., Bagajewicz M.J., 2019. Linear method for the design of shell and tube heat exchangers using the Bell-Delaware method. AIChE Journal, 65, e16602.

Kang L., Liu Y., 2019. Synthesis of flexible heat exchanger networks: A review. Chinese Journal of Chemical Engineering, 27 (7), 1485-1497.

Kazi S., Short M., Biegler L., 2021a. Heat exchanger network synthesis with detailed exchanger designs: Part 1. A discretized differential algebraic equation model for shell and tube heat exchanger design. AIChE Journal, 67(1), e17056.

Kazi S., Short M., Biegler L., 2021b. A trust region framework for heat exchanger network synthesis with detailed individual heat exchanger designs. Computers and Chemical Engineering, 153, 107447.

Mahmood Z., De Mel I.A., Kazi, Z., Isafiade, A.J., Short M., 2021. An optimisation algorithm for detailed shell-and-tube heat exchanger designs for multi-period operation. Chemical Engineering Transactions, 88(1), to appear.

Mizutani F.T., Pessoa F.L.P., Queiroz E.M., Hauan S., Grossmann I.E., 2003. Mathematical programming model for heat exchanger network synthesis including detailed heat exchanger designs. 1. Shell and tube heat exchanger design. Industrial & Engineering Chemistry Research, 42, 4009-4018.

Short M., Isafiade A., Fraser D.M., Kravanja Z., 2016a. Two-step hybrid approach for the synthesis of multi-period heat exchanger networks with detailed exchanger design. Applied Thermal Engineering, 105, 807-821.

Short M., Isafiade A., Fraser D.M., Kravanja Z., 2016b. Synthesis of heat exchanger networks using mathematical programming and heuristics in a two-step optimisation procedure with detailed exchanger design. Chemical Engineering Science, 144, 372-385.

Verheyen W., Zhang N., 2006. Design of flexible heat exchanger network for multi-period operation. Chemical Engineering Science, 61(23), 7730-7753.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50041-5

A mathematical technique for utility exchanger network synthesis and total site heat integration

Jui-Yuan Lee[a,b*], Wilasinee Seesongkram[c]

[a] *Department of Chemical Engineering and Biotechnology, National Taipei University of Technology, 1, Sec 3, Zhongxiao E Rd, Taipei 10608, Taiwan, ROC*
[b]*Research Center of Energy Conservation for New Generation of Residential, Commercial, and Industrial Sectors, National Taipei University of Technology, 1, Sec 3, Zhongxiao E Rd, Taipei 10608, Taiwan, ROC*
[c]*Chemical Engineering Practice School, King Mongkut's University of Technology Thonburi, 126 Prachautid Road, Bangmod, Thoongkru, Bangkok 10140, Thailand*
juiyuan@ntut.edu.tw

Abstract

Fossil energy has been increasingly consumed since the second Industrial Revolution, with economic growth and modernisation. This has led to environmental issues such as resource depletion, pollution and climate change. To enhance energy efficiency as a measure to mitigate climate change, heat exchanger networks (HENs) for heat recovery are widely used in various industrial applications. This work develops superstructure-based mathematical models for direct and indirect HEN synthesis for interplant heat integration. The mixed-integer nonlinear programming model minimises the total annualised cost for utility exchanger network synthesis and total site heat integration, and the results are compared with those from the conventional and unified total site targeting methods. An industrial case study is presented to demonstrate the application of the proposed approach.

Keywords: energy efficiency; heat recovery; mathematical programming; stage-wise superstructure.

1. Introduction

Due to population expansion and the Industrial Revolution as well as rapid economic growth, energy consumption in the world has significantly increased. From 2015 to 2040, world energy consumption is expected to increase by about 28% (Rodriguez, 2018). However, energy resources are limited. Therefore, heat exchanger networks (HENs), as an important heat recovery system, plays a significant role in processing plants. Synthesising HENs in chemical processes allows energy efficiency to improve. HEN synthesis has gained a lot of attention in the process industry. There have been numerous studies on HENs (Klemeš and Kravanja, 2013).

HEN research has focused mostly on a single process because of the inability to recover all the waste heat within chemical plants. Hence, considerable amounts of heat and energy are wasted. Subsequently, some researchers were interested in extending the work for a single plant to multi-plant HEN synthesis, which is also known as interplant heat integration. Both the process streams (direct integration) and the intermediate streams (indirect integration) are involved in multi-plant HEN synthesis. The direct method provides more potential for energy saving because heat is transferred directly

between process streams. However, this method can entail a higher capital cost than the indirect method. This is due to the large requirement of piping and pumping. Although the indirect method can result in a lower capital cost, the achievable energy saving is reduced because of the use of intermediate fluid. Chang et al. (2019) found that considering both the direct and indirect methods could reduce the energy consumption by 3.2% when compared to using only the direct method. To maximise the benefits from multi-plants HENs synthesis, both the direct and indirect methods should be considered in the optimisation. Tarighaleslami et al. (2018) presented a unified total site integration method for HEN synthesis and utility exchange network (UEN) design. This method only allows utility heat exchangers in series within the same process.

In this work, HEN synthesis is carried out in the multi-plant/process context. Both the direct and indirect methods are used to minimise the total annualised cost (TAC) of the plant(s). A mathematical programming model is developed as an alternative to pinch-based techniques and applied to an industrial case study, where the hot-side and cold-side temperatures of the intermediate fluid (hot water) are treated as optimisation variables. The results obtained are then compared with the results reported in the previous works using pinch analysis.

2. Problem statement

- There are a set of hot process streams $i \in \mathbf{IP}_p$ and a set of cold process streams $j \in \mathbf{JP}_p$ in a set of processes $p \in \mathbf{P}$. The supply and target temperatures of the process streams and their heat capacity flowrates are known parameters.
- A set of hot utilities $hu \in \mathbf{HU}$ and a set of cold utilities $cu \in \mathbf{CU}$ are available for heating and cooling demands that cannot be met by heat recovery. Additionally, intermediate streams (e.g. hot water) that act as both a hot stream ($i \in \mathbf{IM}_p$) and a cold stream ($j \in \mathbf{JM}_p$) are also used in the processes. The supply and target temperatures on both hot and cold sides of the intermediate streams and their heat capacity flowrates are to be determined or optimised.
- The objective is to synthesise an optimal HEN, which consists of a heat recovery network and a UEN, for the minimum total annualised cost (TAC).

In this HEN synthesis problem, direct heat integration of process streams is allowed in individual processes. Inter-process heat integration is carried out indirectly through intermediate streams. It is assumed that intermediate streams are split for the processes that require heating/cooling, and mixed isothermally. Also, the supply temperature of the hot side of an intermediate stream equals its cold-side target temperature, whilst its hot-side target temperature equals its cold-side supply temperature.

3. Model formulation

Figure 1 shows a modified stage-wise superstructure, based on which the mathematical model for multi-process HEN synthesis involving heat recovery and utility exchange is formulated. This superstructure considers heat exchange matches between process and intermediate streams in all stages. However, the match between intermediate streams is excluded because intermediate streams are used for the heating and cooling of process streams. The hot side of an intermediate stream loop may use cold utilities to achieve its target temperature, whilst the cold side achieves its target temperature by heat exchange with hot process streams, without using hot utilities.

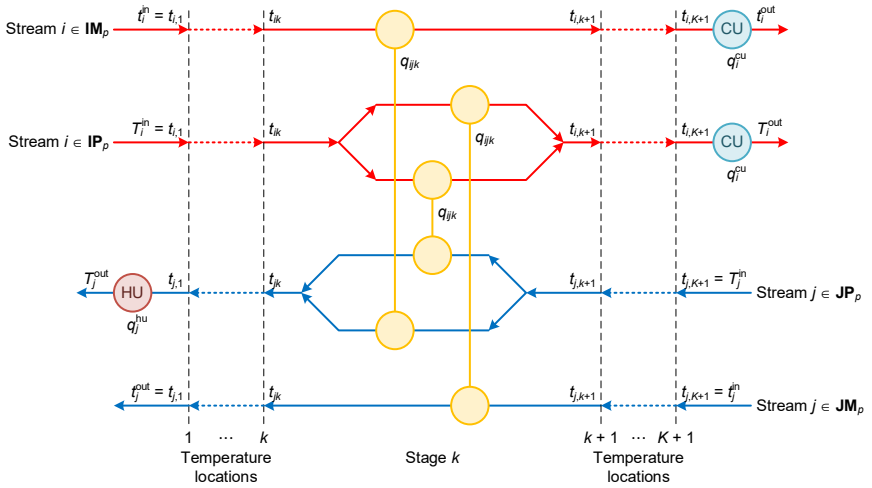


Figure 1. Stage-wise Superstructure for a HEN Involving Utility Exchange

The formulation of the HEN model is as follows. Eqs. (1)-(4) describe the overall heat balances for process and intermediate streams

$$F_i\left(T_i^{\text{in}} - T_i^{\text{out}}\right) = \sum_{j\in\mathbf{J}} \sum_{k\in\mathbf{ST}} q_{ijk} + \sum_{cu\in\mathbf{CU}} q_{i,cu} \quad \forall i \in \mathbf{IP}_p, p \in \mathbf{P} \tag{1}$$

$$F_j\left(T_j^{\text{out}} - T_j^{\text{in}}\right) = \sum_{i\in\mathbf{I}} \sum_{k\in\mathbf{ST}} q_{ijk} + \sum_{hu\in\mathbf{HU}} q_{hu,j} \quad \forall j \in \mathbf{JP}_p, p \in \mathbf{P} \tag{2}$$

$$f_i\left(t_i^{\text{in}} - t_i^{\text{out}}\right) = \sum_{j\in\mathbf{JP}_p} \sum_{k\in\mathbf{ST}} q_{ijk} + \sum_{cu\in\mathbf{CU}} q_{i,cu} \quad \forall i \in \mathbf{IM}_p, p \in \mathbf{P} \tag{3}$$

$$f_j\left(t_j^{\text{out}} - t_j^{\text{in}}\right) = \sum_{i\in\mathbf{IP}_p} \sum_{k\in\mathbf{ST}} q_{ijk} \quad \forall j \in \mathbf{JM}_p, p \in \mathbf{P} \tag{4}$$

Eqs. (5)-(8) describe the heat balances in each stage.

$$F_i\left(t_{ik} - t_{i,k+1}\right) = \sum_{j\in\mathbf{J}} q_{ijk} \quad \forall i \in \mathbf{IP}_p, p \in \mathbf{P}, k \in \mathbf{ST} \tag{5}$$

$$F_j\left(t_{jk} - t_{j,k+1}\right) = \sum_{i\in\mathbf{I}} q_{ijk} \quad \forall j \in \mathbf{JP}_p, p \in \mathbf{P}, k \in \mathbf{ST} \tag{6}$$

$$f_i\left(t_{ik} - t_{i,k+1}\right) = \sum_{j\in\mathbf{JP}_p} \sum_{p\in\mathbf{P}} q_{ijk} \quad \forall i \in \mathbf{IM}_p, p \in \mathbf{P}, k \in \mathbf{ST} \tag{7}$$

$$f_j\left(t_{jk} - t_{j,k+1}\right) = \sum_{i\in\mathbf{IP}_p} \sum_{p\in\mathbf{P}} q_{ijk} \quad \forall j \in \mathbf{JM}_p, p \in \mathbf{P}, k \in \mathbf{ST} \tag{8}$$

Temperature assignments are omitted due to space limitations. Temperature feasibility constraints are given in Eqs. (9)-(12).

$$t_{ik} \geq t_{i,k+1} \geq T_i^{\text{out}} \quad \forall i \in \mathbf{IP}_p, k \in \mathbf{ST} \tag{9}$$

$$T_j^{\text{out}} \geq t_{jk} \geq t_{j,k+1} \quad \forall j \in \mathbf{JP}_p, k \in \mathbf{ST} \tag{10}$$

$$t_{ik} \geq t_{i,k+1} \geq t_i^{\text{out}} \quad \forall i \in \mathbf{IM}_p, k \in \mathbf{ST} \tag{11}$$

$$t_j^{\text{out}} \geq t_{jk} \geq t_{j,k+1} \quad \forall j \in \mathbf{JM}_p, k \in \mathbf{ST} \tag{12}$$

Heat loads for hot and cold utilities are given by Eqs. (13)-(15).

$$F_i(t_{ik} - T_i^{\text{out}}) = \sum_{cu \in \mathbf{CU}} q_{i,cu} \quad \forall i \in \mathbf{IP}_p, p \in \mathbf{P}, k = K+1 \tag{13}$$

$$F_j\left(T_j^{\text{out}} - t_{jk}\right) = \sum_{hu \in \mathbf{HU}} q_{hu,j} \quad \forall j \in \mathbf{JP}_p, p \in \mathbf{P}, k = 1 \tag{14}$$

$$f_i(t_{ik} - T_i^{\text{out}}) = \sum_{cu \in \mathbf{CU}} q_{i,cu} \quad \forall i \in \mathbf{IM}_p, p \in \mathbf{P}, k = K+1 \tag{15}$$

Logical constraints are given in Eqs. (16)-(18).

$$Q_{ij}^{\text{L}} z_{ijk} \leq q_{ijk} \leq Q_{ij}^{\text{U}} z_{ijk} \quad \forall i \in \mathbf{I}, j \in \mathbf{J}, k \in \mathbf{ST} \tag{16}$$

$$Q_i^{\text{L}} z_{i,cu} \leq q_{i,cu} \leq Q_i^{\text{U}} z_{i,cu} \quad \forall i \in \mathbf{I}, cu \in \mathbf{CU} \tag{17}$$

$$Q_j^{\text{L}} z_{hu,j} \leq q_{hu,j} \leq Q_j^{\text{U}} z_{hu,j} \quad \forall j \in \mathbf{JP}_p, hu \in \mathbf{HU} \tag{18}$$

Eq. (19) excludes the matches between intermediate streams. Eq. (20) then excludes the matches between hot and cold streams of different processes.

$$z_{ijk} = 0 \quad \forall i \in \mathbf{IM}_p, j \in \mathbf{JM}_p, p \in \mathbf{P}, k \in \mathbf{ST} \tag{19}$$

$$z_{ijk} = 0 \quad \forall i \in \mathbf{IP}_p \cup \mathbf{IM}_p, p \in \mathbf{P}, j \in \mathbf{JP}_{p'} \cup \mathbf{JM}_{p'}, p' \in \mathbf{P} \wedge p' \neq p, k \in \mathbf{ST} \tag{20}$$

Temperature difference constraints are given in Eqs. (21)-(27).

$$\Delta T^{\min} - \Gamma\left(1 - z_{ijk}\right) \leq t_{ik} - t_{jk} \quad \forall i \in \mathbf{I}, j \in \mathbf{J}, k \in \mathbf{ST} \tag{21}$$

$$\Delta T^{\min} - \Gamma\left(1 - z_{ijk}\right) \leq t_{i,k+1} - t_{j,k+1} \quad \forall i \in \mathbf{I}, j \in \mathbf{J}, k \in \mathbf{ST} \tag{22}$$

$$\Delta T^{\min} - \Gamma\left(1 - z_{i,cu}\right) \leq t_{ik} - T_{cu}^{\text{out}} \quad \forall i \in \mathbf{I}, cu \in \mathbf{CU}, k = K+1 \tag{23}$$

$$\Delta T^{\min} - \Gamma\left(1 - z_{i,cu}\right) \leq T_i^{\text{out}} - T_{cu}^{\text{in}} \quad \forall i \in \mathbf{IP}_p, p \in \mathbf{P}, cu \in \mathbf{CU}, k = K+1 \tag{24}$$

$$\Delta T^{\min} - \Gamma\left(1 - z_{i,cu}\right) \leq t_i^{\text{out}} - T_{cu}^{\text{in}} \quad \forall i \in \mathbf{IM}_p, p \in \mathbf{P}, cu \in \mathbf{CU}, k = K+1 \tag{25}$$

$$\Delta T^{\min} - \Gamma\left(1 - z_{hu,j}\right) \leq T_{hu}^{\text{out}} - t_{jk} \quad \forall j \in \mathbf{JP}_p, p \in \mathbf{P}, hu \in \mathbf{HU}, k = 1 \tag{26}$$

$$\Delta T^{\min} - \Gamma\left(1 - z_{hu,j}\right) \leq T_{hu}^{\text{in}} - T_j^{\text{out}} \quad \forall j \in \mathbf{JP}_p, p \in \mathbf{P}, hu \in \mathbf{HU}, k = 1 \tag{27}$$

The objective function is to minimise the TAC, which consists of the annual operating cost and the annualised capital cost, as given in Eq. (28).

$$\min f_{\text{TAC}} = f_{\text{AOC}} + f_{\text{ACC}} \tag{28}$$

Due to space limitations, detailed operating and capital cost functions are omitted. The complete model is a mixed integer nonlinear programme (MINLP).

In the next section, an industrial case study is presented to demonstrate the proposed HEN model. The MINLP model is solved in GAMS using BARON.

4. Case study

This case study considers a large kraft pulp mill plant, which contains 10 processes and 64 process streams (Bood and Nilsson, 2013). The model for this case study involves more than 10,000 constraints and variables, depending on the number of stages.

In the optimal UEN, high-pressure steam is used for four cold streams in three processes (bleaching, digestion and recovery boiler), with a total requirement of 27,741 kW. This is identical to the target of Tarighaleslami et al. (2018). In addition, low-pressure steam is used for 18 cold streams in nine processes (bleaching, causticizing, digestion, evaporator, district heating, miscellaneous, paper room, stripper and recovery boiler), with a total requirement of 131,128 kW. This is 6.65% less than that of Tarighaleslami et al. (2018). The total cooling water requirement was determined to be 79,309 kW, which is 11.56% more than that of Tarighaleslami et al. (2018).

For indirect inter-process heat integration, low-temperature and high-temperature hot water are used as intermediate streams. The optimised low-temperature hot water loop has a cold temperature of 29.19°C, a hot temperature of 59.96°C and a heat capacity flowrate of 1,051.87 kW/°C, as shown in Figure 2. On the other hand, the optimised high-temperature hot water loop has a cold temperature of 60°C, a hot temperature of 93°C and a heat capacity flowrate of 271 kW/°C, as shown in Figure 3.

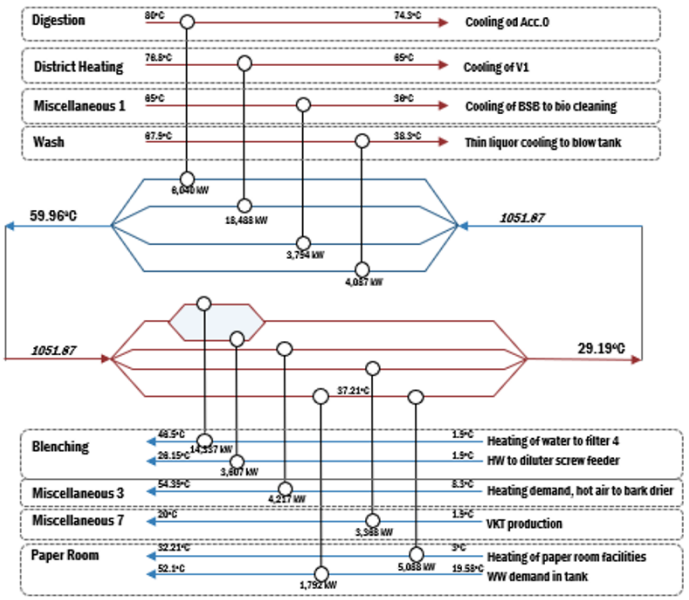


Figure 2. Low-temperature Hot Water Loop in the optimal UEN

Compared to the results of Tarighaleslami et al. (2018), the solution obtained in this work has a lower utility cost ($42,966,983/y versus $44,834,691/y) and a higher capital cost ($2,109,177/y versus $1,856,486/y). This is due to increased heat recovery in total site integration. Consequently, the TAC is reduced by $1,615,017 (3.46%). In addition, compared to the sequential unified total site targeting method, the proposed HEN model considers the trade-off between utility and capital costs, and allows better solutions in terms of TAC to be found through a simultaneous optimisation approach.

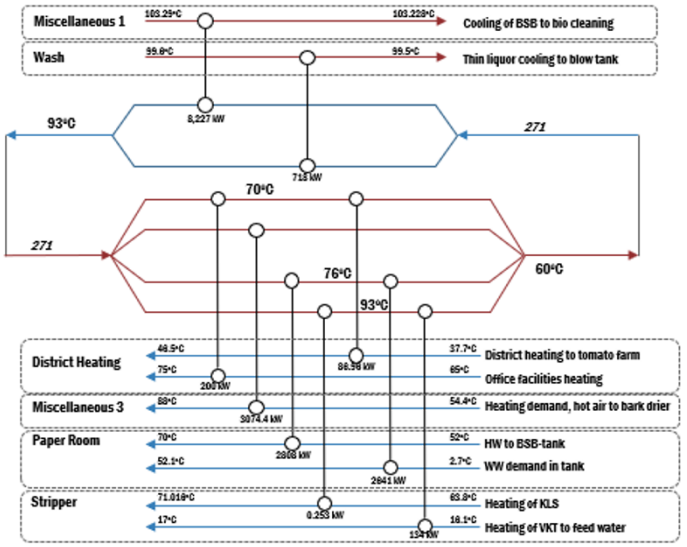


Figure 3. High-temperature Hot Water Loop in the optimal UEN

5. Conclusions

A multi-plant/process heat integration scheme with a HEN model has been presented in this paper. The model comprehensively considers the interactions between utility use, heat recovery and heat transfer area in minimizing the TAC. This model can be used to determine the required heat capacity flowrates of intermediate streams, their supply and target temperatures, and the HEN structure simultaneously. An industrial case study was solved to illustrate the applicability and effectiveness of the proposed model. Compared to sequential targeting approaches, simultaneous optimisation using the proposed model has the capability of finding the minimum-TAC solution. In future work, the objective function will be modified to include further details such as piping requirements for inter-plant/process matches in design.

References

J. Bood, L. Nilsson, 2013, Energy Analysis of Hemicellulose Extraction at a Softwood Kraft Pulp Mill, Case Study of Södra Cell Värö, MSc Thesis, Chalmers University of Technology, Gothenburg, Sweden

C. Chang, Y. Wang, X. Feng, 2019, Optimal Synthesis of Multi-plant Heat Exchanger Networks Considering both Direct and Indirect Methods, Chinese Journal Chemical Engineering, 28, 2, 456-465

J.J. Klemeš, Z. Kravanja, 2013, Forty Years of Heat Integration: Pinch Analysis (PA) and Mathematical Programming (MP). Curr Opin Chem Eng 2, 4, 461-474

J.A. Rodriguez, 2018, EIA Report Projects Fossil and Nuclear Fuels will Provide 83 Percent of Total Global Energy in 2040, Pratt's Energy Law Report, 18, 3, 97-103

A.H. Tarighaleslami, T.G. Walmsley, M.J. Atkins, M.R.W. Walmsley, J.R. Neale, 2018, Utility Exchanger Network Synthesis for Total Site Heat Integration, Energy, 153, 1000-1015

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50042-7

Synthesis and Assessment of NOx to Ammonia Conversion Process in Combined Cycle Power Generation Systems

Hideyuki Matsumoto[a*], Kanako Kurahashi[b] , Haruna Tachikawa[a], Takaya Iseki[b]

[a] *Department of Chemical Science Engineering, Tokyo Institute of Technology, Tokyo 152-8550, JAPAN*
[b] *Fundamental Technology Research Institute, Tokyo Gas Co., Ltd., Yokohama 230-0045, JAPAN*
**Corresponding Author's E-mail: matsumoto.h.ae@m.titech.ac.jp*

Abstract

A purpose of the present paper is to demonstrate feasibility of two NOx to ammonia (NTA) processes in the combined cycle power generation systems by using process simulation. In application of NTA process that is single-stage system for conversion of NO to NH_3 available in existence of oxygen, an effect of NTA process operating temperature on power generation efficiency of two class of gas turbines (1300 °C, 1700 °C) was estimated. Possible configurations of the exhaust gas aftertreatment incorporating the NTA system were also proposed and clarified.

Keywords: Process Synthesis; Nitrogen Cycle; Thermal Power Plant; Heat Exchanger Network.

1. Introduction

Exhaust gas, wastewater, and residues generated from industries and living activities contain harmful nitrogen compounds such as NOx, organic nitrogen, ammonia nitrogen [NH_4^+, NH_3, etc.], NO_3^-, etc. It has been reported the amount of reactive nitrogen (NOx) discharged by combustion of fossil fuel and biomass alone accounts for 20% of the total amount discharged from the human systems (Galloway *et al*., 2008). So far, various selective catalytic reduction (SCR) methods have been reported to remove NOx in combustion exhaust gas. For example, in the urea SCR system, ammonia is utilized as a reductant. Previous study for the SCR system using hydrocarbon (HC-SCR systems) has reported that ammonia formed in the middle of SCR could accelerate the SCR process.

Instead of converting NOx into compound that has no economic value like N_2, possibilities to convert it into valuable product of NH_3 using similar principle as SCR process exists. Recently, development of NOx to ammonia conversion process (NTA process) has been actively promoted, since it is expected that reuse of the produced ammonia as a fuel and a denitration agent will bring about reduction of CO_2 emissions. It is estimated that approximately 250 million ton of ammonia could be produced by using a half of NOx (4%) in exhaust gas from all the thermal power plants located in Japan.

Hence, we consider that the NTA process could achieve both reduction of nitrogen compound emissions and reduction of greenhouse gas emissions, that is, "Cool & Clean

Earth". However, a method of the NTA process incorporation into the plant system and its effect have not been sufficiently investigated. The introduction of the NTA process not only reduces the amount of nitrogen compounds emitted into the environment to zero, but also can be expected to reduce the amount of denitration agent supply, the cost of denitration equipment, and the amount of energy supply to the overall plant system. A purpose of the present paper is to analyse and demonstrate the incorporation of NTA process to combined cycle gas turbine (CCGT) as part of exhaust gas aftertreatment. Process simulation was employed to predict the possible optimum process efficiency, as well as reduction of nitrogen compound emission.

2. Analysis for introducing two-stage NTA process system to CCGT system

A steady-state process simulator for CCGT (Figure 1) was developed by using the free process simulation environment COCO (CAPE-OPEN to CAPE-OPEN: https://www.cocosimulator.org/). The model of the CCGT system consists of a model of a high-pressure steam turbine and a model of a medium-pressure steam turbine. The process simulation of feeding natural gas consisting of CH_4 (about 90%), C_2H_6, C_3H_8, and C_4H_{10} to a gas turbine at about 56 t/h revealed that about 370 MW of energy was recovered in (1) to (4) in Figure 1. The power generation efficiency was estimated to be about 54 %. In this paper, we investigated the introduction methods of the following two types of NTA processes (i) and (ii).

- NTA process (i): Two-stage process system that consists of "adsorption/ concentration of NO" and "conversion of NO to NH_3 that is available in absence of oxygen"
- NTA process (ii): Single-stage process system for conversion of NO to NH_3 that is available in existence of oxygen

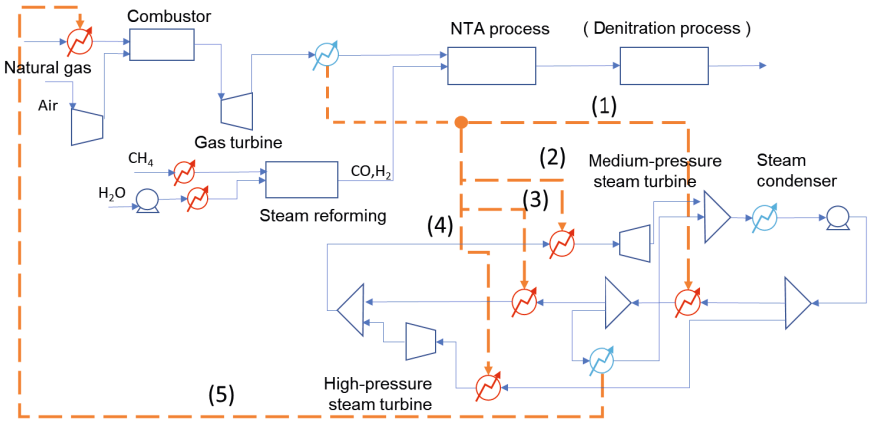


Figure 1 Example of process simulation of combined cycle power generation system.

First, influence of introducing the NTA process (i) in the CCGT system was analysed by the above-mentioned steady-state process simulator. In this simulation, the NTA process model based on the above-mentioned reaction of R1 and R2 was applied.

R1： $2NO+5CO+3H_2O \rightarrow 2NH_3+5CO_2$

R2： $2NO+5H_2 \rightarrow 2NH_3+2H_2O$

In reactions R1 and R2, CO and H_2 act as reductant of NO to produce NH_3. Based on the previous literature data (Kobayashi *et al*., 2019), a correlation equation between the reaction temperature and the NO conversion was estimated (Figure 2), which was applied to the NTA reactor model. In addition, in case of incomplete conversion of NO by the NTA reaction, it was assumed that the generated NH_3 was used for denitration of the remained NO in subsequent conventional SCR process.

The CO and H_2 for these NTA reactions were assumed to be supplied from the process of steam reforming of methane that was utilized from fuel gas to the gas turbine. In this process simulation, the amount of fuel gas used in the CH_4 steam reforming was determined based on the amount of reductant that was stoichiometrically required for conversion of NO in the combustion exhaust gas. In calculation for the steam reforming process, the reaction temperature was 1000 °C, the pressure was 1.4 MPa, and the molar ratio of CH_4 to H_2O was 1.

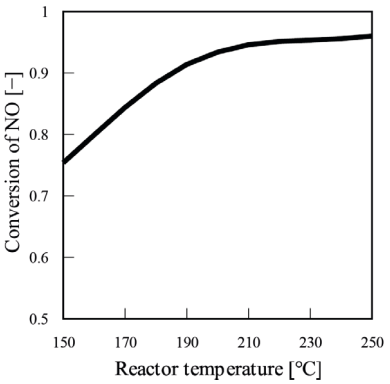
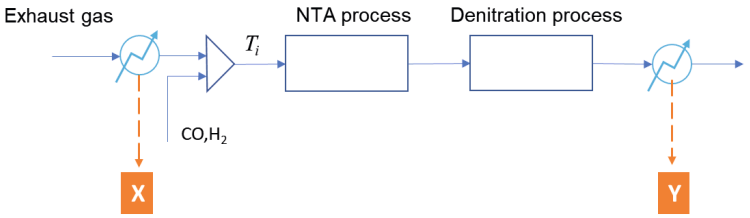


Figure 2 An example of calculation results based on the NTA reactor model.

Simulations for five cases (A－E) were performed as shown in Figure 3 to analyse influence combination of the heat exchanger (1) – (4) (Figure 1) in the heat recovery steam generator (HRSG) and the NTA process. In the cases of A and B, the temperature T_i before the NTA process was too low to commence the NTA reaction. For the case of C, conversion of NO to NH_3 by the NTA reaction was not complete. Subsequently, the SCR post-treatment was required, which consumed a portion of the generated NH_3 and decreased its overall yield. For the cases of D and E, it was seen that the outflow of NO from the NTA process was small. In particular, for case of E, about 1.9 t/h of ammonia, which was the maximum yield, was estimated when the NO concentration in the exhaust gas was about 130 ppm.

In the above simulation analysis, the amount of energy required for the CH_4 steam reforming to produce the reductant (CO, H_2) was estimated to be about 30 MW. Considering that the amount of energy recovered by the high-pressure steam turbine was about 25 MW, it was found that the energy consumption for production of the reductant significantly decreased in the power generation efficiency of the entire CCGT system. In addition, the use of natural gas reforming to produce CO-H_2 reductants offset the benefits of CO_2 reduction from the NTA system. Furthermore, for minimizing the energy consumption of the entire system, it is necessary to optimize position of the installed NTA reactor in the HRSG.



Case	X	Y
A	1,2,3,4	—
B	1,2,3	4
C	1,4	2,3
D	2,4	1,4
E	4	1,2,3

Figure 3 Five cases for combination of heat exchanger and NTA process in heat recovery steam generator (HRSG). Numbers in the table denote the heat exchange combination shown in Figure 1.

3. Design of CCGT system with single-stage NTA process system

We investigated a method for introducing the NTA process (ii) to the CCGT system by using the process simulator. It is considered that the CO_2 emissions derived from power generation can be significantly reduced by raising the combustion temperature of the gas turbine, which was attributed to improvement of the power generation efficiency.

Thus, we analysed influence of combustion temperature for the gas turbine to preferable position of the installed NTA process and the power generation efficiency of the entire CCGT system, by using a simulation system that included the pinch analysis (Figure 4). In the simulation system, the structure of the heat exchanger network in the HSRG that could perform the maximum recovery of heat of steam was derived by pinch analysis. And the amount of power generated by the entire CCGT system was estimated by using process simulation based on the derived heat exchanger network models.

In the present paper, the operating temperature for the NTA reactor was set at 300 °C or higher, by referring to information of the catalyst developed by Prof. Iwamoto research group in Waseda University. Assuming the operating temperature of the conventional ammonia selective catalytic reduction (NH_3-SCR) process in the CCGT system was 350 °C, it was considered that a part of the NH_3-SCR equipment in the HRSG could be replaced by the NTA process. In introduction of the developed NTA process, it is expected that its performance will facilitate revamp of the target CCGT system. Thus, we investigated influence of the operating temperature of the NTA process to energy consumption and power generation efficiency of the entire system. It was also assumed that the outlet temperature of the NTA reactor was different from the operating temperature of NH_3-SCR equipment.

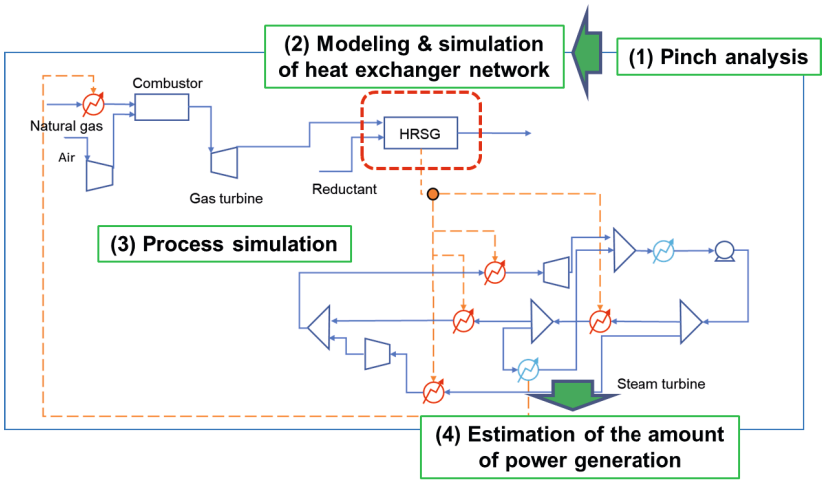


Figure 4 A framework of simulation system for design of CCGT system with one-stage NTA process system.

In a case when the operating temperature of the NTA process was set at 330 °C, which was reported to attain a relatively high NH_3 yield (however, conversion was less than 50%), it was assumed that NH_3-SCR process was placed after the NTA process. Thus, we came up with the two kinds of subprocess (Figure 5) as a means of raising the temperature of outflow from the NTA reactor to 350 °C.

- Subprocess 1: An afterburner is inserted at the midpoint between NTA reactor and NH_3-SCR equipment, and the gas temperature is adjusted to 350 °C.
- Subprocess 2: Part of the exhaust gas is bypassed from the inlet of HRSG to the midpoint between NTA reactor and NH_3-SCR equipment, and the gas temperature is adjusted to 350 ° C.

As mentioned in Section 2, it was found that change of the combustion temperature of the gas turbine from 1300 °C to 1700 °C could increase the NO concentration in the exhaust gas and further reduce CO_2 emissions. Thus, for two case studies with gas turbine combustion temperatures of 1300 °C and 1700 °C, we investigated effects of introducing the above-mentioned two subprocesses on power generation efficiency of the entire system, respectively. In the present simulation analysis, simulation models for two types of gas turbine M701DA, M701JAC (Mitsubishi Heavy Industries, Ltd.) were applied to calculate process data for CCGT system applying 1300 °C class and 1700 °C class, respectively.

Table 1 shows an example of results for optimization of the heat exchanger network in the HSRG based on the pinch analysis for Subprocess 1. For the case study applying a 1700 °C class gas turbine, implementation of the Subprocess 1 increased the amount of power generated by the steam turbine (ST). In contrast, implementation of Subprocess 2 showed decrease in the amount of generated power. As shown in Table 1, positioning of the NTA reactor and the NH_3-SCR equipment differed slightly for Subprocess 1, which depends on the combustion temperature in the gas turbine.

4. Conclusions

For combined cycle power generation systems, we estimated effects of introduction methods for two different types of NTA processes (i) and (ii) on efficiency of the entire system, respectively. In application of NTA process (ii), an effect of the design temperature of the NTA process (around 350 °C) on the decrease in power generation efficiency was estimated, and furthermore the difference of changes in power generation efficiency between two class of gas turbines (1300 °C, 1700 °C) was also clarified. Hence, the optimization of the mass and energy balance and the evaluation of performance of CCGT system from the viewpoint of overall system were demonstrated to be useful for setting target temperature and target performance in the research and development of NTA catalysts.

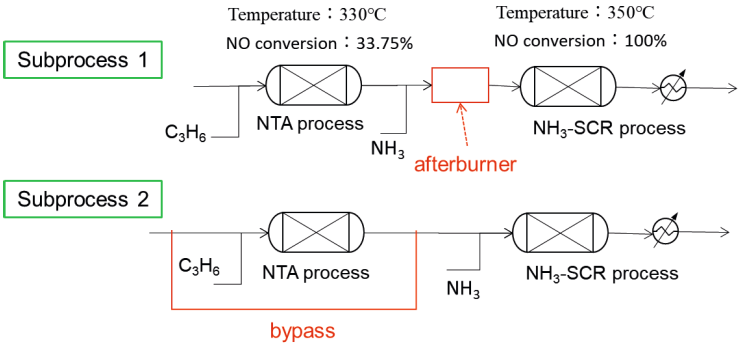


Figure 5 Introduction method of NTA process and NH_3-SCR process.

Table 1 Simulation results for optimization of the heat exchanger network in the HSRG for Subprocess 1.

	1700 °C class gas turbine	**1300 °C class gas turbine**
Additional fuel	17.81 MW	10.55 MW
Changes in the amount of power generated by ST	+4.6 MW (Increase)	+3.67 MW (Increase)
Changes in power generation efficiency	−0.6 pt (Decrease)	−0.4 pt (Decrease)
Optimum position of NTA & NH_3-SCR processes	In the middle of high-pressure secondary economizer & medium pressure overheating	In the middle of high-pressure steam economizer

Acknowledgement

This paper is based on results obtained from a project, P14004, commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

References

J. N. Galloway *et al*., 2008, Transformation of the Nitrogen Cycle: Recent Trends, Questions, and Potential Solutions, Science, **320**, 889-892.

K. Kobayashi *et al*., 2019, Effect of the TiO_2 crystal structure on the activity of TiO_2-supported platinum catalysts for ammonia synthesis via the NO–CO–H_2O reaction, Catal. Sci. Technol., **9**, 2898–2905.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50043-9

Knowledge integrated, deep neural network-based prediction of stress-strain curves of polymer matrix composites for AI-assisted materials design

Nagyeong Lee [a], Jaewook Lee [a] and Dongil Shin [a,b*]

[a] *Department of Chemical Engineering, Myongji University, Yongin, Gyeonggido, 17058, Korea*
[b] *Department of Disaster and Safety, Myongji University, Yongin, Gyeonggido, 17058, Korea*
[*] *dongil@mju.ac.kr*

Abstract

In order to achieve an effective energy transition, development of new materials must be accompanied with the development of new and renewable energy facilities. However, to this day, material design is costly because material development relies on the designer's intuition. Therefore, for the competitiveness of material development, AI-based material design automation must be made through the combination and composition prediction of components. As the first step in the AI-based material reverse engineering system, this study predicts the mechanical properties and behavior of polymer matrix composites (PMC).

The mechanical behavior of a material can be expressed from the strain-stress curve (S-S curve), and the deformation from the elastic section to the plastic section can be judged along with mechanical properties such as tensile strength, elastic modulus, and maximum load. Therefore, this study aims to predict the mechanical behavior of the PMC by learning the minimum tensile test data and information on the components for the two-component PMC based on the deep learning methodology.

Through literature/data analysis, most features that can affect mechanical properties were classified into two predictive models. The first predictive model inputs tensile test data and chemical/mechanical properties, and outputs mechanical properties behavior. And the second prediction model predicts by inputting structural information of each components. Through SMILES of each components, MACCS key was obtained and converted to use functional group information and used as a feature. As a result of comparing the performance of the two predictive models, the second model required less material information than the model that did not learn structural information, and performed better. As a result, it is a model that predicts the behavior of the plastic section beyond the existing prediction model that stayed in the elastic modulus section.

Keywords: Machine learning, Language process model, Polymer matrix composite (PMC), Mechanical property, Chemical language processing

1. Introduction

Stable securing of new materials is one of the important conditions to achieve efficient energy transition. For example, to solve the green mobility issue with low energy efficiency, many companies are paying attention to fuel efficiency improvement through

vehicle weight reduction and participating in material development. In particular, a material attracting attention in the face of energy conversion is polymer matrix composites (PMC).

For material development, it is essential to reflect mechanical properties, and there are tensile strength, modulus of elasticity, maximum load, maximum stress, break point, and stress. These can be easily derived from the S-S curve obtained from the tensile test. Mechanical properties and behaviors are expressed differently depending on the components (matrix, filler), the composition of each components, test conditions, process conditions, etc. Because of the various complexity, to this day, designs are made by the designer's intuition. For the competitiveness of material development, we proposed an AI-based reverse engineering system with the mechanical properties required for the application, away from the material design method that relied on experience.

There have been attempts to predict mechanical properties in the past, but in this study, the S-S curve problem over the entire section of the material, which was difficult to predict due to the large plasticity section and complex response, was predicted based on the deep neural network (DNN). The model predicts the entire S-S curve even in the absence of test data by using complex correlations between vast amounts of experimental data. In this paper, tensile test data in various compositions for a two-component combination using amorphous and partially crystalline polymers, which are often used as materials for transportation equipment, as a base material, and ceramic powder, glass fiber, carbon fiber, etc. as reinforcing materials was used to build a data-based prediction model.

2. Background

In this section, 2.1 describes previous studies and challenge for predicting PMC mechanical properties behavior, and 2.2 describes the theoretical background based on the study.

2.1. Challenge

There are two representative mathematical models that predict the S-S curve through numerical models. First, Ramberg-Osgood relationship:

$$\frac{\varepsilon}{\varepsilon_0} = \frac{\sigma}{\sigma_0} + (\frac{\sigma}{\sigma_0})^n \tag{1}$$

ε = *strain*, $\varepsilon_0 = \left(\frac{\sigma_0}{E}\right)$, $strain, \sigma = stress$, $\sigma_0 = yield\ strength$, $E = elastic\ modulus$. And a second is Hollomon piecewise power law:

$$\frac{\varepsilon}{\varepsilon_0} = \begin{cases} \frac{\sigma}{\sigma_0} & for\ \sigma \geq \sigma_0 \\ (\frac{\sigma}{\sigma_0})^n & for\ \sigma \geq \sigma_0;\ 1 < n \leq \infty \end{cases} \tag{2}$$

n=1, it is a complete elastic material, and when n=∞, it is an elastic-complete plastic material.

The two mathematical relational expressions, most of all, are not suitable for functions with high nonlinearity. In the case of Eq.1, the prediction accuracy is lowered for

materials with rapid changes in the plasticity section because the calculation is performed without separating the elastic/plastic section (H. C. Hyun et al.). On the other hand, in the case of Equation Eq.2, the prediction rate is higher than the Eq.1 using the section power function method, but it is difficult to apply a new material, due to that it is hard to categorise the carbon/plastic section. The mechanical behavior of PMC has more than one directionality and various parameters, so there is a limit to expressing it in a numerical formula.. Therefore, mechanical properties behavior including the plastic section of PMC is predicted through DNN.

2.2. Theoretical background

(1) Chemical structure information

Chemical structure information was expressed through chemical identifiers. Chemical identifiers are strings designed to encode chemical structures, chemicals, and molecular information. Types of formula identifiers include InCHI, SMILES, and SMARTS. Among them, SMILES advantageous for MACCS keys conversion was selected, and SMILES of matrix, filler were obtained through the database source of PubChem, respectively.

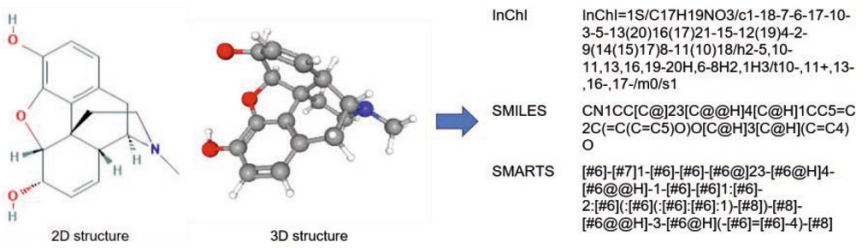


Figure 1. Example of String representation of 2D and 3D chemical structure information

(2) Chemical fingerprints

Fingerprints are the main expression methods that can confirm each molecular information in machine learning, and molecular structures and functional groups can be identified. Recently, predictions using functional groups of Fingerprints have been actively used in the field of drugs and catalysts. However, Binding energy, and Young's Modulus are also highly related to functional groups. When three specific functional groups are placed on Graphene sheets for mechanical properties simulation, the dimension and molecular structure of the functional group affect Binding energy and Young's Modulus. (Qingbin Zheng et al., 2010). In this study, among various methodologies representing Fingerprints of chemical substances, MACCSkeys represented by 0, 1 binary at 166 bits was selected as structural information and used for input. Functional groups represented by MACCS keys are related to the mechanical properties of the material and are used as input features (G. Chen, et al.).

3. Prediction of mechanical behavior based on chemical/mechanical information of components

3.1. Data collection

As for the tensile test data to be used for learning, the tensile test results according to the composition of each type of PMC in provided by the Korea Research Institute of Chemical Technology were used. (Under the same conditions, the test proceeds five times each.) The tensile test results include values such as the type of material, mark distance, and force according to strain. Additionally, the 'Poisson ratio' to reflect the difference in length strain according to the load direction, density for pore reflection, and density and molecular weight for each Matrix/Filler related to mechanical properties were extracted from PubChem's database(https://pubchem.ncbi.nlm.nih.gov/). Through data analysis, it was found that the test conditions were correlated with the S-S curve, so we added the mark distance and type as input features.

3.2. Data preprocessing

First, the experimental error data on the tensile test data is processed. Since the negative value of stress in raw data is physically impossible due to an experimental error, all data of negative values were converted to zero. In addition, values such as molecular weight, density, elastic modulus, Poisson ratio, marker distance, and stress of matrix and filler are pretreated to have a large range of 0.3 to 300. Standardization was performed using the stats module provided by Scipy to improve the performance of the model.

3.3. Model construction and training

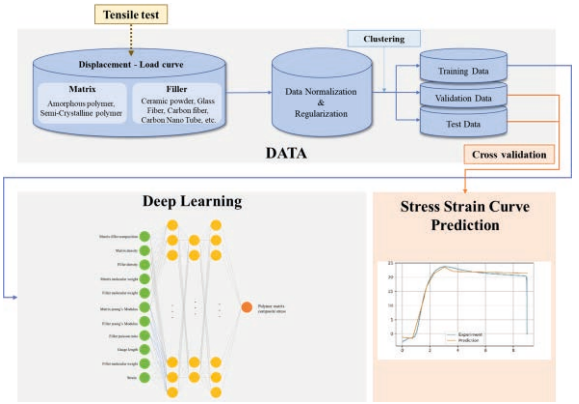


Figure 2. Workflow of mechanical behavior predictive model based on chemical/mechanical information of components. (As input, the density, molecular weight, young's modulus of each components (matrix/filler), matrix composition, filler poisson ratio, tensile test strain data, gauge length are used)

The model learns the remaining composition's data, and predict PMCs mechanical behavior according to the desired composition and test conditions. The learning data set and the test data set are divided into 5:1.The model has a Feed-forward Neural Network (FNN) structure, using TensorFlow Keras, four hidden layers, ReLu as an activation function, Adam as an optimization function, and Mean Square Error (MSE) as a loss function. To further prevent overfitting, Batch Normalization, Regularization, Dropout structure, He-normalization, and L1 Regularization were added to the structure. Bayesian

optimization was applied to optimize each hyperparameter, and the accuracy of the model was evaluated as R^2 (N. Lee, et al.).

4. Predicting mechanical behaviour based on structural information of components

4.1. Data collection

SMILES of each component material scraped from PubChem's database is used to reflect structure information in addition to the tensile test data conducted in Section 3.

4.2. Data preprocessing

Section 3 performed the same preprocessing for the same data. MACCS keys compared eight types of MACCS keys functional groups and used only 72 functional groups with differences among 166 functional groups. We compared eight types of component material's functional groups and used only 72 functional groups with differences among 166 functional groups.

4.3. Model construction and training

To check the influence of features, we completed a model with two different types of inputs. Therefore, the same network structure as the model in Chapter 3 was used in the model that reflects the structural information of the components. However, the prediction model was completed by selecting different optimal hyperparameters through the Bayesian optimizer.

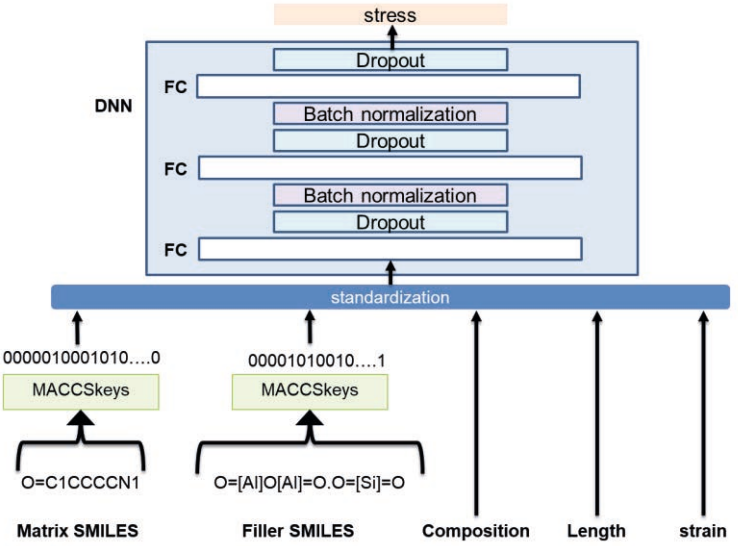


Figure 3. Workflow of mechanical behavior predictive model based on structural information of components.

5. Results

It can be seen that not only different PMC conditions, but also different composition and gauge distance affects the prediction results. Prediction model using the chemical/mechanical properties of the components showed an accuracy of R2≅ 0.45 ~

0.95 depending on the type of polymer composite, and the model using the structural information showed an accuracy of R2≅0.55 ~ 0.93.

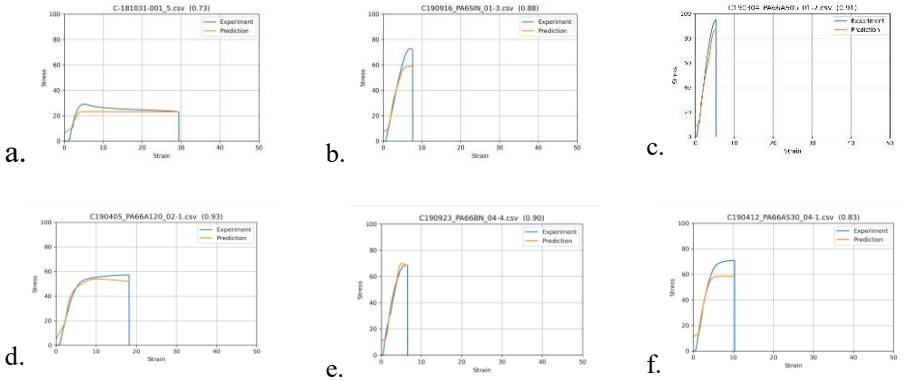


Figure 4. a, b, c above are suggested model based on chemical/mechanical information, and below d,e,f are suggested model based on structural information (a) PP+Al_2O_3, b) PA6+Si_3N_4, c) PA6,6+Al_2O_5Si, d) PA6+Al_2O_3, e),PA6,6+BN, f)PA6,6+Al_2O_5Si) : Orange line is 'prediction', blue line is 'experiment'

6. Conclusion

This study proposes a model for predicting the mechanical properties of a polymer composite resin using deep learning-based material information. Compared with previous studies utilizing chemical/mechanical properties of constituent materials, the model reflecting the component structure has similar performance, but has versatility by using data from fewer components. Compared to the first model that needs to know the seven chemical/mechanical properties of the constituent materials, it is possible to predict using only the SMILES structure information of each constituent material, that is, only two pieces of information about the constituent material. As a result, it was found that the structural information of the molecule had a great influence on the mechanical properties. Accordingly, it is expected that the mechanical properties can be predicted based on the minimum information of the constituent materials. As a follow-up study, the predictive performance of the model will be improved by improving the model structure through Natural Language Processing (NLP).

References

Q. Zheng, Y. Geng, S. Wang, Zh. Li, and J. Kim, 2010, Effects of functional groups on the mechanical and wrinkling properties of graphene sheets, Carbon, vol.48, issue15, 4315-4322

G. Chen, L. Tao, and Y. Li, 2021, Predicting polymers' glass transition temperature by a chemical language processing model, polymers, vol.13, 1898

H. C. Hyun, J.H. Lee and H. Lee, 2008, Mathematical Expressions for Stress-Srain Curve of Metallic Matrerial, Transactions of the KSME A , vol.32, 21-28

N. Lee, Y. Shin, and D.Shin, 2021, Prediction of Mechanical Properties and Behavior of Polymer Matrix Composites Based on Machine Learning. Journal of the Korean Institute of Gas, 25(2), 64-71

PubChem. Retrieved from https://pubchem.ncbi.nlm.nih.gov/

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50044-0

EVALUATION OF ECONOMIC PERFORMANCE OF CO_2 SEPARATION PROCESS USING MIXED MATRIX MEMBRANE

Kakeru FUJITA[a], Ryousuke AKIMOTO[a], Yasuhiko SUZUKI[b], Yuki OGASAWARA[b], Masaru NAKAIWA[a] and Keigo MATSUDA[a*]

[a] Department of Chemistry and Chemical Engineering, Graduate School of Science and Engineering, Yamagata University, 4-3-16, Jonan, Yonezawa, Yamagata 992-8510, Japan
[b]Faculty of Chemistry and Chemical Engineering, Yamagata University, 4-3-16, Jonan, Yonezawa, Yamagata 992-8510, Japan

[*]Corresponding Author's E-mail: matsuda@yz.yamagata-u.ac.jp

ABSTRACT: A multi-stage CO_2 capture process using Mixed Matrix Membrane (MMM) could separate molecules with similar molecular diameter such as CO_2 and N_2 by dissolution and diffusion. The required energy and the membrane area were investigated based on the membrane performance such as CO_2 permeance and CO_2/N_2 selectivity. In this multi-stage CO_2 capture process, the feed gas fed into the first and second membranes. The gas permeated through the second membrane was recycled to Feed. The permeate gas from the first membrane was fed to the adsorption column to separate H_2O. The CO_2-rich dry gas was liquefied by using a compressor and a condenser. The residual gas was separated using a membrane, and the permeate gas was recycled to the dry gas. The flue gas assumed from a coal-fired power plant containing 11.6 mol% CO_2. When the CO_2 permeance of 1000 GPU and CO_2/N_2 selectivity of 50 were used for the first and second membranes of the multi-stage CO_2 capture process, the required energy was 139 MW and the membrane area was 3.6×10^6 m^2, respectively. From these results, the operation, construction, and membrane skid costs were calculated. The CO_2 capture cost per ton of CO_2 was found to be \$38/ton-CO_2. In addition, the membrane area of the second was large, and the membrane area could be reduced by using a membrane with high CO_2 permeance and low CO_2/N_2 selectivity. Therefore, a membrane with CO_2 permeance of 1000 GPU and CO_2/N_2 selectivity of 50 was used in the first. A membrane with CO_2 permeance of 3000 GPU and CO_2/N_2 selectivity of 30 was used in the second. As a result, it was clarified that the required energy was 141 MW, and the membrane area was 1.25×10^6 m^2. The CO_2 capture cost was \$29/ton-CO_2.

Keywords: Membrane separation, Carbon dioxide, Process design

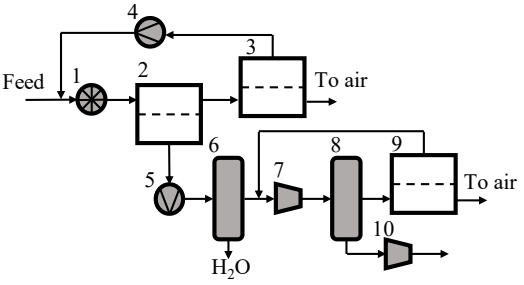
1. Introduction

United Nations was accelerating its efforts on the Sustainable Development Goals (SDGs), and the number 13 climate change has been an urgent issue. The main cause of climate change was the increase of greenhouse gases such as CO_2 (T. M. Lenton *et al*., 2019). IEA has been announced that CO_2 emissions in 2021 would be expected to be about 33 billion tons. Especially, 11 billion tons have been emitted from coal-fired power plant (Ministry of Economy, Trade and Industry, 2019). Recently, to achieve drastic reduction of CO_2 emissions, Carbon dioxide capture and storage (CCS) which is the separation and storage carbon dioxide from large-scale intensive CO_2 emission sources has attracted much attention. Examples of CO_2 separation technologies include chemical absorption and membrane separation. In chemical absorption, gases containing CO_2 have been absorbed in an alkaline aqueous solution in an absorption tower. After that, the absorbed solution has been sent to the stripper where thermal energy is required to strip the CO_2 in the absorbed solution. Although gas absorption could recover CO_2 at a concentration of 99 mol% or higher, it consumes a large amount of thermal energy in the stripper that results in high CO_2 capture cost (D. Leeson *et al*., 2017). On the other hand, in membrane separation, the driving force has been the difference in partial pressure between the permeate and retentate of the

membrane. Therefore, the only energy required for separation was a pressure exchanger such as a compressor or a vacuum pump (A. Stankiewicz *et al*., 2000). Membranes include inorganic membranes such as zeolite which are permeable by molecular sieves. Polymeric membranes have been separated by dissolution and diffusion. Polymeric membranes are used with close molecular diameter such as CO_2 and N_2 because the effect of molecular sieving is less effective and the separation proceeds by dissolution and diffusion. Among polymer membranes research and development of organic-inorganic hybrid membranes (MMM) which have the advantages of durability of inorganic materials and excellent gas permeability of organic materials has been conducted (M. Tanaka, 2016). However, few studies have been conducted on the required energy and membrane area of processes how parameters such as permeance and selectivity. Therefore, process synthesis and integration based on the process systems engineering approach have been demanded as socioeconomic innovations (B. Ghalei *et al*., 2017). Process synthesis consists of three methods. 1. Planning of process that examines the selection and combination of process equipment, 2. Functional design of the process that quantitatively assigns functions to this equipment, 3. Evaluation of these process that meet their intended functions. In this study, the multi-stage CO_2 capture process using MMM was developed via process synthesis method and evaluate the economic performances.

2. Modelling

Fig. 1 shows schematic diagrams of multi-stage CO_2 capture process for CCS (T. C. Merkel *et al.*, 2010). Assuming exhaust gas from a coal-fired power plant, this process separated to a recovery ratio of 90% CO_2 against flow rate of CO_2 on Feed. The feed was assumed following conditions; flow rate of 2.2×10^4 mol/s, a pressure of 100 kPa, a temperature 298 K, a composition CO_2:11.6, N_2:73, H_2O:11, O_2:4.4 mol%. The process consists of a pressure exchanger (blower, compressor, vacuum pump), an adsorption column to remove H_2O, a condenser and three membrane modules. The feed gas pressure is elevated to 200 kPa with a blower which fed into the membranes 2 and 3. The permeate is depressed by 20 kPa with vacuum pomp. The permeate gas from membrane 3 is recycled to the feed gas. The permeation gas of membrane 2 separated H_2O with adsorption tower of 6. The dry gas is compressed to 2250-3800 kPa with compressor of 7 (253 K). The gas is pumped into the ground at 14,000 kPa by compressor 10. The residual gas in the condenser is separated by the membrane of 9, and the permeate gas is recycled to the dry gas. The membrane module is a cross plug flow module in which the permeate gas exits in the flow direction. The CO_2 permeance and CO_2/N_2 selectivity set to be 1000, 3000, 5000 GPU, CO_2/N_2:10-100, CO_2/H_2O:0.03, N_2/O_2:1, respectively. All of simulation was implemented of the Aspen Plus®V11. The Peng Robinson type equation of state was applied to estimate the vapor-liquid equilibrium for the steady state.



1 : Blower 2, 3, 9 : Cross-plug flow module 4, 5 : Vacuum pump
6 : Adsorption tower 7, 10 : Compressor 8 : Condenser

Fig. 1 Schematic diagram of multi-stage membrane CO_2 capture process (CCS)

3. Result and discussion

3.1 Effect of CO_2 Permeance and CO_2/N_2 Selectivity on CO_2 concentration for CCS, required energy, and membrane area

First, the effects of the same CO_2 permeance and CO_2/N_2 selectivity of membranes 2, 3, and 9 on the CO_2 concentration for CCS, the required energy and the membrane area are investigated. In this study, the CO_2 permeance are set to 1000, 3000, and 5000 GPU, and the CO_2/N_2 selectivity is varied in the range of 10-100. Table 1 shows the comparison of CO_2 concentration for CCS, required energy and membrane area. The required energy is the sum of the CO_2 separation process (1-4) and CO_2 storage process (5-10). The membrane area is total of 2, 3 and 9. The CO_2 concentration for CCS increased with high CO_2/N_2 selectivity. The required energy for the storage process decreases with high CO_2/N_2 selectivity due to the higher CO_2 concentration in permeate gas of membrane 2. In addition, the required energy with CO_2/N_2 selectivity 100 decreases by 60% compere to CO_2/N_2 selectivity 10. On the other hand, the CO_2 partial pressure on the permeate is larger for the high CO_2 concentration on the permeate. As a result, the difference between the CO_2 partial pressure on the retentate and that on the permeate is smaller, and the membrane area of the multi-stage CO_2 capture process increases. In addition, the membrane area with CO_2/N_2 selectivity100 increases by 225% compere to CO_2/N_2 selectivity 10. On the other hand, by increasing the CO_2 permeance from 1000 to 5000 GPU, the membrane area is reduced by 80% due to the flow rate of CO_2 on permeates increases.

Table1 Comparison of CO_2 mole fraction, required energy and membrane area

CO_2/N_2 selectivity	-	10	20	30	40	50	60	70	80	90	100
CO_2 concentration	mol%	88.5	93.9	95.6	96.5	97.0	97.3	97.4	97.5	97.7	97.8
Required energy	MW	285	213	170	151	139	133	124	119	117	115
Membrane area (1000 GPU)	$\times 10^6$ m^2	2.40	2.50	2.90	3.20	3.60	4.00	4.30	4.70	5.10	5.40
Membrane area (3000 GPU)	$\times 10^6$ m^2	0.80	0.83	0.97	1.07	1.20	1.33	1.43	1.57	1.70	1.80
Membrane area (5000 GPU)	$\times 10^6$ m^2	0.48	0.50	0.58	0.64	0.72	0.80	0.86	0.94	1.02	1.08

3.2 Evaluate the economic performance for CO_2 capture process for CCS

The cost index is shown in Table 2. The CO_2 capture cost, CC is estimated by Eq. (1)

$$CC = \frac{(P \times T \times E) + (0.2 \times C)}{F_{CO2} \times T} \quad (1)$$

P is the required energy for CO_2 capture process (kW), T is the annual operating time (h/year), E is the cost of electricity (\$/kWh), C is the capital cost of the CO_2 capture process (\$), F_{CO2} is the mass flow rate of captured CO_2 (T. C. Merkel *et al*., 2010). Fig. 2 shows CO_2 capture cost for CO_2 permeance and CO_2/N_2 selectivity. The CO_2 capture cost decreases because the reducing the operating cost is larger than the increasing the membrane skid cost with high CO_2/N_2 selectivity. The high CO_2 permeance has a significant effect on the CO_2 capture cost by reducing the membrane area due to increasing flow rate of CO_2 on permeate. It is clarified that the minimum CO_2 capture cost is 18.6 US\$/ton-CO_2 with the CO_2 performance of 5000 GPU and CO_2/N_2 selectivity of 90 under these conditions.

Table 2 Cost index of CO_2 capture process

Category	Units	Value
Mechanical efficiency	-	0.8
Mechanical cost	\$/kW	500
Membrane skid cost	\$/m^2	50
Cost electricity	\$/kWh	0.04
Annual operating time	h/year	7446

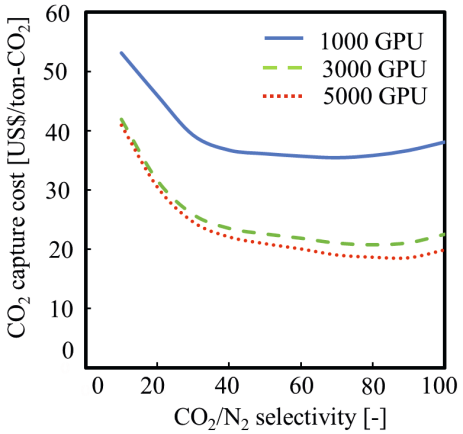


Fig. 2 CO_2 capture cost for CO_2 permeance and CO_2/N_2 selectivity

3.3 Configuring membranes with appropriate separation performance

In this study, the required energy and the membrane area can be reduced with appropriate separation performance in membrane 2 and 3. Table 3 shows the CO_2 permeance and CO_2/N_2 selectivity configured. The membranes used are those with CO_2 permeance of 1000 GPU and CO_2/N_2 selectivity of 50 and those with CO_2 permeance of 3000 GPU and CO_2/N_2 selectivity of 30. In Scenarios 1 and 2, the same CO_2 permeance and CO_2/N_2 selectivity are used for membrane 2 and 3. In Scenarios 3 and 4, membranes with different CO_2 permeance and CO_2/N_2 selectivity are used. Fig. 3 shows the effect of whole scenario on the membrane area. Since the multi-stage CO_2 capture process using MMM requires a larger membrane area for membrane 3 than for membrane 2, membrane 3 with a CO_2 permeance of 1000 GPU and a CO_2/N_2 selectivity of 50 resulted in a larger membrane area due to the pressure difference problem as explained in Section 3.1. In the scenario where the CO_2 permeance and CO_2/N_2 selectivity are changed for each membrane, the membrane area in scenario 3 decreases by 64% compared to scenario 4. Thus, it is found that the membrane area can be reduced by configuring with high CO_2 permeance and low CO_2/N_2 selectivity at membrane 3. Scenario 1 results in the smallest membrane area in whole scenario.

Table 3 Membrane configuration of CO_2 capture process

Category	Unit	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Membrane 2					
CO_2 permeance	GPU	3000	1000	1000	3000
CO_2/N_2 selectivity	-	30	50	50	30
Membrane 3					
CO_2 permeance	GPU	3000	1000	3000	1000
CO_2/N_2 selectivity	-	30	50	30	50

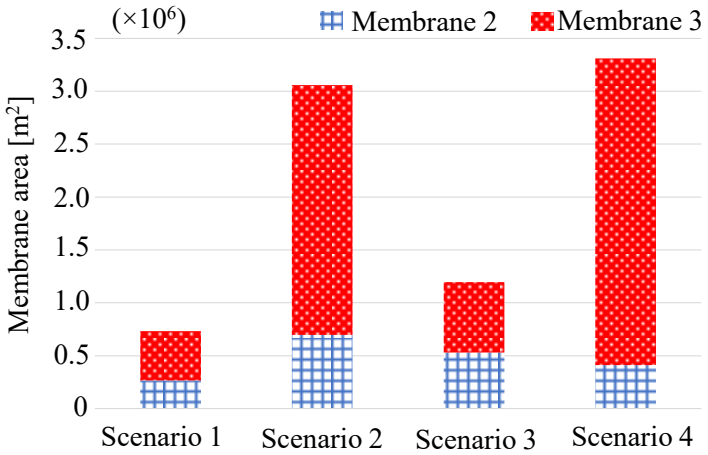


Fig. 3 The effect of membrane area on CO_2 permeance and CO_2/N_2 selectivity

Fig. 4 shows the effect of whole scenario on the required energy. The required energy is the sum of the CO_2 separation process (1-4) and the CO_2 storage process (5-10) as in the study in Section 3.1. The required energy in the CO_2 storage process (5-10) is larger for each scenario in Figure 4. However, required energy for CO_2 storage is reduced in Scenarios 1 and 2. In Scenario 1, the membrane area is the smallest, but the required energy is the largest at 177 MW. In Scenario 3, the required energy is reduced by 33 MW compared to Scenario 1 by setting the CO_2/N_2 selectivity 50 on membrane 2. Furthermore, the required energy in Scenario 3 is reduced by 8 MW compared to Scenario 4 by considering the membrane configuration with appropriate separation performance. In addition, cost evaluation of scenario 3 and scenario 4 is done. The cost of Scenario 3 is \$29/ton-CO_2 and that of Scenario 4 is \$37/ton-CO_2. By considering the appropriate CO_2 permeance and CO_2/N_2 selectivity configuration of the multi-stage CO_2 capture process using MMM as in Scenario 3, the membrane area and required energy are reduced compared to Scenario 4. The CO_2 capture cost is reduced by \$8/ton-CO_2.

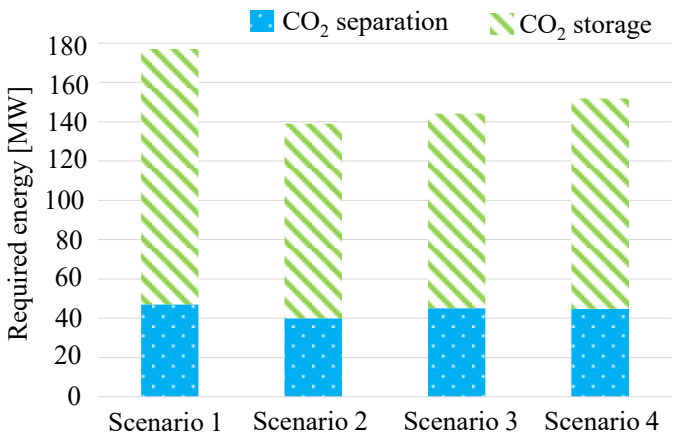


Fig. 4 The effect of membrane area on CO_2 permeance and CO_2/N_2 selectivity

4 Conclusions

The multi-stage CO_2 capture process using MMM were fully developed by Aspen plus V11. The multi-stage CO_2 capture process using MMM was developed via process synthesis method and evaluate the economic performances. The required energy decreased with membranes of high CO_2/N_2 selectivity because of CO_2 concentration on the permeate gas increased. On the other hand, the membrane area increased due to the smaller in the pressure difference as the high CO_2/N_2 selectivity. The CO_2 permeance had a significant effect on the reduction of the membrane area. Regarding the economic evaluation, increasing the CO_2/N_2 selectivity from 30 to 100 reduces the required energy, but because the membrane area increased, the effect of increasing the CO_2/N_2 selectivity above 30 on the cost reduction was small. If a membrane with a CO_2 permeance of 5000 GPU or higher could be developed, the CO_2 capture cost will be less than \$20/ton-CO_2. By configuring membranes with appropriate CO_2 permeance and CO_2/N_2 selectivity in a multi-stage membrane CO_2 capture process, the required energy and membrane area could be reduced.

Acknowledgment

This work was partially supported by JST Mirai project "Development of mixed matrix porous membrane endowed with high performance in CO_2 selectivity and anti-aging".

References

B. Ghalei, K. Sakurai, Y. Kinoshita, K. Wakimoto, A. P. Isfahani, Q. Song, K. Doitomi, S. Furukawa, H. Hirao, H. Kusuda, S. Kitagawa, E. Sivaniah, "Enhanced selectivity in mixed matrix membranes for CO_2 capture through efficient dispersion of amine-functionalized MOF nanoparticles," *NATURE ENERGY,* **2**, 17086 (2017)

D. Leeson, N. M. Dowell, N. Shah, C. Petit, P. S. Fennell, "A Techno-economic analysis and systematic review of carbon capture and storage (CCS) applied to the iron and steel, cement, oil refining and pulp and paper industries, as well as other high purity sources," *Int. J. Greenhouse Gas Control*, **61**, 71–84 (2017)

T. M. Lenton, J. Rockström, O. Gaffney, S. Rahmstorf, K. Richardson, W. Steffen, H. J. Schellnhuber, "Climate tipping points-too risky to bet against," *NATURE ENERGY*, **575**, 592-595(2019)

T. C. Merkel, H. Lin, X. Wei, R. Baker, "Power plant post-combustion carbon dioxide capture: An opportunity for membranes," *J. Membrane Sci*, **359**, 126-139 (2010)

Ministry of Economy, Trade and Industry; White Paper 2019

A. Stankiewicz, J. A. Moulijn, "Process Intensification: Transforming Chemical Engineering," *Chem. Eng. Prog*., **96**, 22-34 (2000)

M. Tanaka, "Gas separation using polymeric membranes," *Chemical and education*, **41**(4), 173-177(2016)

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50045-2

Nature vs engineering: Production of methanol from CO_2 capture

Guillermo Galán,[a] Mariano Martín,[a,*] Ignacio E. Grossmann.[b]

[a]*Department of Chemical Engineering. University of Salamanca. Plz. 1-5, 37008, Salamanca, Spain*
[b]*Department of Chemical Engineering. Carnegie Mellon University, 5000 Forbes Ave, 15213, Pittsburgh PA, USA*
** mariano.m3@usal.es*

Abstract

This work compares integrated facilities to capture CO_2 from the atmosphere and use it for the production of bulk chemicals, methanol. Two different alternatives have been proposed. On the one hand, the use of direct air capture (DAC) employing either alkaline solutions based in KOH or a bipolar membrane electrodialysis (BPMED). The CO_2 captured is subsequently hydrogenated with electrolytic hydrogen produced using solar and/or wind energy. On the other hand, the use of biomass such as switchgrass, corn stover, miscanthus, wheat straw and forest residues, from spruce and pine, are considered. This biomass is pretreated, gasified, either direct or indirect gasification, the raw syngas followed steam reforming or partial oxidation, it is cleaned and its composition is adjusted for the synthesis of methanol. All units are modelled individually to formulate the superstructure as an MINLP optimization model. The results show that the optimal option consists of the use of spruce bark biomass gasification. The direct air capture has production and investment costs almost 10 times higher due to the large consumption of electricity to power the fans.

Keywords: Process design, CO_2 capture, biomass, renewable methanol

1. Introduction

Since the 18th century with the beginning of the industrial revolution, as well as the development of the use of steam and other energy sources, mainly fossil fuels, human growth and its development was linked to the increasing emission of carbon. Because of this, humanity is in a race to reduce emissions to keep the planet's temperature within 1.5 °C. Some efforts to remove CO_2 from the atmosphere are thus being investigated (Allen et al., 2021). Purified CO_2 can be used in the chemical industry for food production, cosmetics and even for the population of intermediate reagents for the production of acids and aldehydes, among others. This work considers the use of CO_2 to obtain intermediate compounds such as methanol. Two major capture technologies can be used. On the one hand, nature captures CO_2 to grow biomass. This biomass is later gasified to produce methanol. On the other hand, direct air capture (DAC) is an engineered alternative that is emerging in our attempt to remove CO_2 from the atmosphere. By using an air-water contactor, it allows a constant flow of air to circulate through alkaline solutions. CO_2 is absorbed and captured by transforming it into $CaCO_3$. This carbonate is subsequently calcined, thus releasing the captured CO_2 to be purified (Keith et al., 2018). A modification of the DAC process consists of the use of a bipolar electrodialysis membrane (Sabatino et al., 2020). Ion exchange membranes allow the recovery of CO_2 through the

use of water and the subsequent regeneration of the H^+ and OH^- species in the respective solutions. The CO_2 is hydrogenated with electrolytic H_2 to produce methanol. This work evaluates, from a techno-economic perspective, both pathways towards the production of methanol comparing the nature and the manmade alternatives. The processes are optimized using a mathematical modelling approach.

2. Process description

The alternative based on biomass requires washing and milling before the gasification. Two technologies are considered: **The Renugas gasifier (R),** direct gasification, operates at medium pressure using oxygen and produces a gas rich in CO_2. It allows large throughput per reactor volume and reduces the need for a downstream pressurization. However, its efficiency is lower (Eggeman, 2005). The low pressure gasifier, **Battelle Columbus (Ferco, F)**, is indirectly heated. The system consists of two chambers, a gasifier and a combustor. Olivine is heated up by burning char to provide the energy for gasification. The syngas shows low CO_2 content but heavier hydrocarbons (Phillips, 2007). Subsequently, the syngas is reformed to remove the hydrocarbons. **Steam reforming (S)** is endothermic but provides a higher concentration of hydrogen in the syngas. **Partial oxidation (O)** is exothermic but its yield to hydrogen is lower. Finally, the raw syngas is cleaned. Two steps are proposed. Cold cleaning by means of a scrubber for low pressure gasification, or a ceramic filter operating at high temperature for high pressure gasification. The second step consists of a multibed PSA system used to remove the last traces of hydrocarbons, H_2S and CO_2 in that order. Once the syngas is purified, we use it for traditional methanol synthesis. The superstructure of alternatives is presented in Figure 1.

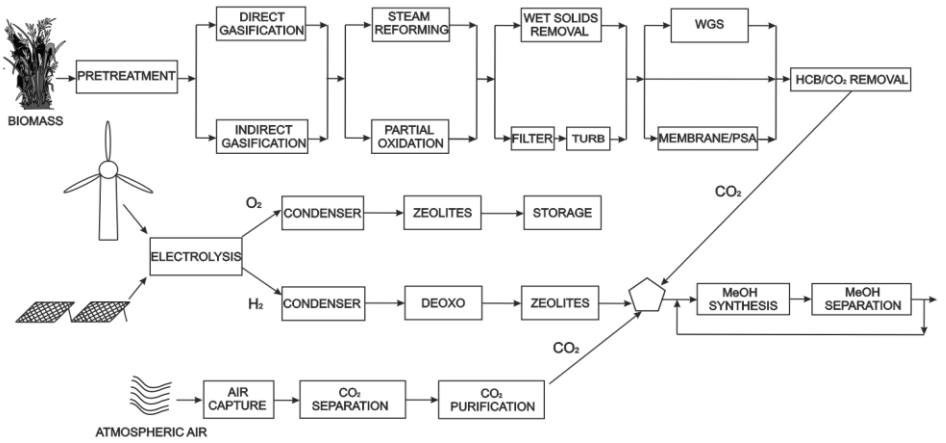


Figure 1.-Superstructure for the production of methanol

The DAC process captures the CO_2 from the air using a counterflow air contactor. The concentration of CO_2 in air is around of 300-400 ppm. This low concentration requires the use of alkaline salts, KOH, to reduce the water flow (Keith et al., 2018). Once the CO_2 has been captured, it can be recovered from the alkaline solution with a pellet-type reactor where CO_2 byproducts react with a stream rich in $Ca(OH)_2$, forming $CaCO_3$ and regenerating the pH of the original solution, releasing OH^-groups. Later the calcination of $CaCO_3$ allows recovering the CO_2 producing CaO, which is sent to a slaker unit where

the $Ca(OH)_2$ is regenerated. The high temperatures reached allow to produce medium-low pressure steam, obtaining energy in a turbine. The alternative process employs the bipolar electrodialysis membrane (BPMED), separating a basic solution rich in CO_2 and an acid solution. The control of pH is important to maintain most of the CO_2 dissolved, the recovery yield in the membranes and the regeneration of the H^+ and OH^- species in the respective solutions, avoiding CO_2 bubbles (Sabatino et al., 2020). The stream with CO_2 recovered is sent to the condenser and molecular sieves units to remove the water content and lately hydrogenated with electrolytic hydrogen produced using energy from PV panels or wind turbines is evaluated. The CO_2 is hydrogenated using renewable H_2 for the production of methanol by eq. (1).

$$CO + 2H_2 \leftrightarrow CH_3OH$$
$$CO_2 + H_2 \leftrightarrow CO + H_2O \quad (1)$$

3. Modelling approach

The different units are modelled using first principles, and mass and energy balances based on detailed simulations and/or experimental data. For particular units such as the direct air capture including the bipolar membrane electrodialysis (Sabatino et al., 2020), surrogate models are developed.

The gas composition produced from the gasifiers is defined by experimental correlations. The reforming stage uses conversions from the literature (Eggemann, 2005; Phillips et al., 2007)

The capture of CO_2 from air needs the use of an air contactor where the air and the water streams are placed in contact. The efficiency of capture and the molar flow rate of CO_2 through the membranes are a function of the concentration of bicarbonate and carbonate ions, $[HCO_3]^-$ and $[CO_3]^{2-}$, the concentration of KOH, [KOH], and current density, i. (Sabatino et al., 2020).

$$\eta_{current} = f([HCO_3]^-, [CO_3]^{2-}, [KOH], i)$$
$$f_{CO_2} = f([HCO_3]^-, [CO_3]^{2-}, [KOH], i) \quad (2)$$

The methanol synthesis reactor is modelled based on chemical equilibrium, mass and energy balances (Cherednichenko, 1953).

The superstructure is formulated in terms of total mass flows, component mass flows, component mass fractions, and temperatures of the streams in the network.

4. Optimization procedure

The superstructure is decomposed into three different flowsheet alternatives based on the pretreatments: direct air capture with alkaline solutions, direct air capture with bipolar membrane electrodialysis and the gasification of biomass. An NLP problem is solved for each one where the objective function consists of simplified production costs given by eq. (3)

$$Z = P_{MetOH} m_{MetOH} + P_{O_2} m_{O_2} - P_{Electricity} \sum_i W_{consumed} - \sum_i P_{utilities} \cdot m_{utilities} \quad (3)$$

subject to the models described in section 3. The NLP's consist of around 2000-2500 eqs and 3000 variables and it was solved with GAMS, CONOPT, requiring 30-60 s of CPU-time. After the optimization, a heat exchanger network is designed to reduce energy consumption. Finally, a detailed economic evaluation of the alternatives is performed to

compute the production and investment costs of the facility using the cost correlations in Martín and Grossmann (2011) and the procedure described in that work.

5. Results

This section shows the principal operating results, and the economics of the different alternatives. For wind and solar capture, we consider Cadiz, to the south of Spain where high solar irradiance is available, and the wind velocity is fairly high. Different types such as switchgrass, corn stover, miscanthus, wheat straw and pruning residues including pine and spruce bark are evaluated

5.1. Process analysis

The superstructure of alternatives is decomposed by technology to evaluate the yield and performance of each alternative. The biomass path follows indirect gasification followed by steam reforming, since the H_2 to CO ratio required for the production of methanol is around 2, After the gas clean up and the adjustment of the composition, the syngas is fed to the synthesis loop. In the case of DAC processes both alternatives are presented. Tables 1 and 2 show the major results. In general, DAC needs more energy than biomass gasification due to the low concentration of CO_2 in air that forces to move large volumes of air through the fans increasing the requirements of electricity, that is generated using wind turbines or solar panels. The conventional DAC is more efficient than the BPMED resulting in 10% lower energy requirements. However, the cost of PV panels and aerogenerators, together with the requirements of a large surface, increase considerably the investment cost of this technology. Thus, the yield to methanol is higher from those wastes with a composition richer in carbon.

Table 1. Major yields for gasification of biomass

	Gasification					
	Switchgrass	Corn Stover	Wheat Straw	Miscanthus	Pine Bark	Spruce Bark
Product cost(€/kg_{MetOH})	0.192	0.216	0.170	0.224	0.169	0.110
Investment (M€)	181.27	175.09	157.67	172.27	144.16	152.79
kg_{MetOH}/ $kg_{Biomass}$	0.658	0.674	0.620	0.687	0.875	0.816
Productivity (t/ha)	12.00	10.92	7.30	10.00	6.94	4.08
Surface required (ha)	51,840	55,925	90,47	59,552	67,334	122,980

Table 2. Major yields for DAC process paths

		DAC			
		Conventional Process PV panels	**Conventional Process Wind**	**BPMED PV panels**	**BPMED Wind**
Product cost (€/kg_{MetOH})		0.934	1.089	1.059	1.233
Investment (M€)	Plant	959.45	959.45	1,218.75	1,218.75
	PV/Aerogenerators	449.86	685.50	503.83	767.75
Surface required (ha)	PV panels	205.12	-	230.52	-
Number of units	Aerogenerators	-	572	-	643
$kgCO_{2\ air\ captured}$/kW		573.60	573.60	867.03	867.03
$kgCO_{2\ available\ captured}$/kW		968.79	968.79	867.03	867.03

5.2. Process economics

Table 1 presents the investment and the production costs for the six different biomass species. The most economic ones are pine and spruce bark since the higher composition in carbon lead to larger yields to methanol, reducing the production costs. Although spruce bark biomass shows an investment cost above that of pine bark and the largest

growing area, the lower cost of biomass, even with a slighter lower yield of kg $_{methanol}$/kg$_{Biomass}$, results in the best option, for a production cost is 0.11 €/kg methanol. The use of biomass is competitive with the production cost of methanol from fossil resources.

Another issue would be to be able to meet the global demand using biomass waste. Although growth of biomass is very efficient to capture CO_2, it is important to indicate that DAC technology is still at an early stage of development and with potential for improvement. Table 2 shows the investment and the production cost of the two DAC alternatives, conventional process and BPMED, both with a renewable power supply from PV panels and/or aerogenerators. These costs could change as a function of the chosen technology and the location. Conventional DAC process has a lower ratio kgCO_2 $_{air\ captured}$/kW than BPMED process due to the use of biogas as fuel in the calciner. The CO_2 from the combustion of this biogas is added to the CO_2 captured from air, decreasing the volume of air and with that the power consumption of the fans, with the corresponding reduction in PV panels and wind turbines.

The location has a direct effect due to the availability of resources. In this case the location was the province of Cádiz (Spain), which shows long sun hours and high wind velocities. The best option corresponds to the use of the conventional DAC power with PV panels. The production and investment costs are the lowest among the DAC alternatives, 0.934 €/kg methanol and 1409.31 M€, which are around 9 and 10 times larger than the values of the best biomass process. Figure 2 shows that the investment cost of the PV panels represents around of 32% of the total, i.e., a third of the investment cost is destined to the energy requirements. The breakdown of the investment in the equipment shows that the fans, the PV panels, and the electrolysis represent around of 90% of the total investment cost, leaving only the remaining 11% destined to capture of CO_2 and synthesis of methanol. The expected improvement in the efficiency of the solar panels would reduce not only the number of panels and the total surface but also the cost.

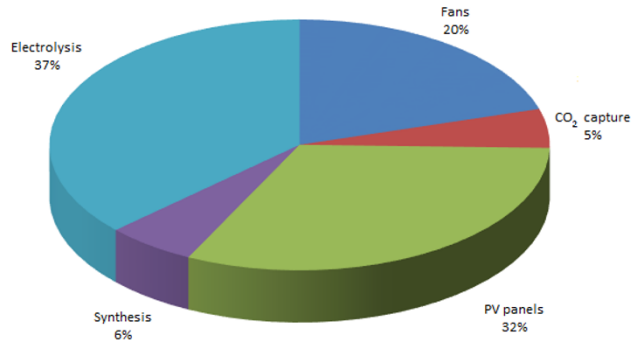


Figure 2.-Distribution of costs in DAC conventional process with PV panels

6. Conclusions

This work systematically compares the capture of CO_2 from atmosphere using a natural, biomass growth, and manmade, direct air capture (DAC), alternatives towards the production of a bulk chemical, methanol, in an attempt to build a sustainable chemicals industry as well. Two different configurations for DAC are optimized, the use of alkaline solutions (Keith et al., 2018) and a membrane (Sabatino et al., 2020), that captures the

CO_2, which is subsequently hydrogenated using electrolytic hydrogen. The facility is powered using wind or solar energy depending on the regional availability. This technology is compared to the Nature's alternative based on biomass, considering switchgrass, corn stover, miscanthus, wheat straw, and forest residues (spruce, pine). The biomass harvested is fed to a gasification-based process, consisting of indirect gasification, steam reforming, syngas clean-up and composition adjustment and methanol synthesis. The most economical alternative to remove CO_2 from the atmospheric air is the gasification of spruce bark biomass due to present a lowest production cost, 0.11 €/kg methanol, a low investment cost, 153M€, and the high yield to methanol, 0.816 kg $_{methanol}$/kg $_{biomass.}$ DAC technologies still show 10 higher production and investment costs, that are expected to decrease with the improvements in PV panels, wind turbines and the capture process itself.

Acknowledgment

The authors thank the PSEM3 GIR at USAL and the CAPD at CMU. GG appreciates the FPU PhD fellowship from the Spanish Government.

References

M. Allen, O. P. Dube, W. Solecki, F. Aragón-Durand, W. Cramer, S. Humphreys, M. Kainuma, J. Kala, N. Mahowald, Y. Mulugetta, et al, (2020). Global Warming of 1.5°C; An IPCC Special Report on the Impacts of Global Warming of 1.5°C above Pre-Industrial Levels and Related Global Greenhouse Gas Emission Pathways, in the Context of Strengthening the Global Response to the Threat of Climate Change, Sustainable Development, And Efforts to Eradicate Poverty; IPPC.

V.M. Cherednichenko, V. M., Dissertation, Karpova, Physico Chemical Institute, Moscow, U.S.S.R., 1953.

T. Eggeman, 2005, Updated Correlations for GTI Gasifier – WDYLD8. Technical memorandum for Pam Spath, National Renewable Energy Laboratory, Golden, Colorado. June 27, 2005.

D. W. Keith, G. Holmes, D. St. Angelo, K. Heidel., (2018) A Process for Capturing CO_2 from the Atmosphere. Joule 2, 1573–1594

M. Martin, I.E. Grossmann, (2018). Towards zero CO_2 emissions in the production of Methanol from switchgrass. CO_2 to methanol. Comp. Chem. Eng. 105, 308–316

S. Phillips, A. Aden, J Jechura, D Dayton, T Eggeman, 2007, Thermochemical ethanol via indirect gasification and mixed alcohol synthesis of lignocellulosic biomass. Technical Report, NREL/TP-510–41168, April 2007.

F. Sabatino, M. Mehta, A. Grimm, M. Gazzami, F. Gallucci, G. J. Kramer, M. van Sint Annaland, (2020) Evaluation of a Direct Air Capture Process Combining Wet Scrubbing and Bipolar Membrane Electrodialysis. Ind. Eng. Chem. Res., 59, 7007−7020

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50046-4

Superstructure Optimization for the Design of an Algae Biorefinery Producing Added Value Products

Maryam Raeisi[a*], Jiawei Huang [a], Thien An Huynh [a], Meik B. Franke[a], Edwin Zondervan[a]
[a]*Sustainable Process Technology, Faculty of Science and Technology, University of Twente, Meander, kamer 216, Postbus 217, 7500 AE Enschede, the Netherlands*
m.raeisi@utwente.nl

Abstract

This study presents a superstructure framework to evaluate processing pathways for the production of omega-3 and pigments in an algae biorefinery. Different stages such as cultivation, harvesting, dewatering, drying, cell distribution, and extraction are considered as processing sections in this superstructure. To simplify and speed up modelling, each of these technologies is grouped in blocks.

The superstructure framework is converted to a mixed-integer nonlinear programming (MINLP) model. It has more than 6.000 constraints/variables. The model is implemented in the Advanced Interactive Multidimensional Modelling System (AIMMS) software. The CPLEX and CONOPT are the selected solvers. The most promising pathways for three types of microalgae are proposed. These have differences in the dewatering section. Furthermore, the different pathways are compared in terms of cost and performance. The results show that the Haematococcus Pluvialis biorefinery leads to the highest profits due to pigments products' high amount and price.

Keywords: Superstructure optimization; algae biorefinery; biochemical; MINLP; techno-economic analysis.

1. Introduction

Biomass has been considered a renewable feedstock to overcome the shortage of petroleum-based fuel sources and handle global warming. Microalgal biomass offers incredible possibilities to be used as feedstock for biochemical and bioenergy production compared to other biomass sources. Microalgae is a non-food biomass feedstock that grows very fast in many types of water (such as freshwater, saltwater, wastewater, etc.) (Gebreslassie et al., 2013).

Algae biomass is composed of pigments, lipids, proteins, and carbohydrates that can be converted into various products (de la Noue & de Pauw, 1988). There is a growing industrial interest in using microalgae for an extensive range of applications, including biofuels and bioenergy, biofertilizers, vitamins, and chemical compounds for food production, nutraceutical dietary supplements, cosmetics, and pharmaceutical products, etc. (Torres et al., 2021). Despite the vast potential to use microalgae as a feedstock for various industries, a technical challenge must be addressed to commercially extend the use of biochemicals and biofuels from algal biomass. The optimization of a superstructure is one approach to enhance the application of microalgae on a large scale by finding a cost-effective pathway.

Rizwan et al. (2015) formulated a superstructure as a mixed integer non-linear program (MINLP), optimizing the net present value (NPV) of an algae biorefinery. Although biodiesel, bio-oil, and biogas are produced in this biorefinery, the capital costs are not considered (Rizwan et al., 2015). Galanopoulos et al. (2019) proposed a superstructure for an integrated algae biorefinery to minimize the price of biodiesel. The total biodiesel costs can be decreased with 20 % by producing bioethanol, glycerol, and levulinic acid (Galanopoulos et al., 2019). Their study showed that the price of biodiesel could be decreased by producing added-value products. Still, the profits of this algae biorefinery are not high enough to scale it up to a commercial level. Furthermore, they considered only a Chlorella Vulgaris biorefinery. Including different types of microalgae with different compositions and investigating various bioproducts will increase the prospect of commercializing the algae biorefinery. For this reason, a superstructure that includes three types of microalgae is developed to optimize the production pathway of added value products such as pigments, biodiesel, biogas, glycerol, omega-3, fertilizers.

2. Methodology

2-1 Process description and superstructure development

By using carbon dioxide and wastewater, microalgae can be cultivated. Four technologies (open pond, flat plate photobioreactor, bubble column photobioreactor, turbo column photobioreactor) are available for this cultivation stage. Subsequently, microalgae are separated from water in harvesting (including sedimentation and flotation/filtration), dewatering (flocculation, centrifugation, filter press), and a drying section. Next, the cells are disrupted to extract pigments and various lipids. There are a number of technologies for cell disruption, such as bead beating, high-pressure homogenization, microwaving, sonication, and hydrothermal liquefaction. After cell disruption, the pigments, (which are the most expensive products) are extracted. This stage is commonly done by using organic solvents or supercritical carbon dioxide. The lipids are extracted with appropriate solvents (n-butanol, Hexane, supercritical carbon dioxide) and forwarded to the lipid production stage to produce omega-3, biodiesel, and glycerol. Finally, the remaining parts of the microalgae are transported to the remnant treatment section to produce biogas and biofertilizer. Based on the current technologies, different process pathways can be selected to produce added-value components and bioenergy. All the alternatives are considered in the superstructure, as shown in Figure 1. Each block represents one of the technologies mentioned before.

2-2 Problem statement

Given is a superstructure with all current technologies and pathways. The specifications of products and raw materials are extracted from the literature. Furthermore, the equipment data includes performance (split factors and yields), cost factors (CAPEX/OPEX, Lang factors, and interest rates). The superstructure is optimized under the condition that the mass and energy balance hold and that costs display in economy of scale. The decision to be made is to select one technology at each stage and to determine the mass and energy flows at each stage. Then, the cost-effective pathway and related

technologies are decided by optimizing this superstructure to maximize profits and minimize the cost-based as the objective function.

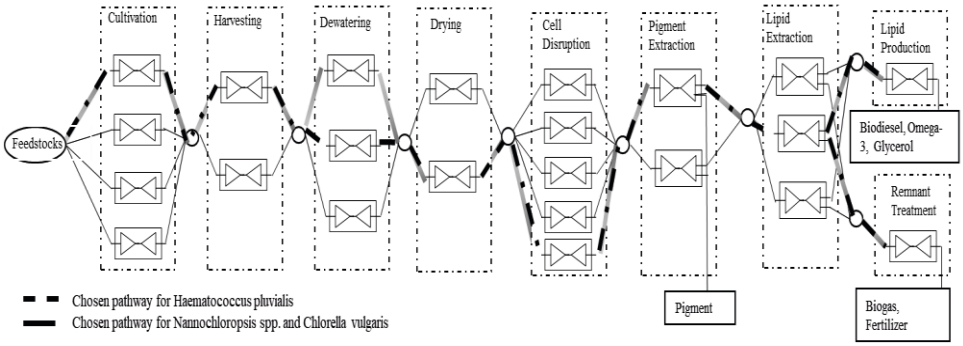


Figure 1: Superstructure of algae biorefinery and cost-effective production pathway for each type of algae

2-3 mathematical model

A mathematical model can be used to optimize the superstructure. This model contains an objective function and various constraints and variables (such as mass and energy balances and equipment limitations). The logical constraints are defined to allow only for the selected one option of each stage. There are nine intervals and 23 options in total (as shown in Figure1).

All flows that can enter/leave each option (j) are shown in figure 2, schematically. In the first part of each block, there is a mixing process to produce input flow (IN). The mass flow of mixing section ($m_{k.j}^{IN}$) for each component (k) is a mixture consisting of two parts, the upstream stream (U) mass flow ($m_{k,j}^{U}$) from the previous stage or feedstock (for four options of cultivation stage) and the reactant stream mass flow ($m_{k,j}^{R}$), which could be used to add solvents or reactants. The concentration factor $x_{k,j}$ is defined for calculating the reactant stream. It is a weight fraction based on the basic component k in the upstream flow. All these flows are added up in the Eq. (1).

$$m_{k,j}^{IN} = m_{k,j}^{U} + m_{k,j}^{R} = m_{k,j}^{U} + x_{k,j} \cdot m_{k,j}^{U} \tag{1}$$

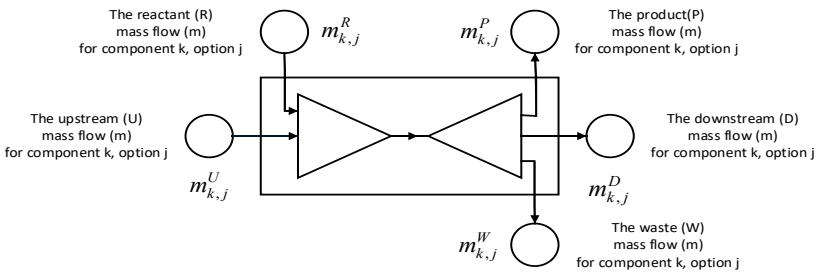


Figure 2. Mass balances in option J

To consider the reaction inside each block, the output (OUT) mass flow ($m_{k,j}^{OUT}$) (Eq. (2)) can be calculated either by a mass stoichiometric coefficient $S_{k,j}$ or by distribution coefficients $D_{k,j}$ in the case of remnant treatment. If no reaction or distribution takes place inside the option, the outlet flow should equal the inlet flow.

$$m_{k,j}^{OUT} = m_{k,j}^{IN} + S_{k,j} \cdot m_{k,j}^{IN} + D_{k,j} \cdot m_{k,j}^{IN} \quad (2)$$

Then the output mass flow ($m_{k,j}^{OUT}$) can be separated into three streams: A downstream (D) mass flow ($m_{k,j}^{D}$) going to the next stage, a waste(w) flow ($m_{k,j)}^{W}$, and products(P) flows ($m_{k,j)}^{P}$. These streams are calculated by using split factors $SF_{k,j}$as shown in Eq. (3).

$$m_{k,j}^{OUT} = m_{k,j}^{D} + m_{k,j}^{W} + m_{k,j}^{P} = SF_{k,j}^{D} \cdot m_{k,j}^{OUT} + SF_{k,j}^{W} \cdot m_{k,j}^{OUT} + SF_{k,j}^{P} \cdot m_{k,j}^{OUT} \quad (3)$$

For the energy balances, three elements (electricity (U_j^E), heating (U_j^H), and cooling (U_j^C)) are considered. For the utility consumption of each option (U_j), it is assumed that the energy demand is proportional to the total inlet stream mass flow going through the option. (SUC_j) is the specific utility consumption factor (Eq. (4)).

$$U_j = \sum_k m_{k,j}^{IN} \cdot SUC_j \quad (4)$$

The profit margin is defined as the difference of the annualized investment cost (AIC), annualized operating cost (AOC), and the product sales (PS).The AIC are calculated from the total plant installation cost ($TIPC$), the interest rate (IR), and the lifetime (LT), as shown in Eq. (5). The $TIPC$ can be calculated from the equipment cost with an engineering coefficient (K^{ENG}) and the land cost (LC_j) for the cultivation stage, which is shown in Eq. (6). The equipment cost is calculated using the economy-of-scale principle (f_j), with a reference cost(E_j^{ref}), a reference mass flow (m_j^{ref}), a cost index in 2020 (IDX_j^{2020}) and a reference cost index (IDX_j^{ref}). The land cost LC_j is calculated with the land price (P^{Land}) and the productivity for algae cultivation ($Productivity$) by Eq. (7).

$$AIC = TIPC \cdot \frac{IR \cdot (IR+1)^{LT}}{(IR+1)^{LT}-1} \quad (5)$$

$$TIPC = K^{ENG} \cdot \sum_j EC_j^{ref} \cdot \left(\frac{\sum_k m_{j,k}^{IN}}{m_j^{ref}}\right)^{f_j} \cdot \left(\frac{IDX_j^{2020}}{IDX_j^{ref}}\right) + LC_j \quad (6)$$

$$LC_j = P^{Land} \cdot \frac{m_{Algae,1-4}^{OUT}}{Productivity_{Algae,1-4}} \quad (7)$$

The annualized operating cost includes the raw material cost (RMC), the utility cost (UC), the operating and maintenance cost (OMC), and the waste treatment cost (WTC), which are presented in Eq. (8). The RMC and UC are calculated from multiplying the operating hours per year and the material prices and utility price, respectively. The operating and maintenance cost (OMC) are calculated from multiplying the operating and maintenance factor and the AIC. The waste treatment cost (WTC) is linear to the waste stream mass flow with a price for waste treatment.

$$AOC = RMC + UC + OMC + WTC \quad (8)$$

The product sales are calculated using the product prices ($P^{Product}$) , the operating hours per year (H) and the total product mass flow as shown in Eq. (9).

$$PS = H \cdot \sum_p P_p^{Product} \cdot \sum_j \sum_k m_{k,j}^{P} \quad (9)$$

To decrease the number of variables, parameters, and constraints and to relax the model a block integration is generated in this study. With this approach, the whole process of each technology would be considered as one integrated option with one data set. The

block integration parameters are calculated in advance based on the parameters for each sub process within the option, ensuring that there is only one series of data for each option.

3. Results

Three types of microalgae (Chlorella Vulgaris (B. Wang et al., 2008), Nannochloropsis spp. (X. Wang et al., 2017), and Haematococcus Pluvialis (Ba et al., 2016)) are considered in this study. These algae grow in influent wastewater in the Netherlands. The required carbon for growing is prepared with pure carbon dioxide gas. Since daylight hours are another factor that influences algae growth, an average of 12hr sunlight per day is assumed in this study.

The Advanced Interactive Multidimensional Modelling (AIMMS) software version 4.82.3.29 64-bit is used to set up a mixed-integer non-linear programming (MINLP) model. It is solved with the Outer Approximation Algorithm (AOA) that consists of the CONOPT 4.1 solver for the non-linear part and the CPLEX 20.1 solver for a mixed-integer part. Furthermore, the model contains 6710 variables, 23 integers variables, and 6161 constraints.

For the Chlorella Vulgaris and the Nannochloropsis spp., the open pond, sedimentation and flotation, flocculation, without a dryer, hydrothermal liquefaction, organic solvent pigment extraction, N-butanol lipid extraction, lipid production, and anaerobic digestion are selected as the most cost-effective pathway (as shown in Figure 1). The optimal process pathway for Haematococcus Pluvialis is different only in the dewatering section. Centrifugation is chosen for this step in the biorefinery. Based on the productivity and cultivation reaction of Haematococcus Pluvialis, the amount of this algae is higher than other types, and it is not economically beneficial to use flocculant for separation water from them.

The Haematococcus Pluvialis biorefinery has the highest profit due to the high amount of pigments. During one year, 1 Mt of influent wastewater and 2 Mt of carbon dioxide are approximately used. 0.7 Kt of pigment and 3 t of Omega-3 can be produced. The daily profit margin of Haematococcus Pluvialis biorefinery is 28 and 34 times higher than Chlorella Vulgaris and Nannochloropsis spp. biorefineries, respectively.

The pigment is one of the expensive bioproducts. Depending on pigment composition, its price is about 2500-7000 ($/t) (Panis & Carreon, 2016). The amount of pigment that can be produced in a Haematococcus Pluvialis biorefinery are 5 and 28 times higher than Chlorella Vulgaris biorefinery and Nannochloropsis spp. biorefinery, respectively. The annual profits Haematococcus Pluvialis biorefinery for this bioproduct is approximately 200M$.

To validate the model, the superstructure is simplified to produce only biodiesel. Furthermore, one common microalgae (Chlorella Vulgaris) with 25 % lipid composition are studied in this comparison. The results are (an estimated biodiesel price of 5.2 $/L) in good agreement with data found in(Davis et al., 2011)who report biodiesel prices of 2.6$/L, as well as data from (Richardson et al., 2021), who found biodiesel prices of 9.2 $/L.

As recycles were not included and an open pond was considered for cultivation, the investment costs are relatively low. To separate large amounts of water, 31 % of total investment costs are attributed to harvesting and dewatering stages. The cell disruption stage is the most expensive part due to disrupting a massive amount of algae (39 % of

total investment costs). In addition, the operating costs contribute to approximately 83% of the total costs. Utilities are about 50 % of the total operating costs.

4. Conclusion

A superstructure of an algae biorefinery is developed to produce added-value products from microalgae (Haematococcus Pluvialis, Chlorella Vulgaris, Nannochloropsis spp). These superstructures are optimized in the AIMMS to find cost-effective production pathways. The optimal pathways consist of an open pond, sedimentation and flotation, flocculation/centrifugation, without a dryer, hydrothermal liquefaction, organic solvent pigment extraction, N-butanol lipid extraction, lipid production, and anaerobic digestion. Types of microalgae have an important role in finding the appropriate technology for the dewatering step (flocculation/centrifugation). The profit of Haematococcus Pluvialis biorefinery is more than 28 times higher that of the Chlorella Vulgaris biorefinery and more than 32 times higher than that of Nannochloropsis spp biorefinery. Haematococcus Pluvialis can produce 0.7 Kt of pigment and 3 t of Omega-3 using 1 Mt of influent wastewater and 2 Mt of carbon dioxide.

References

F. Ba, A. V. Ursu, C. Laroche, G. Djelveh, 2016, Haematococcus pluvialis soluble proteins: Extraction, characterization, concentration/fractionation and emulsifying properties. Bioresource Technology, 200, 147–152. https://doi.org/10.1016/j.biortech.2015.10.012

R. Davis, A. Aden, P. T. Pienkos, 2011, Techno-economic analysis of autotrophic microalgae for fuel production. Applied Energy, 88, 3524–3531. https://doi.org/10.1016/j.apenergy.2011.04.018

J. de la Noue, N. de Pauw, 1988 , The potential of microalgal biotechnology: A review of production and uses of microalgae. Biotechnology Advances, 6(4), 725–770. https://doi.org/10.1016/0734-9750(88)91921-0

C. Galanopoulos, P. Kenkel, E. Zondervan, 2019, Superstructure optimization of an integrated algae biorefinery. Computers & Chemical Engineering, 130, 106530. https://doi.org/10.1016/j.compchemeng.2019.106530

G. Panis, J. R. Carreon, 2016, Commercial astaxanthin production derived by green alga Haematococcus pluvialis : A microalgae process model and a techno-economic assessment all through production line. Algal Research, 18, 175–190. https://doi.org/10.1016/j.algal.2016.06.007

M. Rizwan, J. H. Lee, R. Gani, 2015, Optimal design of microalgae-based biorefinery: Economics, opportunities and challenges. Applied Energy, 150, 69–79.https://doi.org/10.1016/j.apenergy.2015.04.018

G. F. Torres, B. P. Elisabeth, J. Pittman, C. Theodoropoulos, 2021 Microalgae strain catalogue: A strain selection guide for microalgae users: cultivation and chemical characteristics for high added-value products. The University of Manchester. https://doi.org/10.5281/zenodo.3780067

B. Wang, Y. Li, N. Wu, C. Q. Lan, 2008, CO2 bio-mitigation using microalgae. Applied Microbiology and Biotechnology, 79(5), 707–718. https://doi.org/10.1007/s00253-008-1518-y

X. Wang, L. Sheng, X. Yang, 2017, Pyrolysis characteristics and pathways of protein, lipid and carbohydrate isolated from microalgae Nannochloropsis sp. Bioresource Technology, 229, 119–125. https://doi.org/10.1016/j.biortech.2017.01.018

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50047-6

Process simulation of continuous biodiesel production catalysed by a high stability solid in a reactive distillation

Chatchan Treeyawetchakul[a,*]

[a] *Department of Chemical Engineering, Mahanakorn University of Technology, Bangkok 10530, Thailand*
chatchan@mut.ac.th

Abstract

KF/(Ca/Al) catalyst developed in laboratory robust activity and stable than a conventional CaO catalyst for biodiesel production from refined palm oil. The highest conversion yield of 94.7% wt. obtained by employing KF/(Ca/Al) catalyst (10% wt.) in a 350 ml batch reactor. It was operated at methanol to oil molar ratio of 15:1, reaction temperature of 65°C, and reaction time of 3 hrs, the pseudo-first-order rate law could be used to fit the palm oil transesterification reaction (WongSree et al., 2016). In order to produce biodiesel in industrial scale and intensification, in this work, approximately 1,050 kg hr-1 of biodiesel production rate was considerably basis, and the transesterification reaction was occurred in the reactive distillation (RD) column as shown in Figure 1. The thermophysical parameters of all the components were computed and validated with available experimental data with reasonable accuracy. In addition, the kinetic model obtained from laboratory experiment was also used and modelled in the RD column simulation. The optimal conditions, with a maximum conversion of approximately 90% wt., are a methanol to oil molar ratio of 5:1, a reflux ratio of 0.2, a total number of trays of 15, a reboiler heat duty 20 kW, and the number of reactive trays should not be less than 7.

Keywords: Biodiesel production; KF/(Ca/Al) catalyst; Reactive distillation; Process Simulation.

1. Introduction

Biodiesel from palm oil has been promising and continuously supported by the Thai government since 2010 as an alternative and sustainable energy. Generally, biodiesel is produced from the transesterification of crude palm oil (CPO) or palm stearin with short-chain alcohol (methanol or ethanol) involving homogeneous catalyst (KOH or NaOH) under appropriate conditions and reaction time. However, the use of homogeneous catalysts leads to the continuous catalyst consuming reaction thus reducing the catalytic efficiency over the reaction period. In addition, it is technically difficult to remove unreacted catalyst after the reaction completion and a large amount of wastewater is produced to and it needs to be separated and cleaned the products, which increases the overall cost of the process. Thus, the biodiesel production cost based on homogeneous catalysis, is not yet sufficiently competitive as compared to the cost of diesel production from petroleum (Zhang, 2003).

The development of heterogeneous catalysts is an alternative choice that could eliminate the additional operation costs associated with the aforementioned (separation and

purification step). Furthermore, biodiesel production with heterogeneous catalysts does not produce soap as a by-product. Therefore, the study of heterogeneous catalysts has been focused on by several researchers since it leads to a possibility of another pathway for biodiesel production development.

WongSree et al. (2016) developed and studied the effects of KF/(Ca/Al) on biodiesel production from purified palm oil compared with CaO and KF/CaO catalysts on the laboratory scale. They concluded that the highest oil conversion of 94.7% could be achieved by employing the KF/(Ca/Al) catalyst in a batch reactor together with methanol to oil molar ratio 15:1, catalyst loading of 10 wt.%, reaction temperatures of 65 ^{o}C, and reaction time 3 hours. They also proposed that the kinetic model of the transesterification could be explained and fitted with the experiments by the pseudo-first order model as shown in Eq. (1)

$$\frac{dc_A}{dt} = \left[5.1 \times 10^6 \exp\left(\frac{58.47}{RT} \right) \right] c_A \tag{1}$$

Where c_A is concentration of tripalmitin

The aim of this study is in preliminary proposing and applying an RD column simulation of the biodiesel production process via Aspen Plus® as a simulator by using the aforementioned laboratory results as basic parameters for industrial production.

2. Description of RD and methodology

For simulation experiments, based on 1,000 kghr^{-1} biodiesel production rates, the CPO was fed approximately 1,167 kghr^{-1} at 25 ^{o}C, 1 atm. The RD column for the biodiesel production process consists of three sections: rectification, reaction, and stripping as shown in Fig. 1. In the reaction stage, the transesterification reaction scheme is:

$$C_{51}H_{98}O_6 + CH_3OH \rightarrow 3C_{17}H_{34}O_2 + C_3H_8O_3 \tag{2}$$

$$C_{57}H_{104}O_6 + CH_3OH \rightarrow 3C_{19}H_{36}O_2 + C_3H_8O_3 \tag{3}$$

In the reaction zone, the pseudo-first order kinetic model as shown in eq.1 was used in the simulation. For this case, methanol excess was released at the top of the RD column while both glycerol and methyl palmitate were separated via decanter.

The Aspen Plus® tool was used for simulation of this process. Regarding the CPO as raw material which is high content of triolein ($C_{57}H_{104}O_6$) and tripalmitin ($C_{51}H_{98}O_6$), then both of them were represented in this simulation. Accordingly, methyl oleate ($C_{19}H_{36}O_2$) and methyl palmitate ($C_{17}H_{34}O_2$) were taken as the FAME products and their properties were available in the library of the Aspen tool. Due to the highly polar components presenting in this process such as methanol and glycerol, the UNIF-DMD and NRTL thermodynamic/activity model were used to estimate the activity coefficients in a liquid phase.

The important parameters such as molar reflux ratio, column pressure, number of the column tray, and so on would be adjusted in order to get highest biodiesel yield and lowest RD column duty.

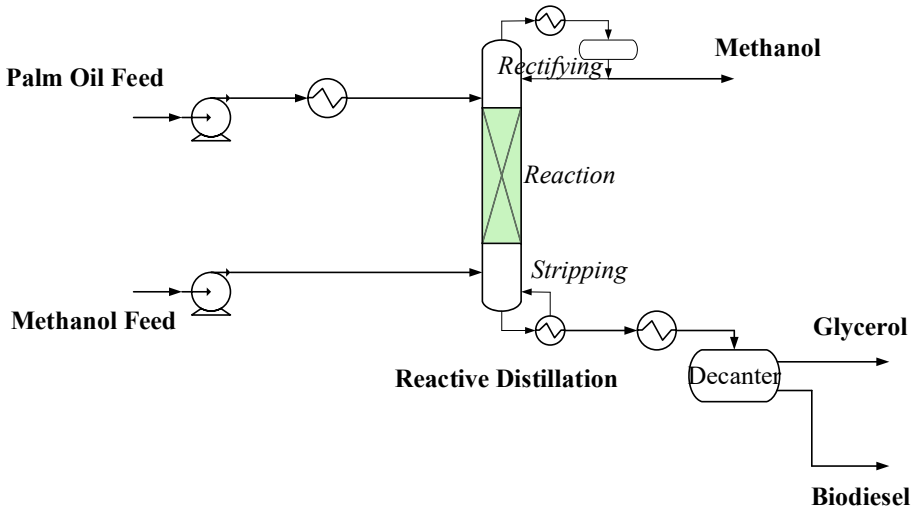


Figure 1. Biodiesel production with RD column

3. Results and Discussion

3.1 Effects of operating pressure and molar reflux ratio

At steady-state simulation, with the 15-theoretical stage of the RD column, the effect of column pressure and molar reflux ratio were shown in Fig. 2 and 3, respectively. Fig. 2 shows that the reboiler duty of the RD column continuously increases with operating pressure increasing and it is the same trend for molar reflux ratio (Fig. 3). It can also be interpreted that the column pressure 0.2 bar and molar reflux ratio 0.5 are optimal operating conditions.

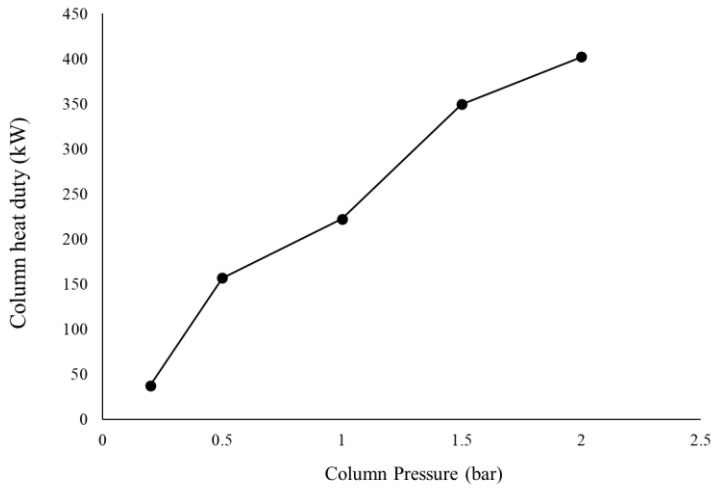


Figure 2. Effect of column pressure on the column duty

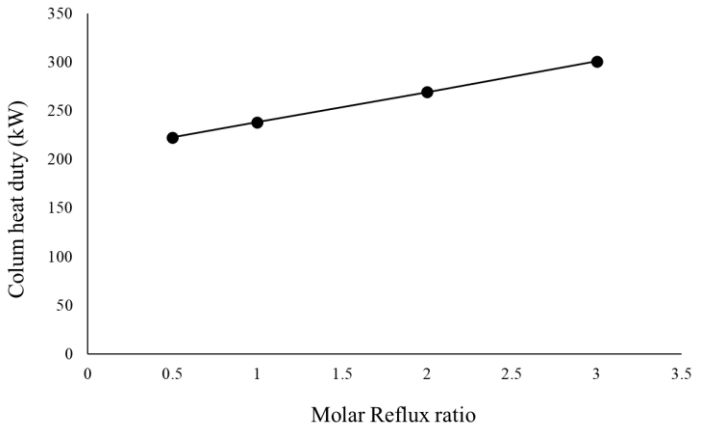


Figure 3. Effect of molar reflux ratio on the column duty

Besides the duty of the column considering, the proper pressure reduction (0.2 bar) was also kept the reboiler temperature lower than 200 ºC and condenser temperature higher than 20 ºC as shown in Fig. 4. Furthermore, at this pressure, the reaction zone temperature was kept between 60-80 ºC which corresponds to the experiments (WongSree et al., 2016). However, the temperature at the 12th tray (methanol feed tray) is rapidly decrease which is different from Karacan and F Karacanb (2014) have reported. This can be explained that this study used methanol fed at the room temperature (25 ºC) without preheater.

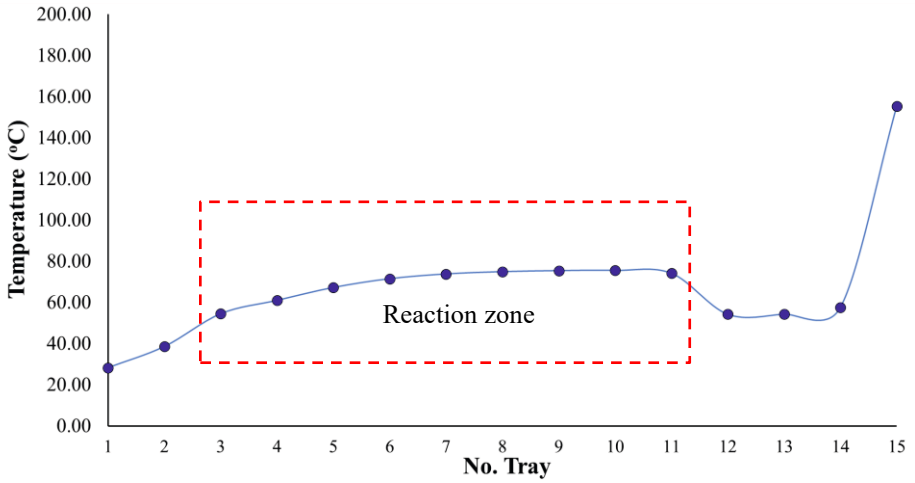


Figure 4. Temperature profiles of the RD column for the operating pressure 0.2 bar

3.2 Effects of methanol to oil molar ratio as feed

At the preliminary optimal conditions of RD column (reflux ratio =3.0, operating pressure 0.2 bar, methanol to oil ratio = 5:1, number of total stages = 15, and number of the reaction stage = 8), the biodiesel product (mixture of methyl oleate and methyl palmitate) was achieved at the flow rate of 1,050 $kghr^{-1}$ with the highest yield at 0.903 as shown in Fig. 5. It can be seen that the yield of biodiesel production from the simulation was closed to the experiments.

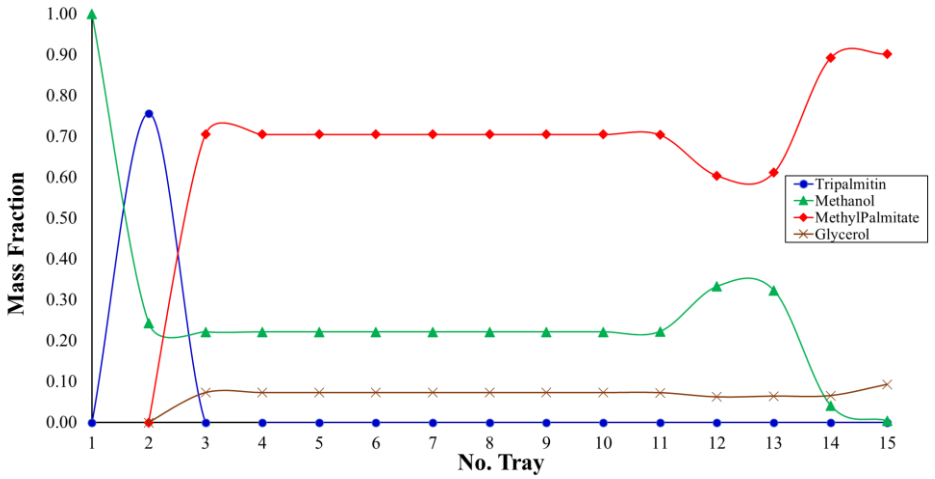


Figure 5. Liquid compositions profiles of the RD column at optimal condition (methanol to oil molar ratio = 5:1)

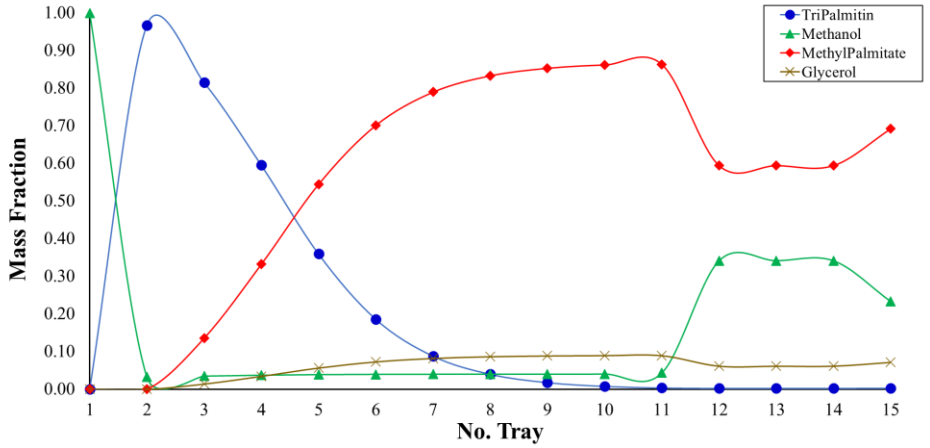


Figure 6. Liquid compositions profiles of the RD column at optimal condition (methanol to oil molar ratio = 15:1)

Fig.5 shows the concentration distribution of all chemicals within RD column. The figure shows that, with the oil feed stage is 2, the rapid reaction rates are occurring on stages 3-4 while the transesterification is occurred continuously between stage 4 and stage 11. Then the biodiesel produced along with glycerol flowed downward to the bottom of the RD column.

However, at the same biodiesel yield, in an experiment, the methanol to oil molar ratio should be 15:1, while this ratio from the simulation is only 5:1. This can be explained that when a large amount of methanol is fed at stage 12, it would be affecting on the bottom temperature to be lower than 150 ºC (approximately 80-100 ºC). Compared with methanol boiling point temperature (approximately 65 ºC), some methanol might not be vaporized to the above stage and poured and mixed with biodiesel product. This phenomenon was proved in Fig. 6.

4. Conclusions

In this paper, the biodiesel production catalysed by KF(Al/CO) was simulated by Aspen Plus with the biodiesel production rate at 1,050 $kghr^{-1}$. The simulation results show that, the biodiesel yields is approximately 90.3% which closed to the experimental result at optimal conditions (molar reflux ratio =3.0, operating pressure 0.2 bar, methanol to oil ratio = 5:1, number of total stages = 15, and number of the reaction stage = 8), and the RD column heat duty is around 50 kW. Then these preliminary results show that this modified process can be interesting to both of energy saving concern.

5. References

N. WongSree, S. Dokmai and T. Prachoomphon, 2016, Continuous production of biodiesel in a reactive distillation catalyzed by a high stability solid catalyst, B.Eng. Thesis, Chemical Engineering Department, Mahanakorn University of Technology, Thailand.

S. Karacan and F Karacanb, 2014, Simulation of Reactive Distillation Column for Biodiesel Production at Optimum Conditions, Chemical Engineering Transactions, 39, 1705-1710

X. Yang, L. Haoyang; X. Guomin, G. Lijing, and P. Xiaomei, 2013, Simulation of the catalytic reactive distillation process for biodiesel production via transesterification, IEEE International Conference on Materials for Renewable Energy and Environment, Chengdu, China. DOI:10.1109/ICMREE.2013.6893646

Y. Zhang, M.A. Dubé, D.D McLean and M.Kates, 2003, Biodiesel production from waste cooking oil: 1. Process design and technological assessment, Bioresource Technology, 89, (1), 1-16

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50048-8

Generative Approaches for the Synthesis of Process Structures

Tahar NABIL[a], Jean-Marc COMMENGE[c], and Thibaut NEVEUX[b,c*]

[a] *EDF R&D, Boulevard Gaspard Monge, F-91120 Palaiseau, France*
[b] *EDF R&D, 6 quai Watier, F-78401 Chatou, France*
[c] *Université de Lorraine, CNRS, LRGP, F-54000 Nancy, France*
thibaut.neveux@edf.fr

Abstract

In process synthesis, generative approaches are algorithmic strategies able to produce new structures, which differs from conventional optimization techniques consisting in choosing among a predetermined set of structures (e.g. heuristics and superstructure optimization). The development of these approaches has only intensified recently with the rise of both evolutionary computation and machine learning techniques. This paper aims at introducing some recent experiments, categorized into reward-driven and data-driven algorithms; and discussing key aspects of the generative steps such as: required initial database, process data representation, generative model architecture, reward design, optimization strategy and post-processing for the engineer.

Keywords: Process Synthesis; Machine Learning; Artificial Intelligence; Evolutionary Programming; Optimization

1. Introduction

The essence of process synthesis implies to propose a process structure, i.e. a set of unit operations for the transformation of mass and energy interconnected in a network (process flowsheet), associated with degrees of freedom such as equipment design and operating conditions. Solutions to solve this problem evolved from heuristics and expertise, through iterations with process simulators and experimental works, towards mathematical optimization techniques. A common approach consists in optimizing the process structure among a postulated set of alternative paths, called superstructure, using optimization (Mixed Integer Non-Linear Programming, Generalized Disjunctive Programming, etc.) to select the best structures with respect to given objective(s) and constraints. These approaches are proving to be very complex to implement in real industrial cases and intrinsically introduce an inductive bias due to the restriction to a search space limited by the defined superstructure: if the optimal structure is not in the defined space, the optimizer cannot find it.

Recently, generative approaches have used advanced algorithms to propose process structures directly from a set of available unit operations. These approaches differ from conventional optimization techniques by their capacity to generate new structures, instead of choosing among a predetermined set of structures. This paper aims at introducing some of those recent experiments, discussing the key aspects of the generative steps, and highlighting some remaining challenges.

2. Recent prospective works on "Generative Approaches"

The idea of using algorithmic strategies to generate process structures is not new (Nishida et al., 1981), but its development has only intensified recently with the rise of both evolutionary computation and machine learning techniques (see examples in Figure 1). Yet, most of the works on generative approaches remain prospective with academic proof-of-concept and/or specific case-studies. We provide here a non-exhaustive overview of the various techniques proposed to generate process structures and highlight their key features.

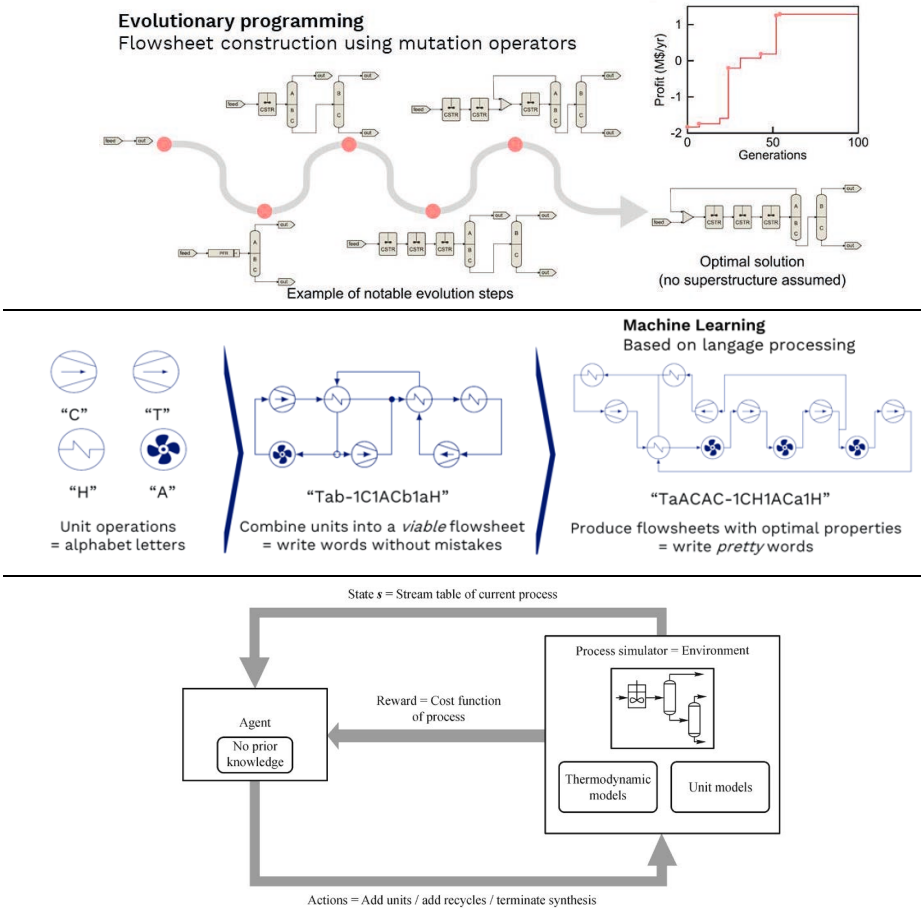


Figure 1. Examples of generative approaches to build flowsheets: evolutionary programming using mutation operators (top), machine learning using language processing (middle), and reinforcement learning using a two-player game (bottom); respectively adapted from Neveux (2018), Nabil *et al.* (2019), and Göttl *et al.* (2022).

Generative approaches usually rely on a two-level decomposition of the process synthesis problem, an upper level for designing the topology (i.e. the process flowsheet) and a lower level for evaluating the flowsheet (i.e. optimizing the degrees of freedom for a given structure). In this paper, we discuss only the topological problem, i.e. the

generation of process structures. The more specific task of generating structures with certain desired properties is called targeted process generation; it consists in producing flowsheets which will maximize a given fitness function. Some recent contributions in the field of process synthesis suggest using artificial intelligence algorithms to generate new flowsheets. They could be categorized into:

- Data-driven algorithms, learning from an initial database of known process structures and their performance indicators to generate new structures (see Figure 1, middle). This category is based on Machine Learning and Deep Learning techniques. For instance, Nabil *et al.* (2019) used a string representation of a power cycle and applied Natural Language Processing to generate new structures.
- Reward-driven algorithms, performing a topological optimization driven by a set of objectives and constraints. This category includes Evolutionary algorithms (EA) and Reinforcement Learning (RL) models. For instance, EA could apply mutation operators to modify a population of process structures (Figure 1, top); operators could be elemental such as unit addition, unit removal, stream permutation (Neveux, 2018) or hierarchical with function and technology levels (Wang *et al.,* 2015); while RL techniques (Figure 1, bottom) maximize a reward by sequentially modifying the structure of the process. For instance, Göttl *et al.* (2022) formulated the problem as a turn-based two-player game and applied RL for ethyl tert-butil ether synthesis. Midgley (2020) used RL with deep neural networks to optimize a sequence of reactors and a distillation column train.

Process design could also benefit from initiatives in other research fields with similar data representation and network generation problems. In particular, machine learning generative models have obtained significant successes in the field of molecular design (Elton et al., 2019), and their approaches could be adapted to processes.

3. Discussions on generative aspects

In this section, we formalize the notion of a generative machine learning model and study how to apply it to process synthesis, highlighting some key challenges.

3.1 Generative machine learning models

Consider a set of observed data $\mathcal{D}$. We assume that each element in $\mathcal{D}$ is a sample from an underlying -unknown- data distribution p^*. The goal of a generative machine learning model is to learn a parametric approximation p_θ of p^*, where the model distribution p_θ belongs to a model class $\mathcal{M} = \{p_\theta : \theta \in \Theta\}$, parameterized by a vector θ from the model family Θ. The following optimization problem is thus solved: $\min_{\theta \in \Theta} d(p^*, p_\theta)$, with $d(\cdot)$ a distance between probability distributions. Once it is learned, one may sample from the model distribution to generate new elements: $x_{new} \sim p_\theta(\cdot)$. Hence, one advantage of generative models is that they can produce (infinitely many) new samples, extending $\mathcal{D}$.

3.2 Application to process synthesis: concept and challenges

Applied to process synthesis, e.g. in (Nabil et al., 2019), the generative approach starts by gathering an initial database of flowsheets, determining the set of unit operations to include and their numerical representation. Next, the chosen model architecture is optimized to represent the underlying data distribution and maximize a certain fitness

function for targeted generation. The outcome is a large pool of flowsheets with good properties, to be analysed by the design engineer. We call this approach data-driven.

Another complementary approach is also emerging, namely a purely reward-driven optimization including evolutionary programming (Neveux, 2018) and reinforcement learning models (Göttl *et al.*, 2022, Midgley, 2020). In practice, the two approaches can be combined to obtain better performances. For instance, reinforcement learning can accelerate the convergence of the generative model towards good regions of the search space, as in (Olivecrona *et al.*, 2017) for molecules.

In the sequel, we discuss several practical challenges that need to be addressed in order to apply successfully the machine learning approach to process synthesis.

3.2.1 Initial database

The first task is to collect data, i.e. a set $\mathcal{D} = \{x_i, y_i\}_{i=1}^{n}$, where each process structure x_i is associated with a label y_i, e.g. the economic cost, for a large integer n ~ 10^6. Whereas such datasets already exist for molecule synthesis (Elton et al., 2019), it is not the case for processes. We suggest three ways to overcome this issue. Firstly, one can generate processes as random coloured graphs - it was shown by Nabil et al. (2019) that starting with a small-size set of random layouts could yield promising results by iteratively augmenting the training set with the generated data. Secondly, former expert knowledge can be exploited by listing every layout included in a given superstructure. The third option is to re-use artificial samples generated by another method, e.g. reward-driven methods not requiring an initial database (Neveux, 2018).

Besides, each label y_i is not an intrinsic property of the layout x_i but depends instead on certain boundary conditions. y_i is usually found by solving an optimization problem whose complexity depends on the physics of the process. This adds to the computational burden of generative models, since for each new use case with specific boundary conditions, the label y_i should be computed again. Hence, a typical dataset of processes might be sparse (unlabelled x_i's) and noisy or heterogeneous (y_i's obtained from different optimization algorithms). Finally, one last challenge is to create a dataset that is not trivially separable on process structures while covering a large, non-local, fraction of the search space.

3.2.2 Data representation

Generative approaches rely on an abstract representation of a process structure to be processed by a machine learning algorithm, and it remains an open question to determine which format is most suitable for process structures. Two formats stand out particularly, namely a graph-based or a string-based representation. Both formats are bijective and can benefit from advances in artificial intelligence on graphs and Natural Language Processing, respectively. For instance, the underlying representation used by Neveux (2018) or Göttl *et al.* (2022) is a graph, whereas Nabil *et al.* (2019) explicitly created a language with an alphabet and syntactic rules to represent a power cycle. Language encoding could in the future exploit the more generic SFILES (simplified flowsheet-input line-entry system) format (d'Anterroches and Gani, 2005). On the other hand, a graph-based representation, typically a directed graph with node colouring, is closer to the flowsheets known to chemical engineers and removes the need to learn syntactic rules. However, graph generative models have a greater computational cost, since graph isomorphism is not solvable in polynomial time.

It is interesting to note that generative models for molecules were historically based on molecular fingerprints, then strings (so-called SMILES format), and the recent trend is to shift towards graph representations (Elton et al., 2019).

3.2.3 Generative model architecture

Once the observed data $\mathcal{D}$ is available and the data representation chosen, it remains to determine the architecture of the generative model, i.e. to select a class of models $\mathcal{M}$. The deep learning literature is rich of such generative models, which have achieved state-of-the-art performances. The three main classes are the recurrent neural networks (RNNs), the variational autoencoders (VAEs) and the generative adversarial networks (GANs). The field of graph or string generative models is also driven by the application to molecule synthesis, which constitutes thus an excellent first set of architectures to adapt to process synthesis: see e.g. Elton et al. (2019).

3.2.4 Reward design

In targeted process generation, the generative model is biased towards a certain region of the search space by a reward function. The reward should be designed carefully to obtain processes with desirable features such as (i) diversity and novelty (ii) feasibility and respect of the constraints and (iii) high fitness. Diversity and novelty metrics can be found in (Elton et al., 2019). The feasibility and constraints are to be defined on a per use-case basis. In particular, the performances of generative models might decrease whenever the problem is severely constrained, reducing thus the space of valid flowsheets. As an example, if 99% product purity is expected, a process reaching 90% is not valid yet better than 0%; the reward should therefore encode this designer expertise, e.g. by penalizing the objective function (such as an economic cost) by the constraints violation (distance from 99% purity). Hence, softening the constraints by adding adequate penalty terms in the reward function can facilitate the convergence of the model. See also (Göttl *et al.*, 2022) for a discussion on how the reward function might impact the diversity of the generated flowsheets. Finally, the ability to evaluate the fitness of a flowsheet is a nonlinear optimization problem itself. Since generative models are data-intensive and require the evaluation of thousands of flowsheets, efficient evaluation is necessary for the application to real use cases. A promising track to alleviate this computational cost is to develop surrogate meta-models (Gorissen et al., 2010) or train a machine learning regression or classification model on separate data (Nabil et al., 2019).

3.2.5 Optimization strategy

Reward-driven methods are optimization algorithms, as such they are explicitly designed for targeted process generation. Machine learning generative models can also produce focused libraries of processes, although in a less straightforward manner. For instance, transfer learning is used by Nabil et al. (2019) to bias the RNN network towards regions of high efficiency and shaft power output for power cycles while reinforcement learning is combined with generative models by Olivecrona et al. (2017). Whereas the space of process flowsheets is discrete and large, another approach is to take advantage of the latent space built by VAEs or GANs, enabling thus the optimization in a low-dimensional continuous space (Gómez-Bombarelli et al., 2018).

In the future, the ease of the optimization process, for instance the sensibility of the model to hyperparameter tuning, should also be a criterion for discriminating between

models, in particular between string-based and graph-based approaches: robust models adaptable to new use cases are preferable for the adoption by the process engineer.

3.2.6 Post-processing for the engineer

The finality of generative models is to produce pools of relevant flowsheets to help the process engineer at the design stage. Beyond the selection of the best topology achieved by the model, it is also beneficial to gain knowledge from the generated data, to extract new heuristic rules. For instance, Zhang et al. (2019) propose a methodology to compare process flowsheets and detect structural similarities between them, by applying text pattern mining algorithms to SFILES.

4. Conclusive remarks

Applying generative approaches for the synthesis of process structures is a new field, with various recent experiments using evolutionary and machine learning techniques. They have been tested on a limited number of use cases, which calls for further investigation to evaluate their most appropriate use and define practical tips for the process engineer. The search space being virtually infinite, the computation effort could become prohibitive. Therefore, there is a need both for results reproducibility and for a comprehensive comparison on identical synthesis problems, to assess the potential and drawbacks of these techniques; whether they could substitute to established techniques (such as heuristics or superstructure optimization) or serve as new search heuristics to better define a restricted search space for conventional synthesis approaches.

References

d'Anterroches, L. and Gani, R. (2005), Group contribution based process flowsheet synthesis, design and modelling. *Fluid Phase Equilibria,* 228:141-146.

Elton, D. C., Boukouvalas, Z., Fuge, M. D. and Chung, P. W. (2019), Deep learning for molecular design—a review of the state of the art. *Mol. Syst. Des. Eng.,* 4:828-849.

Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernández-Lobato, J. M., Sánchez-Lengeling, B., Sheberla, D., Aguilera-Iparraguirre, J., Hirzel, T. D., Adams, R. P. and Aspuru-Guzik, A. (2018), Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Cent. Sci*., 4:268-276.

Gorissen, D., Couckuyt, I., Demeester, P., Dhaene, T. and Crombecq, K. (2010), A surrogate modeling and adaptive sampling toolbox for computer based design. *J. Mach. Learn. Res.*

Göttl, Q., Grimm, D.G. and Burger (2022), J. Automated synthesis of steady-state continuous processes using reinforcement learning. *Front. Chem. Sci. Eng*. 16, 288–302.

Midgley, L. I., (2020), Deep Reinforcement Learning for Process Synthesis, *Preprint arXiv:2009.13265*

Nabil, T. and Le Moullec, Y. and Le Coz, A., (2019), Machine Learning based design of a supercritical CO2 concentrating solar power plant, 3rd European scCO2 Conference, Paris.

Neveux, T., (2018), Ab-initio process synthesis using evolutionary programming, *Chemical Engineering Science,* 185:209–22.

Nishida, N., Stephanopoulos, G. and Westerberg, A.W. (1981), A review of process synthesis. *AIChE Journal*, 27(3):321-351.

Olivecrona, M., Blaschke, T., Engkvist, O. and Chen, H. (2017), Molecular de-novo design through deep reinforcement learning. *J. Cheminform,* 9:48

Wang, L., Voll, P., Lampe, M., Yang, Y. and Bardow, A. (2015), Superstructure-free synthesis and optimization of thermal power plants. *Energy*, 91:700-711.

Zhang, T., Sahinidis, N.V. and Siirola, J.J. (2019), Pattern recognition in chemical process flowsheets. *AIChE Journal*, 65: 592-603.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50049-X

Metal-Organic Framework Targeting for Optimal Pressure Swing Adsorption Processes

Xiang Zhang[a], Teng Zhou[a,b,*], Kai Sundmacher[a,b]

[a]*Process Systems Engineering, Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1, D-39106 Magdeburg, Germany*
[b]*Process Systems Engineering, Otto-von-Guericke University Magdeburg, Universitätsplatz 2, D-39106 Magdeburg, Germany*
zhout@mpi-magdeburg.mpg.de

Abstract

Gas separation accounts for a major production cost in chemical industries. So far, pressure swing adsorption (PSA) has been widely used for gas separation applications such as H_2 purification and CO_2 capture. For PSA processes, the adsorption efficiency is greatly affected by the selected adsorbent and process operating conditions. Over the past decade, porous metal-organic frameworks (MOFs) have been recognized as innovative adsorbents featuring tunable properties. For achieving a high separation efficiency, a novel two-step integrated MOF and PSA process design approach has been recently proposed. In the first step, MOF is represented as a set of geometric and chemical descriptors. The MOF descriptors and process operating conditions are simultaneously optimized to maximize the process performance. In this work, the second step, namely MOF targeting, is presented. The objective is to use various computational tools to synthesize hypothetical MOFs and identify potential candidates based on the optimized MOF descriptors. The involved computational tools include Tobacco for computational MOF synthesis, Poreblazer for geometry characterization, and RASPA for rigorous adsorption isotherm simulation.

Keywords: MOF targeting, Hypothetical MOF, Machine learning, Adsorption process design, Gas separation

1. Introduction

Pressure swing adsorption (PSA) has been widely used in chemical and energy industries for gas separation. It exploits different gas affinities on solid adsorbents at different pressures to achieve the separation. Currently, multiple types of porous materials are used as adsorbents such as metal-organic framework (MOF), zeolite, and activated carbon. Among them, MOFs are formed via the self-assembly of various molecular building blocks (i.e., metal nodes and organic linkers) in different topologies. Due to the large variety of building blocks, a near-infinite design space exists for MOFs (Yao et al. 2021). In addition, MOFs have many superior properties such as high porosity, tunable pore geometry, and functional pore surface. Therefore, MOFs have a great potential to enhance process efficiency in adsorption-based gas separation (Wang et al. 2020).

When MOFs are used for adsorption-based gas separation, four alternative mechanisms exist: equilibrium separation, kinetic separation, molecular sieving, and gate-opening separation. Among these, equilibrium separation is the most common method, which is

based on the difference in gas equilibrium loadings. So far, numerous MOFs with diverse isotherm characteristics have been synthesized in the laboratory by varying MOF chemistry and structure (e.g., pore geometry and topology). Unfortunately, this experimental trial-and-error approach is time-consuming and inefficient. Importantly, many MOFs cannot lead to good process performance (Burns et al. 2020). It is well-known that adsorbents ultimately serve a specific adsorption process. In this case, the design of a MOF for use in PSA processes is indeed a multiscale design problem that incorporates the inter-linked material, phase, and process levels. Variations of materials and process operating conditions affect the adsorption behavior and thus jointly decide the process performance. With this in mind, a computational approach is desired to expedite the identification of promising MOF adsorbents that can best serve the PSA process.

Recently, focusing on the equilibrium separation, our group has proposed a novel two-step integrated MOF and PSA process design framework to tackle the above challenges (Zhang et al. 2021). As shown in Figure 1, it consists of two steps: descriptor optimization and MOF targeting. Due to the large number of MOF building blocks, it is difficult to build mathematical models to predict adsorption isotherms directly from building blocks. Therefore, in the published first step, a MOF is represented by a set of chemical and geometry descriptors. Several data science techniques are utilized to select proper descriptors, define a valid design space, and build data-driven models for the prediction of adsorption isotherms. This enables an explicit formulation of the integrated MOF and PSA process design problem, where MOF descriptors and process operating conditions are simultaneously optimized to maximize the process performance. In the present work, the second step namely MOF targeting is addressed. We will show how the optimal results obtained in Step 1 can be used to guide the discovery of high-performance MOFs. Same as the first step, this second step is demonstrated on the separation of propene (PE) and propane (PA), which is currently achieved with energy-intensive cryogenic distillation. Clearly, an optimal design of MOF and PSA process is of great importance for energy saving. The paper is organized as follows. First, the specific workflow of the MOF targeting step is introduced, followed by the introduction of MOF decomposition and building block construction. Afterwards, the workflow will be applied to identify the optimal MOFs for the PE/PA separation.

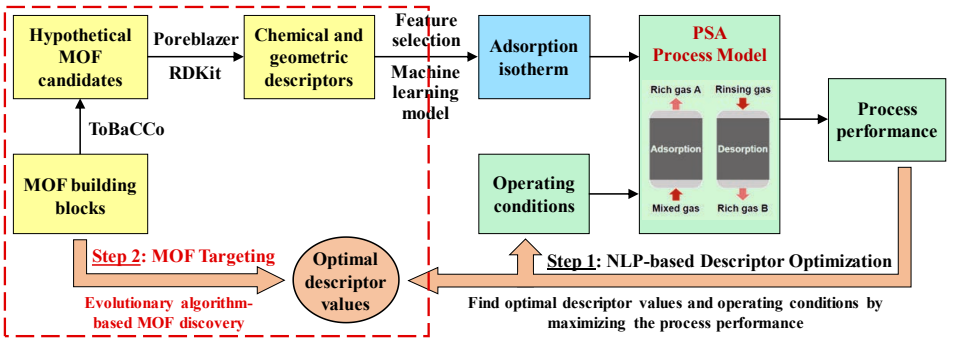


Figure 1. General methodology for integrated MOF and PSA process design

2. Workflow of Descriptor-based MOF Targeting

Figure 2 shows the explicit workflow of the descriptor-based MOF targeting. First, hypothetical MOF candidates are generated by retrieving existing MOF database and computational MOF synthesis via building blocks (BBs). The computational MOF synthesis is performed using genetic algorithm (GA) that continuously generates potential combinations of BBs. The BB combinations are sent to Tobacco that can validate MOF structural feasibility (e.g., connection feasibility, size feasibility, bond feasibility, etc.) based on certain construction rules (Colón et al. 2017). In this case, the computationally feasible MOFs can be obtained. For each MOF candidate generated, its key descriptors (i.e., 9 geometry and 10 chemical descriptors pre-specified in Step 1) are calculated using Poreblazer and RDKit. Afterwards, the feasibility of calculated descriptors are verified, including bound feasibility, design space feasibility, and tailor-made feasibility criteria (see Section 3). For those MOFs whose descriptors are reasonable, rigorous Grand Canonical Monte Carlo (GCMC) simulation is performed to obtain the single-component adsorption isotherms for both PA and PE. Again, these isotherms are verified via the above feasibility criteria. If feasible, the MOFs are used for final PSA process optimization. This can enable the generation of high-performance MOFs and better adsorption process. In the following context, the screening of MOFs from a known database is first demonstrated for defining the benchmark process. Then, the construction of MOF BBs are introduced, followed by the computational synthesis of new advanced MOFs.

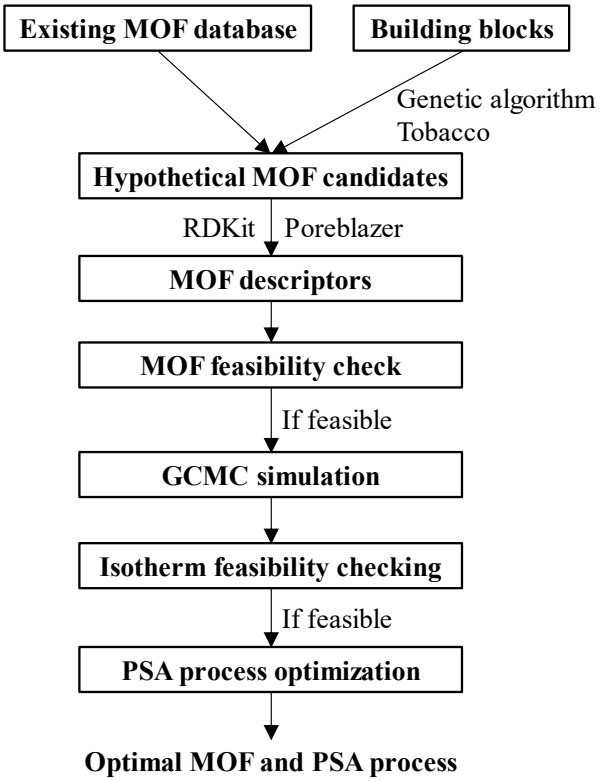


Figure 2. Workflow of descriptor-based MOF targeting

3. Descriptor-based MOF Screening

In Step 1, 471 different MOFs are already selected from the CoRE MOF database where stable and synthesizable MOFs with available atomic coordinates are collected. These 471 MOFs include considerable diversity of chemistry, geometry, and topology. Here, to demonstrate the above workflow, the potential MOFs are first screened out from the 471 MOFs for obtaining a benchmark for subsequent comparison. Before this, two feasibility criteria are established in advance for enhancing screening efficiency. First, based on the optimal isotherm obtained in Step 1, 250 pairs of PA and PE isotherms are sampled using the Latin hypercube sampling approach. The hypothetical isotherms are sent to perform rigorous PSA process optimization. The same as in Step 1, the separation specifications are 99% PE with recovery larger than 30%. The result shows that 116 sets of isotherms can successfully separate PA and PE. Based on the process feasibility data, two criteria can be concluded.

$$PA_2 \leq 2.01 \tag{1}$$

$$PE_{0.01} \leq 6.5879 \times PE_{0.01} - 0.3525 \tag{2}$$

The first criteria is that the adsorption loading of PA at 2 atm should be less than 2.01 mol/kg. The second is that the adsorption loading of PA and PE at 1013 Pa should fulfill a linear constraint. For each of the 471 MOFs, their isotherms have already been estimated using rigorous GCMC simulations and given in our previous publication. Based on those isotherms and the two feasibility criteria, it can be found that only 19 MOFs can be regarded as feasible for PA/PE separation. These 19 MOFs are then sent to PSA process optimization and it can be found that only 9 MOFs can meet the separation specifications. Table 1 lists the top 5 candidates out of the 471 MOFs. Meanwhile, all of the other 452 MOFs are also sent for PSA process optimization. It turns out that none of the 452 MOFs can meet the separation requirements. Thus, from these results, we can conclude that the two criteria (eqs. 1 and 2) can be used as necessary conditions, instead of sufficient conditions, for efficient MOF screening.

Table 1. Feasible and most promising candidates screening from the 471 CoRE MOFs

MOFs	Energy consumption (kWh/kg PE)
SEYDUW	92
QUJFUX	98
XOVVIO	214
VISTUM	251
XEHTUB	257

4. Construction of MOF Building Blocks

In general, MOF can consist of metal vertexes, organic vertexes and organic linkers. The vertexes and linkers are assembled in certain topology. Here, the 471 CoRE MOFs are decomposed into the corresponding building blocks. The decomposition procedures are as follows. First, the topology of each MOF is characterized using Topospro that is a open-source software for topology characterization of periodic structures including MOFs. Then, referring to the introduction of topology templates given in the Reticular Chemistry Structure Resource (RCSR) database, the types of vertexes and the corresponding number of connections are identified. Based on these information, the

metal and organic vertexes can be easily identified and the organic linkers between vertexes can be obtained. Following these procedures, Figure 3 shows that 81 metal vertexes, 85 organic vertexes, 133 organic linkers, and 68 regular topologies can be obtained from the 471 CoRE MOFs. The obtained BBs can be subsequently used for computational MOF synthesis. Note that a fraction of MOFs possessing irregular topology that cannot be described by three letters from the RCSR database are simply discarded, since those topologies cannot be used for computational MOF synthesis.

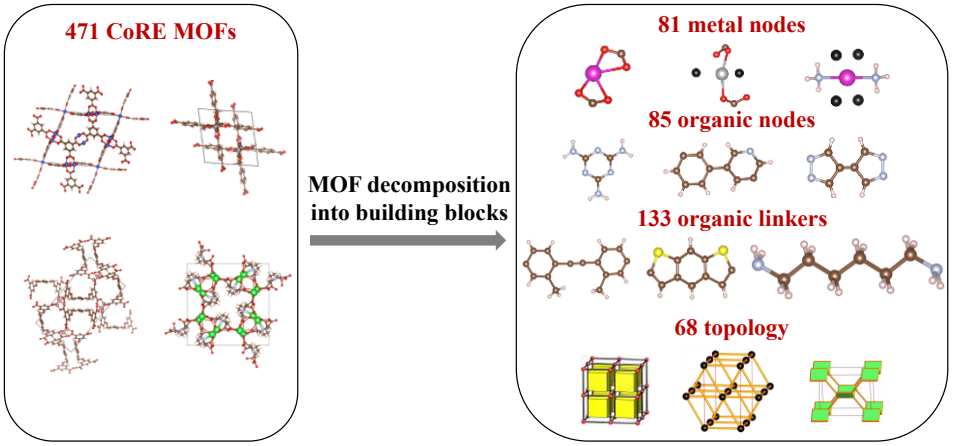


Figure 3. Decomposition of 471 CoRE MOFs into various building blocks

5. Computational MOF Synthesis and MOF Targeting

5.1. GA-based MOF synthesis

For easy implementation, 26 topologies that can usually support the use of only one type of metal vertex are selected out of the 68 topologies. Based on these 26 topologies, GA is adopted to efficiently generate potential combinations of BBs. Each gene consists of 3 integer variables denoting the selection of metal node, organic linker, and topology. The produced BB combinations are directly sent to Tobacco to generate hypothetical MOFs (Colón et al. 2017). In this process, a series of MOF construction feasibility rules will be verified. There are several possible infeasible outcomes such as node connections mismatching with topology, bond infeasibility, etc. Only if a BB combination successfully pass all the feasibility rules, a hypothetical MOF written in the form of crystallographic information framework (.cif) can be obtained. The cif file can be directly used to calculate the corresponding chemistry and geometry descriptors and perform GCMC simulation using the open-source software RASPA (Dubbeldam et al. 2016). Clearly, it is computationally demanding to perform GCMC simulation for each generated hypothetical MOFs. In this case, the synthesized MOFs can go through the screening procedures described above and only the survived MOFs are sent for GCMC simulation as shown in Figure 2.

5.2. Optimal MOF targeting

After over 10 generations of GA operations, a set of potential MOFs that can survive after the feasibility verifications are obtained such as topology *acs* with 6-connection-

based two-Zn metal vertex and bromobenzene organic linker and topology *rob* with 6-connection-based two-Ag metal vertex and benzene organic linker.

6. Conclusions

As a continuation, the present work elaborates the second step of our previous proposed integrated MOF and PSA process design framework. The objective is to use computational tools to generate optimal MOFs based on the optimized MOF descriptors obtained from the first step (Zhang et al. 2021). To achieve this, a detailed descriptor-based MOF targeting workflow is invented. First, hypothetical MOF candidates can be either retrieved from known MOF databases or synthesized computationally via building blocks. According to the optimal descriptor values, a set of MOF building blocks are selected and defined in ToBaCCo. From the building blocks, a series of hypothetical MOF candidates with detailed chemical and structural information can be generated. With this information, their descriptors can be computed and used to predict the adsorption isotherms via the machine learning models developed in the first step. Then, the isotherms go through a preliminary screening step where improper isotherms and the corresponding MOFs are discarded. Finally, the remaining candidates can be sent for rigorous GCMC simulation to obtain adsorption isotherms, which are subsequently used in the process evaluation in order to find the real optimal MOFs. As demonstrated by the PA/PE separation example, hypothetical MOFs with a great potential of leading to better process performance can be computationally synthesized. The largest novelty of the present work is the use of multiscale modeling approach to integrate the variations of MOF chemistry and structure into P/VSA process design. This provides a reliable and efficient way for computational adsorbent design to maximize the practical adsorption process performance.

References

D. Dubbeldam, S. Calero, DE. Ellis, RQ. Snurr, 2016, RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials, Molecular Simulation, 42, 81-101

T. Wang, E. Lin, Y-L. Peng, Y. Chen, P. Cheng, Z. Zhang, 2020, Rational design and synthesis of ultramicroporous metal-organic frameworks for gas separation, Coordination Chemistry Reviews, 423, 213484

TD. Burns, KN. Pai, SG. Subraveti, SP. Collins, M. Krykunov, A. Rajendran, TK. Woo, 2020, Prediction of MOF performance in vacuum swing adsorption systems for postcombustion CO_2 capture based on integrated molecular simulations, process optimizations, and machine learning models, Environmental Science & Technology, 54, 4536-4544

X. Zhang, T. Zhou, K. Sundmacher, 2021, Integrated MOF and P/VSA process design: Descriptor optimization, AIChE Journal, doi: 10.1002/aic.17524

YJ. Colón, DA. Gómez-Gualdrón, RQ. Snurr, 2017, Topologically guided, automated construction of metal–organic frameworks and their evaluation for energy-related applications. Crystal Growth & Design, 17, 11, 5801-5810

Z. Yao, B. Sánchez-Lengeling, NS. Bobbitt, BJ. Bucior, SGH. Kumar, SP. Collins, T. Burns, TK. Woo, OK. Farha, RQ. Snurr, A. Aspuru-Guzik, 2021, Inverse design of nanoporous crystalline reticular materials with deep generative models, Nature Machine Intelligence, 3, 1, 76-86

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50050-6

Energy integration through retrofitting of heat exchanger network at Equinor Kalundborg Oil Refinery

Niels Normann Sørensen[a], Haoshui Yu[b*], Gürkan Sin[c*].

[a]*Chemical and Biochemical Engineering, Technical University of Denmark, Kgs. Lyngby 2800, Denmark*
[b*]*Department of Chemistry and Bioscience, Aalborg University, Niels Bohrs Vej 8A, Esbjerg 6700, Denmark*
[c*]*Process and Systems Engineering Center (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Kgs. Lyngby 2800, Denmark,*

hayu@bio.aau.dk.,gsi@kt.dtu.dk

Abstract

Heat integration studies are commonly performed in the wider chemical industry to identify current energy utilization and detect potential improvements with respect to energy efficiency. In this regard, there are several established methodologies, such as: Pinch analysis, Mathematical Programming (MP) and Hybrid methods. In pinch analysis, the objective is to remove cross pinch heat transfer and configure appropriate utilities, based on a minimum approach temperature ΔT_{min}. The Minimum Energy Required (MER) for the network can then be calculated. However, a drawback is that the user must specify the changes of the HEN to achieve MER, which may not be the best economical solution. In MP the latter problem can be expressed as an optimization problem. However, due to the complexity of HEN in the industry, pinch analysis is typically the preferred method (Sreepathi and Rangaiah 2014). A similarity for all 3 retrofit solutions, are the challenges regarding data collection and the associated uncertainty. To accommodate for this, we present a methodology that involves an iterative application of a process simulator with plant data (to match the heat flows) and the uncertainty of the pinch point(s).
The pinch analysis was constrained to 1 reforming section and 3 hydrofining sections. Average temperature, pressure, volume flow rate and assay of the heavy feeds and residues were taken over a month. One month was selected, when the refinery had been cleaned and flowrates were in the normal ranges of operation. After balancing mass and energy based on the SRK EOS, enthalpies were segmented and exported into UniSim ExchangerNet. Based on a minimum approach temperature of 20°C, the cold pinch temperature was found to be 127.5°C and the hot pinch temperature was 147.5°C, with a total of 9MW cross pinch. A feasible retrofit solution could not be achieved for the heat exchanger with the highest cross pinch of 2.88 MW. Nonetheless, a retrofit solution was possible for the heat exchanger with the second-highest cross pinch at 1.16MW. However, the payback time exceeded the specified requirement, which made the retrofit economically infeasible. Nevertheless, the uncertainty analysis showed that 2 possible pinch points existed. The uncertainty of the pinch point would change the retrofit considerably and therefore also the economical potential of the retrofit.

Keywords: Pinch analysis, Retrofit, HEN, Process simulation, Oil refinery.

1. Introduction

In 2017 approximately 19% of energy consumed in the EU industrial sector, was by crude oil refining (European Energy Agency 2015, Bourgeois et al. 2012). The potential to increase energy efficiency and decrease CO_2 emissions in the downstream process, is therefore substantial. One way to optimize energy efficiency, is to increase heat recovery in the HEN. Several established methodology exists to achieve MER for an existing HEN, such as: Pinch analysis, mathematical programming (MP) and hybrid methods (Kemp 2006). MP is the ideal solution, as MER is achieved by optimizing capital cost and ΔT_{min}. The drawback is the implementation of MP, due to constraints and complexity in an industrial HEN. Pinch analysis is a proven method to reach MER, in an industrial HEN, due to its simplicity. However, in pinch analysis, the user has to suggest the best option to achieve MER, in contrast to MP.

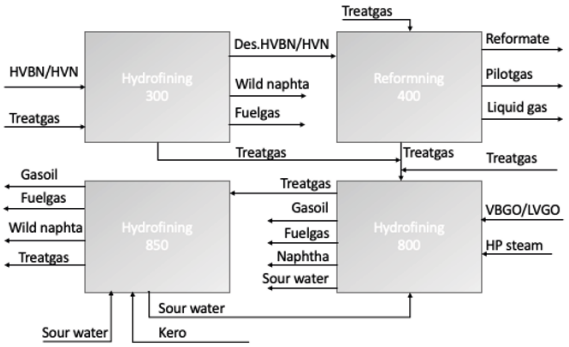


Figure 1: Scope of the retrofit, which includes 3 hydrofining sections and 1 reforming section.

To increase heat recovery and therefore decrease fuel consumption, a retrofit is investigated of a heat exchanger network in Denmark's largest oil refinery, located in Kalundborg. The retrofit will be based on simple pinch analysis, due to the scale of the HEN, which includes 1 reforming section and 3 hydrofining sections (see figure 1).

2. Methodology

The novelty of this study is the clear methodology of using pinch analysis for an industrial retrofit. In studies on industrially HEN retrofits such as: (Alhajri et al. 2021) and (Joe and Rabiu 2013) no emphasis was on: Data extraction, balancing mass, balancing energy and the uncertainty of the pinch point. Real process data has to be balanced by mass and energy due to a combination of sensor calibration errors and taking averages of process data. Process fluctuations have to be accounted for, as it will affect the pinch point and therefore the economical feasibility of the retrofit. Furthermore, in retrofit studies, the estimation of capital cost is usually only based on the HE. However, pipping and instrumentation can be more expensive than the HE, which will increase the payback time considerably.
To accommodate this, we present a systematical approach (see figure 2), to analyze and suggest a retrofit, of the HEN.

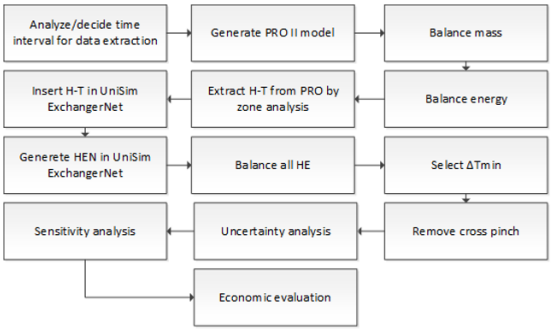


Figure 2: The overall method used for analyzing and suggesting a retrofit, based on pinch analysis.

The first step is to determine, the time interval of data extraction. In general, the time interval for the data extraction, is a trade-off between difficulty in balancing mass/energy and gathering enough data to reflect the uncertainty of the system. To easier balance mass and energy, time interval can be chosen, when key streams are in normal operation. For this HEN, process values were extracted based on 1-month of averages process values, such as: Assay, temperature, pressure, densities and volumetric flow-rate, where major heavy streams were in normal operation. Whereas, data extraction for the uncertainty analysis was based on daily averages values, which meant it was easier to balance mass and energy in contrast to the base case.
In general, for the petroleum industry, it is well known that cubic equations of state can be used as a valid thermodynamic model, as the majority of components are simple hydrocarbons. The Soave-Kwong equation of state (SRK), was therefore the chosen thermodynamic model for all conducted simulations in PRO II.
The properties of the stream in the 4 sections were acquired by modelling all unit operations of the block in PRO II. Where the unit operations included: Distillation's columns, flash drums, heat exchangers, pumps, compressors, and valves.
The composition of the effluent of reactors, for the HF sections, was estimated by using an assay of the remanence. The saturation of olefins could somehow be accounted for, by assuming that most of the product would end up as fuelgas in the stripper. As the volumetric flowrate of fuelgas was known and the molecular weight was expected to be in the range of 25-35g/mol, then the amount and composition of product could be fitted by iterating.
The effluent composition of the reactor in the reforming section was estimated by using GC measurements.
In each section the inlet and outlet massflow, based on averages PV were balanced. Next, all HE was integrated by balancing energy, by using a temperature correction term. When balancing the first HE in the cascade, 4 streams can be corrected. The temperature effluent will then be cascaded to the next HE, which means that the inlet temperature of that HE is fixed. The cascade then continues until all HE's are balanced.
Enthalpy and temperature of cold and heat stream of the respective HE's were then extracted from PRO II, by using zone analysis to account for the nonlinearity. The H-T segments were then inserted into UniSim ExchangerNet. ΔT_{min} of 20°C was then chosen as the minimum approach temperature, as this was the minimum LMTD in the HEN.

Furthermore, a ΔT_{min}= 20°C is typically applied, when implementing a retrofit in the petrochemical industry (March 1998).
The cross pinch heat transfer of each HE, were then calculated and removed based on the methodology of Li and Chang 2010. However, inlet and outlet temperatures were not available for HE in series. To account for the cross-pinch heat transfer for multiple HE in series, rigorous HE was implemented in PRO II by inserting the properties of the shell and tube heat exchanger. The rigorous HE in PRO II uses the well-known design equation, where PRO II can predict the heat transfer coefficient U, correction factor F and the pressure drop. The duty for each HE was then normalized in relative to the total duty, by adjusting the fouling factor.

3. Results

3.1. Removing cross pinch heat transfer

In table 1 one can see that if MER is achieved, one can achieve 9.58MW heat recovery.

Table 1: MER at ΔT=20°C

	Network [MW]	Target [MW]	% Deviation
Q_{hot}	24.61	15.03	163.7
Q_{cold}	27.64	18.07	153
Q_{rec}	59.58	69.16	86

However, one can see in table 2, that the cross-pinch heat transfer is unfortunately distributed on many HE. The highest cross pinch is 2.88MW for E-808D, 1.16MW for E-851 B and 1.11MW for E-411A. A retrofit for E-808D was then investigated.

Table 2: Cross pinch at ΔT_{min}=20°C.

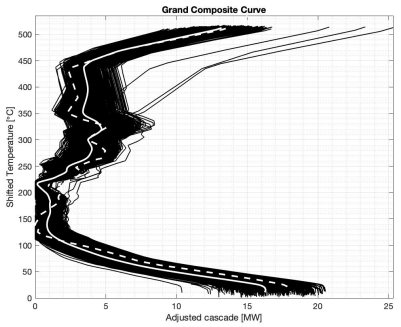
Section	800	850	400	300	800	400	300	300	300	850	800	850	300
HE	E808D	E851B	E411A	E306	E802C	E414	E301F	E301C	E307	E855A	E808B	E856	E302
Cross pinch [MW]	2.88	1.16	1.11	0.95	0.7	-0.387	0.326	0.318	0.237	0.219	0.212	0.114	0.141

First, the required duty to heat the cold stream to the cold pinch. However, the only hot stream below pinch point, that had enough Q, was an air cooler, in the reformate section. If a HE was to be inserted in 0400 then the compressor would compensate for the expected significant pressure drop, which would decrease reformate yield. The option to insert a HE before the air cooler was then discharged.
Another idea was to split the hot streams from several air-coolers. However, the cost of pipping would be too expensive. A feasible hot stream could not be found from a utility stream, so the stream would be needed from another HE. However, due to the relatively high duty required, several splits would be required, which would be expensive, due to the price of pipping. A feasible hot stream could not be found, to heat the cold stream to the cold pinch temperature.
A retrofit solution was not found for E-808D. However, a retrofit was proposed for E-851B, which has the second-highest cross pinch (see table 2). Two options were proposed

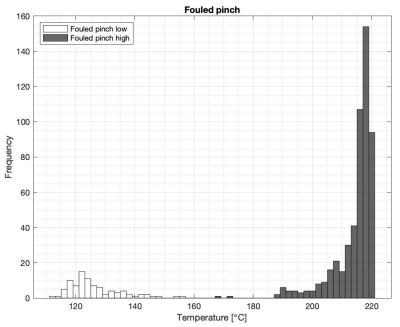
to remove the cross pinch. Both options would heat the cold pinch by using air coolers. However, from experience the refinery wanted at least 1.5MW of cross-pinch, due to a specified payback time, for the retrofit to be economically feasible. The ΔT_{min} was then decreased to see if the cross pinch for the 3 mentioned HE would increase, but the increase of cross pinch was not significant. As a result, an economically feasible retrofit solution could not be obtained, due to a combination of cost of pipping and the only HE with a cross pinch of 1.5 MW was E-808D.

3.2. Uncertainty analysis

The 2 retrofit solution, analyzed in UniSim ExchangerNet was based on average PV's, taken over a month, which resulted in a pinch temperature of 137.5°C. The composite curves were rather parallel to each other, in which a specific pinch point was not well-defined. Rather, than a pinch point a "pinch region" exists, in a range of 110-220°C. To evaluate the uncertainty of the pinch point, MATLAB was used to see the effect of daily averages mass and temperature fluctuations, considering 614 days. Iterating over 614 days, by using the same methodology as described in figure 1, resulted in figure 3a and figure 3b.



(a) Dashed white lines is the clean HEN, solid white line mean fouled HEN and black solid lines pinch over 614 days.



(b) Histogram evlauted over of 614 days.

Figure 3: Uncertainty analysis of pinch point at ΔT_{min}=20°C.

The uncertainty analysis confirms the pinch region as seen in figure 3. However, as seen in 3b it seems that 2-pinch point exists – a high pinch with a mean of 214°C and low pinch temperature with a mean of 126.5 °C. Thermodynamically, it is possible to have multiple pinch points, where the same methodology apply for identifying and removing cross heat transfer. Furthermore, in figure 3b one can see that the higher pinch-point is more frequent and less dispersed than the lower pinch point. In terms of numerical values, the low pinch is in a range of 120-131°C and the high pinch temperature in a range of 213-219°C. The combination of higher frequency and lower uncertainly makes the higher pinch in, relative to the lower-pinch, more significant in terms of calculating the cross-heat transfer. However, even-though all streams were segmented by making polynomial fits, several assumptions were made, in the uncertain analysis. i.e., the pressure and composition were

assumed to be the same as the PRO II simulation. However, there will be pressure changes and the composition of: Treatgas, HVN/HVBN, reformate, VBGO and KERO will change over time. These changes will i.e. influence the volatility for the flash, distillation units and the T-H curves. In contrast to a well-defined pinch point, then it is in this HEN, important to decrease the uncertainty of parameter estimation, as the pinch point(s) in this HEN, is very sensitive to the changes in heat flow, as seen in figure 3a. Ideally, the next step is to synchronize the PRO II simulation, with i.e., MATLAB or Python, to reduce the uncertainly/assumption of the input parameters. Furthermore, the cross heat transfer would be calculated again, based the on the 2 pinch point, by using the methodology from Li and Chang 2010.

4. Conclusions

In this study, it has been shown how mass and temperature fluctuation effect the pinch point and therefore the economic feasibility of an industrial retrofit. When considering the base case of averages process values, a pinch temperature of 137.5°C, was found. The base case network hot utility was 24.61MW, which was 163% above target. The network cold utility was found to be 27.64MW, which was 153% above target. To reach MER with a corresponding 59.58MW of heat recovery. The heat recovery would have to be increased with 9.56MW. However, an economically feasible retrofit could not be found, for the HE with the largest cross pinch of 2.88MW. Two retrofit solution could be found for the HE with the second-highest cross pinch of 1.16 MW. However, the relatively low cross pinch meant that the proposed retrofit would exceed the specified payback time. However, when using process values over 614 days, 2 pinch temperatures were found. One temperature with a mean of 126.5 °C and a pinch at a mean of 214°C, which frequency was significant higher than the pinch at 126.5°C. The amount of cross pinch heat transfer would then be different from the base case, in which the same methodology (Li and Chang 2010) can be applied to suggest a retrofit, when dealing with 2 pinch points. Furthermore, to improve the uncertainty analysis, one could combine MATLAB and PRO II to use the properties generated from PRO II and simulate daily averages values and thereby get a better estimation of the ”true” pinch point(s) in the HEN.

References

Alhajri, Ibrahim H. et al. (2021). “Retrofit of heat exchanger networks by graphical Pinch Analysis – A case study of a crude oil refinery in Kuwait”. In.

Bourgeois, L et al. (2012). “EU refinery energy systems and efficiency”. In.

European Energy Agency (2015). “Final energy consumption by sector and fuel”. In: *Indicator Assessment — Data and maps*, p. 20.

Joe, John M. and Ademola M. Rabiu (2013). “Retrofit of the Heat Recovery System of a Petroleum Refinery Using Pinch Analysis”. In.

Kemp, Ian (2006). *Pinch Analysis and Process Integration*. Elsevier Ltd.

Li, Bao Hong and Chuei Tin Chang (Apr. 2010). “Retrofitting heat exchanger networks based on simple pinch analysis”. In: *Industrial and Engineering Chemistry Research*.

March, Linnhoff (1998). *Introduction to Pinch Technology*. Tech. rep.

Sreepathi, Bhargava Krishna and G. P. Rangaiah (July 2014). “Review of Heat Exchanger Network Retrofitting Methodologies and Their Applications”. In.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50051-8

Modeling and Optimization of Ionic Liquid Enabled Extractive Distillation of Ternary Azeotrope Mixtures

Alejandro Garciadiego[a], Mozammel Mazumder[a], Bridgette J. Befort[a] and Alexander W. Dowling[a*]

[a] *University of Notre Dame, Notre Dame, IN 46556, USA*
* *adowling@nd.edu*

Abstract

To help slow climate change, international efforts have begun to mandate the phase-out of high global warming potential (GWP) hydrofluorocarbons (HFCs) throughout the next decade. Most HFC refrigerant mixtures form azeotropes, complicating separation into the individual HFC components for reuse and recycling. In this paper, we design and analyze ionic liquid (IL)-enabled extractive distillation processes for ternary HFC separations using AspenPlus. Specifically, we design processes to separate three commercially important HFC refrigerant mixtures (R-404A, R-407C, and R-410A) into high purity HFC streams. We find added value of the separation of R-410A of 0.58 $/kg with current market conditions, specifically laboratory-scale IL manufacturing costs (1000 $/kg of IL) and a low-price differential of 1.00 $/kg between raw materials and separated products. If the IL purchase cost decreases 90 % due to mass production, consistent with prior adoption of ILs for niche separations, the added value increases to 0.76 $/kg. Moreover, under proposed reductions in HFC manufacturing, the price of recovered products may dramatically increase in the future. For example, if the price of R-32 increases by 50 %, the added value would reach 3.08 $/kg. In summary, we find IL-based recycling of HFCs is economically viable based on simple technoeconomic analysis. Moreover, this paper reports capital and operation cost curves and a general analysis framework to analyze evolving market conditions.

Keywords: Extractive Distillation; Azeotrope; Ionic Liquid; Modeling; Economic Analysis.

1. Introduction

Thousands of tons of HFC refrigerant mixtures, commonly used in industrial, commercial, and residential applications, are scheduled for phase-out worldwide under the 2016 Kigali amendment to the Montreal Protocol, the European Union F-Gas regulations (2015), and the American Innovation and Manufacturing (AIM) Act of 2020. The latter directs EPA to phase down production and consumption of HFCs in the US by 85 percent over the next 15 years. Common HFC mixtures such as R-410A (50 % R-32, 50 % R-125), R-404A (44 % R-125, 4 % R-134a, 52 % R-143a), and R-407C (23 % R-32, 25 % R-125, 52 % R-134a) are targeted for phase-out because of their high global warming potential (GWP): R-410A with 2088 GWP, R-404A with 3922 GWP, and R-407C with 2107 GWP, where CO_2 has a GWP of 1 by definition. However, R-32 and other HFCs have a low GWP and could be reused after a global phase-out. R-134a is used in R-450A, offering similar performance but with a lower GWP (547) (Honeywell, 2021). Unfortunately, there is no means to easily separate HFC mixtures due to their azeotropic

or near azeotropic nature. Without a new economically viable separation process, the phase-out will require all HFCs to be collected and incinerated.

Extractive distillation, the most common method for separating azeotropic or close-boiling mixtures, is a promising approach to separate HFC mixtures. Moreover, tailored IL solvents can enable extractive distillation of near-azeotropic HFC mixtures. In 2003, Lei et al. first proposed extractive distillation with ILs as entrainers, and Lei et al. (2005) discussed the use of ILs in extractive distillation in detail. ILs have exhibited high capacity as entrainers to separate azeotropic or close-boiling mixtures (Pereiro et al., 2012). ILs can be recycled in separation processes, reducing the material demands and improving the economics (Zhao et al., 2005, Zhao et al., 2017). Shiflett and Yokozeki (2006) proposed extractive distillation to separate fluorinated refrigerant mixtures using ILs.

2. Methods

2.1. HFC Separation Process Development and Modeling

In this work, we design three extractive distillation processes to separate three ternary azeotrope mixtures, R-404A, R-407C, and a mixture of R-410A and R-22 using an IL entrainer. Table 1 summarizes these three case studies (Finberg and Shiflett 2021). We use the Peng-Robison (PENG-ROB) equation of state to calculate thermodynamic properties. We fit the HFC binary interaction parameters similar to Shiflett and Yokozeki (2006, 2007).

Table 1. Compositions of HFCs mixtures separated and IL used.

	R-404A	R-407C	R-410A
HFC	mol/mol	mol/mol	mol/mol
R-32	0.00	0.23	0.45
R-125	0.44	0.25	0.45
R-134a	0.04	0.52	0.00
R-143a	0.52	0.00	0.00
R-22	0.00	0.00	0.10
IL used	[emim][Tf_2N]	[bmim][PF_6]	[bmim][PF_6]
IL (kg/h)	2000	400	750

Figure 1 shows the process flow diagram for the R-404A case study. The process flow diagrams for the other case studies are nearly identical and not shown for brevity.

2.2. Sensitivity analysis

We performed single-parameter sensitivity over eight variables. The base case used 20 theoretical stages, a flowrate of IL of 1000 kg/h, the IL is fed in stage 2, the HFC mixture fed at stage 15, the inlet temperature of 25 °C, the pressure of 10 bar, a reboiler temperature of 130 °C, and a reboiler ratio of 2.5. We found that the extractive distillation column's pressure and reflux ratio are most important to minimize energy consumption while obtaining 99.5 mol% purity of all HFC products. Through our sensitivity analysis, we look to obtain the desired purity (99.5 mol%) of R-134a in the distillate of the extractive distillation in the presence of [emim][Tf_2N] with moderate energy consumption. We found that it is impossible to reach the required purity without 25 theoretical stages and a flowrate of IL of 2000 kg/h, even though they significantly influence the capital cost. The IL is fed in stage 2, and the HFC mixture is fed in stage 20 at a temperature of 20 °C. We selected a pressure of 7 bar in the column and a reboiler

temperature of 90 °C to ensure the energy consumption was as low as possible while reaching the purity target. Finally, following the same analysis, we selected a reflux ratio of 3. Aspen equipment sizing tools were used to size the equipment.

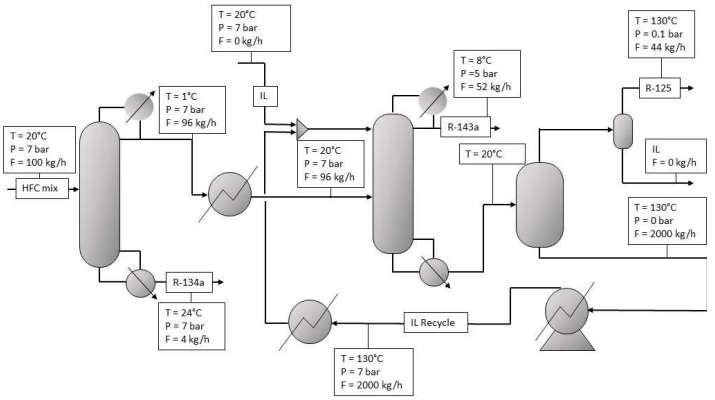


Figure 1. Process flow diagram of the developed HFC separation process.

3. Economic Performance Evaluation

We now analyze the economics of the design HFC separation processes. We evaluate the capital cost, shown in Eq. (1), which includes equipment, installation cost, and the price of the IL as expressed in units of M$/y.

$$\text{Capital cost}\left(\frac{\text{M\$}}{\text{y}}\right) = \text{Equipment cost}\left(\frac{\text{M\$}}{\text{y}}\right) + \text{Installation cost}\left(\frac{\text{M\$}}{\text{y}}\right) + \text{IL price}\left(\frac{\text{M\$}}{\text{y}}\right) \quad (1)$$

We assume a 20-year (N) plant lifetime and a salvage value of 20 % of the cost of the plant assets (excluding the IL). We assume 24 hours a day workload for 330 days in a year for all calculations. We calculate the annualized capital cost (C_{anm}) using Eq. (2), in which CRF is the capital recovery factor, and C_{NPC} is the net present cost estimated in AspenPlus. We assume a nominal discount rate (i') of 8 % and an expected inflation rate (f) of 3.5 % to calculate the real discount rate (i). With the assumptions above, we calculate a capital recovery factor (CRF) of 0.077 using Eqs. (3) and (4). We estimate 2,000 kg/h of IL, which corresponds to a column fill of 65 %. We estimate operation costs using AspenPlus and the following utility costs: electricity (0.07$/KW), cooling water (120 $/MMGAL), and high-pressure steam (8.22 $/Klb).

$$C_{anm} = (CRF)(C_{NPC}) \quad (2)$$

$$CRF = \frac{i(1+i)^N}{(1+i)^N - 1} \quad (3)$$

$$i = \frac{i' - f}{1+f} \quad (4)$$

As shown in Figure 2, the capital and operating costs ($/kg of HFC feed) of the R-404A, R-407C, and R-410A separation increase as we decrease the mixture feed flow rate. We observe that the significant increase in the capital cost is due to the amount of IL necessary to achieve the 99.5 mol% purity of HFCs desired in the separation. The cost of equipment and installation have minor variations as the size of the equipment is nearly minimum or standard size. As observed in Figure 2, the capital cost may increase to up to 60 % of the total cost. The rise in total capital cost is dependent on the ratio of IL/HFC mixture required for each process. 1 % to 2 % of IL degradation per year corresponds to an increase of the operating cost of 0.03 $/kg to 0.05 $/kg, respectively.

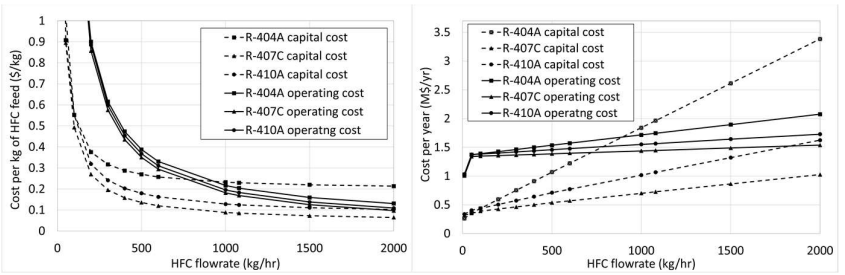


Figure 2. Capital and operating cost of the R-404A, R-407C, and R-410A AspenPlus model. The capital cost of the separation process increases rapidly as we increase the inlet flowrate.

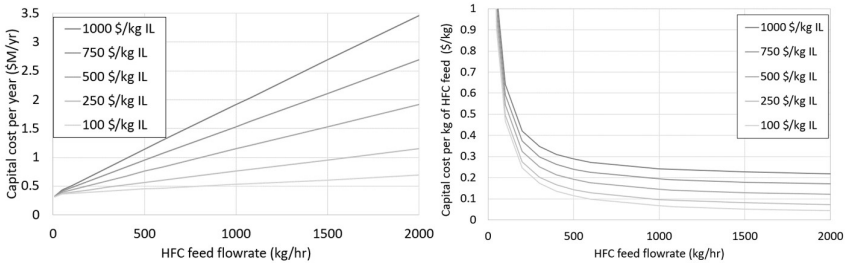


Figure 3. Influence of the ionic liquid price in capital cost for the separation of R-404A.

Currently, most ILs are only available in high purity for laboratory-scale experiments at high prices of $1,000/kg. Historically, after an IL is selected for a commercial application and production increases, the price decreases by 90-92 % (Shiflett et al., 2020). In anticipation of a similar economy of scale, we consider five IL price scenarios: 1,000 $/kg, 750 $/kg, 500 $/kg, 250 $/kg, and 100 $/kg. Figure 3 shows the impact of IL price on capital costs. Specifically, the capital cost (M$/y) increases linearly with the IL flowrate (kg/h) at a given IL price. As expected, changing the IL price changes the slope of this relationship. Moreover, the capital cost is extremely sensitive to the IL price. For example, at 5000 kg/h IL flowrate, decreasing the IL price from 1000 $/kg (laboratory scale specialty chemical) to 100 $/kg (commercial IL) decreases the capital cost from 8

M$/y to 1.5 M$/y. We reiterate that previous commercialization of ILs suggests a 90 % reduction in IL price is reasonable (Shiflett et al., 2020).

$$\text{Added value}\left(\frac{\$}{\text{kg}}\right) = \underbrace{\text{Sell price low GWP components}\left(\frac{\$}{\text{kg}}\right) - \text{Cost of recovery HFC mix.}\left(\frac{\$}{\text{kg}}\right)}_{\text{Price differential}} \underbrace{- \text{Capital Cost}\left(\frac{\$}{\text{kg}}\right) - \text{Operating cost}\left(\frac{\$}{\text{kg}}\right)}_{\text{Costs}} \quad (5)$$

Next, we propose added value, with units $/kg of HFC feed, as a metric to easily compare different hypothetical scenarios. Eq. (5) calculates added value from the price differential and costs. The selling price of low GWP components is the value of the recycled products, and the cost of recovery HFC mixture corresponds to the value of the used HFC refrigerant mixtures (half of the cost of production and transportation of the HFC mixture used as a base and worst-case scenario). A negative cost of recovery HFC mixture is possible with government subsidies incentivizing HFC recycling (instead of illegal venting). Figure 2 reports the operating and capital costs ($/kg) as a function of the HFC feed rate. Similarly, Figure 3 shows the dependence of capital cost ($/kg) on IL price. Because the added value metric represents profit per kilogram of HFC processed, it allows quick evaluation of different market scenarios (e.g., HFC and IL prices).

Using values from these plots, the added value metric can quickly be used to evaluate the benefits of new ILs for the separation process. For example, if a new hypothetical IL required 20 % less mass than the analyzed ILs, the cost in Figure 3 can be proportionally reduced. Likewise, if a new hypothetical IL reduces the separation energy requirement by 50 %, the operating cost value used in Eq. (5) can be reduced by approximately 50 %. This metric gives valuable insights and enables fast “what if” analyses to guide IL and process design.

Under current market conditions, we found that R-410A separation has an added value of 0.55 to 0.72 $/kg with an IL price of 1000 $/kg and 100 $/kg, respectively. Under a futuristic scenario where phase-outs in production doubles the market price for R-32, the added value of the separation of R-410A could be as high as 5.60 $/kg to 5.78 $/kg with an IL price of 1000 $/kg and 100 $/kg respectively. If the price of R-32 increases by 50 %, the added value would reach 3.08 $/kg. We found that the price of ILs has the most significant impact on the capital cost, and the price differential between the HFC mixture and the pure HFC impacts the added value and the payback period.

4. Conclusions

In this paper, we show that separating and recycling HFCs with extractive distillation utilizing ILs is economically attractive, especially under anticipated future scenarios. It is important to note that ILs are viscous, and a rate base model is needed for rigorous design and more accurate technoeconomic analyses. This is left as future work.

The presented results are based on currently available ILs [bmim][PF_6] and [emim][Tf_2N]. However, ILs can be tailored for specific purposes due to the vast diversity of anions and cations available. For example, tailored ILs with higher selectivity would

reduce the amount of IL required and thus capital costs. Tailoring other properties of the ILs, such as the density, viscosity, and thermal capacity could reduce the operating costs of the process.

There are also unexplored opportunities to optimize the extractive distillation process. While the one and two-dimensional sensitivity analyses presented here show 25 theoretical stages and the amount of IL necessary for the separation, rigorous optimization may find additional opportunities for improvement by exploiting interactions across multiple design decision variables. Moreover, simultaneous process optimization (e.g., flowrates, temperatures) and heat integration may further reduce the energy intensity of the process by systematically balancing reboiler duty and compression costs (e.g., by changing column pressure). This is left as future work.

5. Acknowledgments

We acknowledge support from the National Science Foundation under grant no. CBET-1917474, University of Notre Dame and funding from the Richard and Peggy Notebaert Premier Fellowship. We also thank Ethan Finberg from the University of Kansas for his insights of Aspen Plus modeling.

6. References

Honeywell, Solstice® N13 (R-450A), (DoA: 2021). URL https://www.honeywell-refrigerants.com/europe/wp-content/uploads/2015/03/Solstice-N13-TDS-141027-LR-vF.pdf

Z. Lei, B. Chen, Z. Ding, 2005. Special Distillation Processes, Elsevier Science, Amsterdam, Netherlands

Z. Lei, , C. Li, B. Chen, 2003, Extractive distillation: a review, Separation & Purification Reviews, 32, 121–213.

E. A. Finberg, M. B. Shiflett, Process designs for separating R-410a, R-404a, and R-407c using extractive distillation and ionic liquid entrainers, 2021, Industrial & Engineering Chemistry Research 60, 44.

A. Pereiro, J. Araujo, J. Esperança, I. Marrucho, L. Rebelo, 2012, Ionic liquids in separations of azeotropic systems - a review. J. Chem. Thermodyn. 46, 2–28.

M. B. Shiflett, A. Yokozeki, 2006, Separation of diuoromethane and pentafluoroethane by extractive distillation using ionic liquid, Chimica Oggichemistry, 24, 28-30.

M. B. Shiflett, M. A. Harmer, C. P. Junk, C.P, A. Yokozeki, 2006, Solubility and diffusivity of 1,1,1,2-tetrafluoroethane in room-temperature ionic liquids. Fluid Phase Equilib 242, 220–232.

M. B. Shiflett, A. Yokozeki, 2007, Solubility differences of halocarbon isomers in ionic liquid [emim][Tf_2N]. J. Chem. Eng. Data 52, 2007–2015.

M. B. Shiflett, 2020, Commercial Applications of Ionic Liquids, Springer, Switzerland.

Y. Zhao, R. Gani, R. M. Afzal, X. Zhang, S. Zhang, 2017, Ionic liquids for absorption and separation of gases: An extensive database and a systematic screening method. AIChE Journal, 63, 1353–1367.

H. Zhao, S. Xia, P. Ma, 2005, Use of ionic liquids as ‘green’ solvents for extractions, Journal of Chemical Technology & Biotechnology. 80, 1089–1096.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50052-X

Optimal Design of Hybrid Distillation/ Pervaporation Processes

Dian Ning Chia and Eva Sorensen*

Department of Chemical Engineering, University College London (UCL), Torrington Place, London WC1E 7JE, united Kingdom

** Corresponding author: e.sorensen@ucl.a.cuk*

Abstract

Hybrid distillation/pervaporation processes have the potential to reduce the energy consumption and cost of standard distillation for difficult separations such as that of azeotropic systems. Current optimization strategies for such hybrid processes either considers only a simplified membrane system or requires repeated (sequential) optimization for each potential number of membrane stages. This work proposes a superstructure optimization strategy for the optimal design of hybrid distillation/pervaporation processes, and discusses different solution alternatives for how to handle the integer nature of the membrane network, as well as proposes a procedure for systematic initialization, simulation, and optimization of the process. The strategy is illustrated for an azeotropic separation, demonstrating that the optimal design can be obtained in a fraction of the time compared to repeated simulation.

Keywords: Distillation, Hybrid Distillation, Membrane network, Superstructure, Optimization

1. Introduction

The optimization of a hybrid separation process is a challenging task due to the highly integrated and complex design. Current optimization studies typically simplify the membrane system to consider only the membrane area (Singh and Rangaiah, 2019), and/or very limited (typically up to 3) number of membrane stages in series (Koch et al., 2013). A more holistic superstructure optimization of membrane systems was proposed by Marriott and Sorensen (2003), however, the superstructure optimization was solved for n superstructure sizes (from one to n membrane stages), then compared. This strategy reduces computational burden, but requires some manual effort in reconstructing the superstructure for each stage addition, so the maximum number of membrane stages that can be considered is limited. Moreover, when applying this strategy in a hybrid process, which may potentially involve more than one recycle between units, the number of iterative procedures increases and becomes challenging.

Most studies on the optimization of hybrid processes often reported only their main membrane equations, the flowsheet of the superstructure, and the optimization method used, without clearly describing how to overcome the inevitable numerical/ mathematical issues faced during the initialization/convergence of the optimization.

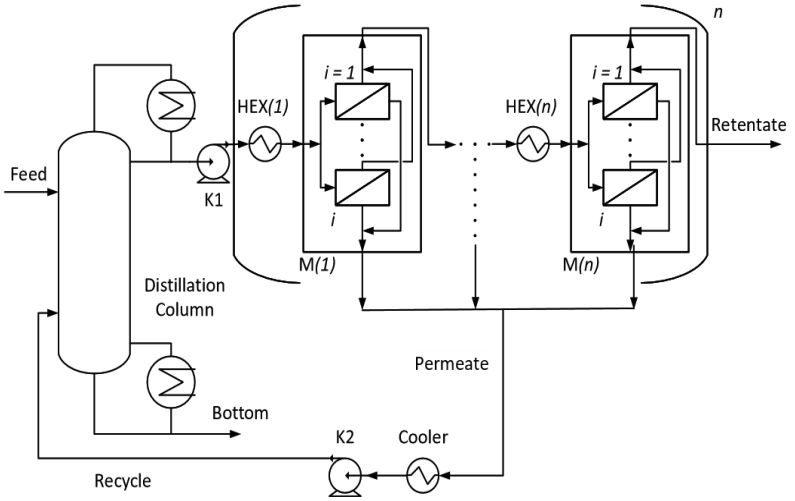


Figure 1: Flowsheet of the hybrid distillation/pervaporation process with a membrane system further purifying the distillate. The membrane system depicts the first and last ($n = n_{max}$) membrane stages connected in series and the i membrane modules connected in parallel within each membrane stage.

This work therefore aims to: (1) propose a membrane superstructure and optimization strategy which improves the convergence and allows the simultaneous optimization of the full membrane system, and (2) apply the full membrane superstructure in a hybrid process with recycle streams.

2. Methodology

2.1. Membrane System Superstructure

A lumped hollow fiber pervaporation membrane model is developed in gPROMS ProcessBuilder (Process Systems Enterprise, 2020) and the model is validated against the work of Tsuyumoto et al. (1997) (not shown). The superstructure of the membrane system extends from the work of Marriott and Sorensen (2003). Due to the small scale plant considered in this work and the fact that the area needed by the heat exchangers are very small for cases with and without recycle streams (outlet streams from a membrane stage recycled back to the previous heater/membrane stage), the capital costs of the heat exchangers are almost constant. Therefore, recycle streams are not considered in this work because a plant with a large recycle flow but with few heaters supplying all the heat would be unrealistic (Marriott and Sorensen, 2003). The membrane stages are connected sequentially in series and in each membrane stage the feed is evenly distributed between the number of membrane modules in parallel (Figure 1), thus greatly reducing the computational burden as only one mathematical model is needed to describe the membrane module (i.e., the membrane stage feed stream is divided by the number of membrane modules in parallel) (Marriott and Sorensen, 2003). A membrane stage feed heater can potentially be added to improve the separation performance. Thus, the optimization task of this membrane system includes the number of membrane stages connected in series (n), the number of membrane modules connected in parallel in a membrane stage (i_n), the existence of the membrane

stage feed heater (HEX), and if in existence also the heater temperature, totalling 4n optimization variables.

2.2. Membrane System Superstructure Simultaneous Optimization Strategy

The membrane superstructure is a composite model of n_{max} membrane stages, where n_{max} should be set as a parameter which cannot be varied as this might introduce convergence difficulties when optimizing n_{max}. There is a lack of open literature discussing the optimization of a membrane network, thus this work introduces three modelling/optimization strategies which can overcome the computational difficulties.

The first strategy is to eliminate the non-existing membrane stages from the solution. This can be done by setting, for example, the membrane fluxes of each component or membrane length/area in the non-existing membrane stages to zero. This strategy is straight-forward and only requires the related variables to be set to zero (directly or via a binary variable). The simulation results for the outlet of the membrane section can still be collected at the n_{max} stage even if n_{max} is not the optimal number of stages. However, this strategy has difficulties at the initialization stage and often fails to converge into a feasible solution. The membrane model involves differential equations, and a good set of initial values is essential to ensure convergence to a solution. However, it is impossible to provide a different set of initial values for each structure (i.e. for each potential number of membrane stages) in the superstructure, and instead, a single set of initial values is typically provided for the whole superstructure. The existence of zeros for non-existing stages may therefore cause large difference between the initial and final values, and will cause numerical errors such as division by zero.

To avoid using zeros, the non-existing membranes can instead be given a feasible non-zero pseudo-feed. This can be achieved by providing the feed to the non-existing membranes a user-defined pseudo-feed or a copy of the feed from the last-existing membrane. The simulation results can still be taken from the n_{max} membrane stage. This strategy does improve the convergence, but it requires a number of additional "if-else" statements which increases the computational costs and difficulties.

The last strategy is to assume that all membrane stages exist, regardless of the optimization result for the number of membrane stages, but the results are collected at the optimized number of membrane stages. This strategy can avoid using zeros, and has fewer "if-else" statements and smoother simulations. However, although not encountered in this work, theoretically, this strategy may face a situation where the feed streams to the non-existing membrane stages are overly pure if the product purity constraint is high, and may therefore cause mass balance convergence issues. The chance of this issue happening can be minimized by forcing the number of membrane modules in parallel in the non-existing membrane stages to one, thus reducing the separation performance of the non-existing membrane stages (which are after all just theoretical rather than actual). From the authors' experience, this strategy does have the best convergence performance and is therefore recommended and applied in this work.

2.3. Hybrid Process Simulation and Optimization

In this work, the hybrid process shown in Figure 1 is considered following the procedure shown in Figure 2. The membrane system (including heaters) is user-defined using the recommended modelling structure discussed in Section 2.2., whilst the other unit models required are modeled within gPROMS ProcessBuilder.

As the hybrid distillation/pervaporation process is often used for handling separation tasks involving azeotropes, rough mass balance calculations around each unit (without considering the recycle stream) can initially be performed by assuming that the column distillate is at the azeotropic point and all product streams are at the required purity. Then, the initial design of the distillation column can be obtained using a proper shortcut method. The membrane system can initially be set with a large number of membranes stages and membrane modules in parallel, so that the product purities are achieved (some trial-and-errors may be needed). Next, the hybrid process is constructed including the recycle stream from the membrane unit back to the column, then simulated with the simulation results obtained from the individual unit simulations as initial values and providing initial values for the recycle stream. If the simulation failed, the values of the key design variables (e.g., reflux ratio, distillate, and number of membrane stages) should be varied and the simulation rerun. Else, the optimization of the whole process can be carried out. In this work, a user-defined genetic algorithm (GA) coded in MATLAB is used, and the details of the settings and strategies applied can be found in our previous study (Chia et al., 2021). The tool gO:MATLAB (Process Systems Enterprise, 2019) is used for data transfer between gPROMS ProcessBuilder and MATLAB. An improvement is made by using parallel computing (18 workers) in MATLAB to speed up the optimization.

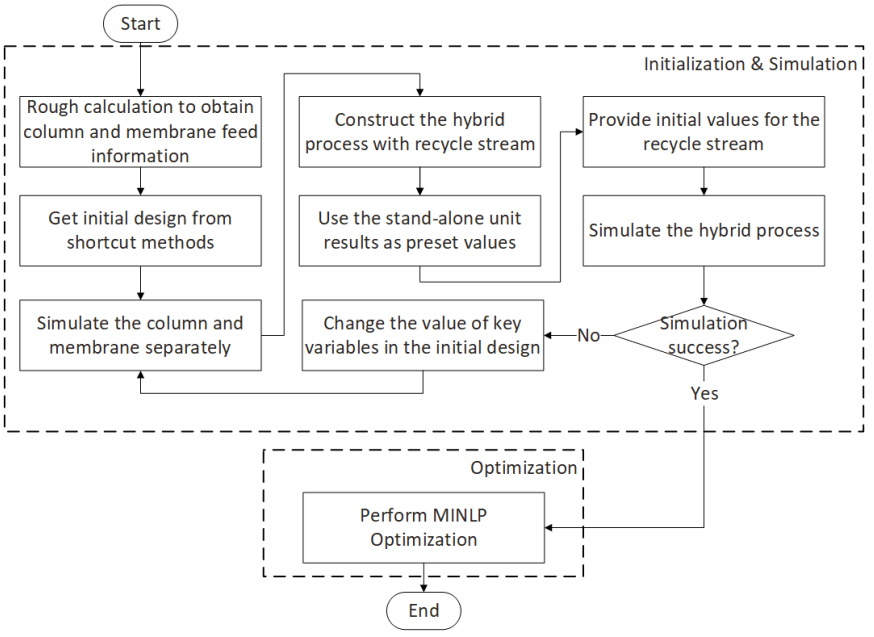


Figure 2: Procedure followed in this work for the initialization, simulation, and optimization of the hybrid process.

Table 1: Main optimization results obtained from repeated optimization and optimization recommended in this work, where n is the number of membrane stages. $n = 1,2,3$ are also optimized, but cannot achieve the product specifications, so the results are not shown here. (Purity specifications in all product streams are *99 mol %*.)

Items	**Repeated Optimization**					**This Work**
	***n* = 4**	***n* = 5**	***n* = 6**	***n* = 7**	***n* = 8**	
Column						
Total stages	23	18	20	20	19	19
Feed stages (Main/Recycle)	19/19	14/17	17/17	15/17	15/17	15/18
Distillate ($kmol\ h^{-1}$)	22.84	23.05	22.72	23.24	23.04	23.23
Molar reflux ratio	1.17	1.21	1.3	1.15	1.23	1.14
Membrane Network *						
No. membrane stages	4	5	6	7	8	6
No. modules in stage 1	13	8	6	8	6	5
No. modules in stage 2	18	12	7	7	2	9
No. modules in stage 3	18	10	11	8	10	9
No. modules in stage 4	19	15	17	17	10	20
No. modules in stage 5	-	18	3	9	12	9
No. modules in stage 6	-	-	17	5	9	12
No. modules in stage 7	-	-	-	10	6	-
No. modules in stage 8	-	-	-	-	8	-
Total no. modules	68	63	61	64	63	64
Total membrane area (m^2)	408	378	366	384	378	384
Fitness and Time						
TAC ($M\ \$\ y^{-1}$)	0.7669	0.7573	0.7588	0.7609	0.7605	0.7577
CPU time (s) †	802	1193	875	1164	1370	1233
Total CPU time (s)			5404			1233

* The existence of membrane stage feed heater is also optimized but not shown here

† Parallel computing used to speed up optimization, number of workers/cores = 18

3. Case Study

The separation of an azeotropic ethanol-water mixture is used as a case study, with UNIQUAC as the thermodynamic model. The feed is provided at 200 $kmol\ h^{-1}$ with 10 *mol%* ethanol, and is a saturated liquid at 1 *bar*. The optimization task is to minimize the total annualized cost (TAC), where the TAC calculations can be found in Sinnott and Towler (2020) and Seider et al. (2016). To examine the performance and reliability of the proposed optimization strategy, the hybrid process is also optimized using the strategy by Marriott and Sorensen (2003) (where optimization is repeated at each number of membrane stages and the design with minimum TAC is selected as the optimal design) and is termed as "repeated optimization" in this work. For the optimization task, the existence of the heater before each membrane stage is optimized but the temperature is fixed at 343 *K* which is the maximum tested temperature in the experiment (Tsuyumoto et al., 1997).

The main optimization results are shown in Table 1. Due to space limitation, the existence of the heater before each membrane stage is not shown but all the optimization results show that heaters should exist between n = 2 to n = 5 where possible. The first stage does not require a heater as a subcooled condenser is used in the distillation column to cool the distillate to 343 *K*. From stage six onwards

(optimization with $n = 6,7,8$), the temperature drop across the membrane stages are low, thus the feed heaters are not needed. The best design obtained using the repeated optimization method is when n = 5 with TAC as 0.7573 $M \$ y^{-1}$. The simultaneous optimization strategy recommended in this work gave the optimal structure when n = 6 with a TAC of 0.7577 $M \$ y^{-1}$, with slightly different column and membrane system structures when compared to the repeated optimization, showing that the proposed optimization strategy in this work is reliable and accurate. Moreover, the recommended optimization strategy is more time efficient by considering the total CPU time for the optimization task where the proposed optimization strategy can save 77% time. (This time saving is underestimated as the time taken for the optimization for $n = 1,2,3$ were also performed but not considered as they could not achieve the product purities.)

4. Conclusions

This work proposes a superstructure optimization strategy for the optimal design of hybrid distillation/pervaporation processes, and discusses different solution alternatives for how to handle the integer nature of the membrane network, as well as proposes a procedure for systematic initialization, simulation, and optimization of the process. The optimization strategy is applied to a case study considering a binary azeotropic separation. The optimization results obtained are compared to solution by repeated optimization (optimize the superstructure at each number of membrane stages). The superstructure optimization strategy is found to be superiority in terms of CPU time (at least 77% time saving) given its ability to simultaneously optimize the distillation column and membrane system superstructures. Theoretically, for membrane systems with recycle streams, the same methodology can be applied but some modifications (e.g., adding stream selectors) may be required which will increase the computational difficulty, and the performance of this methodology will be tested in future work.

References

Chia, D. N., Duanmu, F., and Sorensen, E. (2021). Optimal Design of Distillation Columns Using a Combined Optimisation Approach. In Turkay, M. and Gani, R., editors, *31st European Symposium on Computer Aided Process Engineering*, pages 153–158. Elsevier B.V.

Koch, K., Sudhoff, D., Kreiß, S., Gorak, A., and Kreis, P. (2013). Optimisation-based design method´ for membrane-assisted separation processes. *Chemical Engineering and Processing: Process Intensification*, 67:2–15.

Marriott, J. and Sorensen, E. (2003). The optimal design of membrane systems. *Chemical Engineering Science*, 58(22):4991–5004.

Process Systems Enterprise (2019). gO:MATLAB.

Process Systems Enterprise (2020). gPROMS ProcessBuilder version 1.4.

Seider, W. D., Lewin, D. R., Seader, J. D., Widagdo, S., Gani, R., and Ng, K. M. (2016). *Product and Process Design Principles: Synthesis, Analysis and Evaluation*. Wiley, 4 edition.

Singh, A. and Rangaiah, G. P. (2019). Development and optimization of a novel process of doubleeffect distillation with vapor recompression for bioethanol recovery and vapor permeation for bioethanol dehydration. *Journal of Chemical Technology & Biotechnology*, 94(4):1041–1056.

Sinnott, R. and Towler, G. (2020). *Chemical Engineering Design*. Elsevier, 6 edition.

Tsuyumoto, M., Teramoto, A., and Meares, P. (1997). Dehydration of ethanol on a pilot-plant scale, using a new type of hollow-fiber membrane. *Journal of Membrane Science*, 133(1):83–94.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50053-1

Design and analysis of a single mixed refrigerant natural gas liquefaction process integrated with ethane recovery and carbon removal using cryogenic distillation

Ting He[a,b], Truls Gundersen[a], Wensheng Lin[b*]

[a] *Department of Energy and Process Engineering, Norwegian University of Science and Technology (NTNU), Kolbjoern Hejes v. 1A, NO-7491 Trondheim, Norway*
[b] *Institute of Refrigeration and Cryogenics, Shanghai Jiao Tong University, Shanghai 200240, China*
linwsh@sjtu.edu.cn

Abstract

Currently, the CO_2 purification specification for natural gas liquefaction is fixed as 50 ppm based on the solubility of CO_2 in liquid methane. However, for unconventional natural gas with high ethane content like shale gas and oilfield associated gas, the CO_2 solubility in these cryogenic fluids may increase considerably due to the azeotropic properties of ethane-CO_2 mixture. In this study, a novel integrated process is proposed to simultaneously realize natural gas liquefaction, ethane recovery and CO_2 separation, in which high purity methane and ethane products are obtained through a cryogenic distillation column and an extractive distillation column. The proposed process with refrigeration supplied by a single mixed refrigerant (SMR) cycle is designed, optimized, and comprehensively evaluated through performance indicators such as specific energy consumption, exergy efficiency, CO_2 removal rate as well as ethane recovery rate. Based on a thermodynamic analysis of the CH_4-CO_2-C_2H_6 ternary mixture, the maximum allowable CO_2 content in a feed gas with 2 - 20 mol% ethane is 1.8 - 17 mol%, which is much larger than 50 ppm. In addition, the recovery rate and purity of the ethane product reached 99.5% with a CO_2 removal rate larger than 99.3%. The results show that the specific power consumption and system exergy efficiency corresponding to the maximum allowable CO_2 content are around 0.41 kWh/Nm^3(NG) and 53.1 - 56.4 %, respectively.

Keywords: natural gas liquefaction, ethane recovery, CO_2 removal, cryogenic distillation, extractive distillation

1. Introduction

The world energy system is accelerating the transition to a clean and efficient energy system, and natural gas will play an important role in this transition process before large-scale application of renewable energies. In recent years, unconventional natural gas, has promoted the rapid growth of world natural gas production. In particular, shale gas has successfully transformed the United States from a natural gas importer to an exporter (Shcherba et al., 2019). Unlike conventional natural gas, the ethane content of shale gas in the US is significantly higher (Kort et al., 2016). As an important raw material for ethylene (Yang and You, 2017), the ethane recovery from shale gas can provide additional revenue. The recovery of ethane from natural gas usually requires cryogenic distillation, and this process consumes a large amount of cold energy. If it can be integrated with other

parts in the natural gas chain, for example natural gas liquefaction, considerable investment reductions and energy savings can be achieved through integration within the process and the energy system (Ansarinasab and Mehrpooya., 2017). For ethane recovery, the integration of various natural gas liquefaction processes has been considered in our previous studies (He and Lin, 2020), and the results show that when the ethane content in the feed gas is 10-40 mol%, the proposed processes realized desirable separation effects, with both the purity and recovery rate of ethane reaching 99.5%. Besides, much attention has been paid to the integration of natural gas liquefaction and natural gas liquids (NGL) recovery. Vatani et al., (2013) also proposed an integrated process system for NGL-LNG co-production, and when it is applied to a typical feed gas rich in heavy hydrocarbons (75 mol% methane and 23 mol% heavy hydrocarbons), the specific power consumption is 0.414 kWh/kg (LNG).

In general, natural gas contains a certain amount of CO_2 that is causing calorific value reduction, equipment corrosion, even blockage in cryogenic conditions (Park et al., 2021). Thus, strict standards for carbon content are set for commercial natural gas, which lead to the development of carbon removal technologies for natural gas. The widely used methods for carbon removal in natural gas include physical absorption, chemical absorption, adsorption, cryogenic and membrane technologies (Babar et al., 2019). For LNG production, the purification specification of 50 ppm makes most carbon removal methods unsuitable, while chemical absorption and cryogenic methods (Baccioli et al., 2018) stand out. Although cryogenic carbon removal is considered to be environmentally friendly, it has not been widely used due to high energy consumption. More importantly, the freeze-out problem of CO_2 during cryogenic processes brings another challenge. However, the disadvantage in energy consumption can be overcome if it can be combined with the natural gas liquefaction process (Lin et al., 2018). In addition, because of the azeotropic properties of ethane and CO_2 (Gugnoni et al., 1974), the problem of blockage inside the distillation column due to CO_2 freeze-out may also be solved. However, there are few reports on the natural gas liquefaction process integrated with cryogenic carbon removal, especially by distillation. Focused on natural gas with high ethane content, this study proposes a novel single mixed refrigerant (SMR) liquefaction process integrated with ethane recovery and carbon removal. The separation of CO_2 and ethane is realized through cryogenic distillation and extractive distillation.

2. Process simulation and optimization

2.1. Process description

The flow diagram of the entire process is described in Figure 1. Focusing on the liquefaction and CO_2 removal process, the upstream natural gas processing, such as water removal, are excluded in this study and the feed gas (101) is simplified as a mixture of methane, ethane and CO_2.

The feed gas (101) first passes through a multi-stage compression unit (C-101, C-102) equipped with interstage coolers, and then passes through heat exchangers HEX-101, H-101, and HEX-102 to be partially condensed. Next, it enters the cryogenic distillation column (D-101) to obtain the enriched methane flow (109). After further pressure increase by the cryogenic compressor (C-103), it is completely condensed in HEX-102 and subcooled in HEX-103. Finally, it enters the storage tank (T-101) after throttling. The liquid flow from D-101, a mixture of CO_2 and C_2H_6, is first throttled by valve V-201, and then provides cooling capacity in heat exchangers H-203 and H-205. Finally, it enters the

extractive distillation column (D-102), where C_2H_6 is produced in liquid form by the extractant isobutane, and CO_2 gas is obtained at the top of the column. The mixture of ethane and isobutane (205) enters the distillation column D-103 to obtain high-purity ethane, and the separated isobutane (301) enters D-102 for recycling. High-purity ethane (207) is also subject to further condensation, subcooling, throttling and finally storage as a liquid product with a pressure slightly above atmospheric. The refrigeration needed in heat exchangers and condensers in this process is provided by a standard SMR cycle (the black lines in Figure 1).

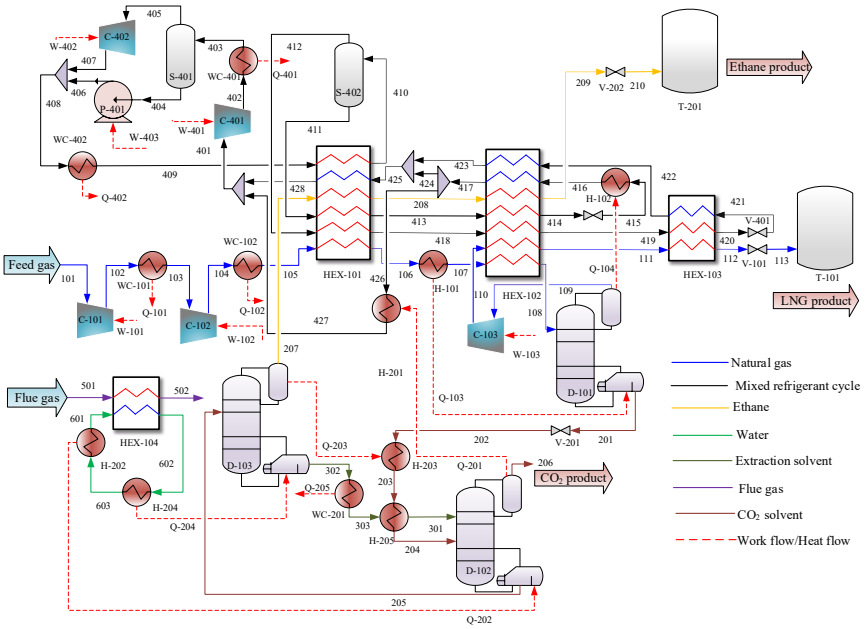


Figure. 1. Diagram of the SMR natural gas liquefaction process integrated with ethane recovery and carbon removal by cryogenic distillation. The process units are classified as follows: C: compressor, D: distillation column, H: heat exchanger, HEX: multi-stream heat exchanger, P: pump, Q: heat flow, S: separator, T: tank, V: valve, W: work; WC: water cooler

In addition to the cooling demand, the reboilers of D-102 and D-103 in this process require heat load with a temperature up to 112°C. Waste heat in the flue gas from the combustion-driven compressor unit is integrated in HEX-104 with circulating hot water that provides the required heat to the reboilers.

2.2. Initial settings and assumptions

The proposed process is modeled in Aspen HYSYS V11 by utilizing the Peng–Robinson equation of state to calculate thermodynamic properties of the feed gas and the mixed refrigerant. To simply the simulation, some parameters need to be set or assumed as presented in Table 1 according to initial conditions, product requirements or industry standard.

Table 1 Initial parameter settings and assumptions (Δp: Pressure drop)

Initial parameter settings	Value	Assumptions	Value
t_{101} /°C	40	Δp in water coolers /kPa	0

p_{101} /kPa	120	Δp in heat exchangers /kPa	0
n_{101} /kmol/h	1000	Δp in separator/mixer /kPa	0
Products storage pressure /kPa	120	Adiabatic efficiency of compressors	85 %
CO_2 in LNG	< 50 ppm	Adiabatic efficiency of pump	75 %
Ethane purity	> 99.5 mol%	Temperature after water cooling / °C	40

2.3. Process evaluation and optimization

In this study, the system evaluation involves the calculation of energy efficiency, carbon removal effect, ethane recovery rate, etc. The definition of each performance indicator used is shown in Table 2.

Table 2 Definition of evaluation indexes

Evaluation index	Definition	Annotation
ideal minimum work ($W_{\min}$)	$W_{min} = W_l + W_s$	W_l: minimum theoretical liquefaction work, kW;
total power consumption (W)	$W = \sum W_P + W_C$	W_S: minimum theoretical separation work, kW;
specific power consumption (w)	$w = \frac{\sum W_P + W_C}{N_{NG} V_M}$	W_P: power consumption of pump, kW;
exergy efficiency (η)	$\eta = \frac{W_{\min}}{W} = \frac{W_l + W_s}{W}$	W_C: total power consumption of compressors, kW;
methane loss rate (α)	$\alpha = \left(1 - \frac{N_{109} C_{109,\,C1}}{N_{101} C_{101,\,C1}}\right) \times 100\%$	V_M: nominal molar volume, $Nm^3/kmol$;
CO_2 removal rate (β)	$\beta = \frac{N_{206} C_{206,\,CO2}}{N_{101} C_{101,\,CO2}} \times 100\%$	N: molar flow, kmol/h;
ethane recovery rate (γ)	$\gamma = \frac{N_{207} C_{207,\,C2}}{N_{101} C_{101,\,C2}} \times 100\%$	C: mole fraction

In this study, sequential search and a genetic algorithm (GA) are combined to find the optimal solution for the parameters that have an influence on the energy consumption of the proposed process. The objective function is minimum specific power consumption.

3. Results and discussion

3.1. Calculation of the maximum allowable CO_2 content

First, this study analyzes the maximum allowable CO_2 content under different ethane fractions by comparing the CO_2 freeze-out temperature in both gas and liquid phases with tray temperature under given operating conditions. To be specific, for a certain ethane content, the CO_2 fraction in the feed gas is gradually increased, and then the freeze-out temperature of CO_2 is calculated based on a thermodynamic analysis of the CH_4-CO_2-C_2H_6 ternary mixture. By determining whether the CO_2 freeze-out temperature is lower than the tray temperatures of the distillation column, the corresponding maximum allowable CO_2 content without blockage due to freeze-out can be found in Table 3.

Table 3 Maximum allowable CO_2 content under different ethane contents

Ethane content	0.02	0.05	0.1	0.15	0.2
Maximum allowable CO_2 content	0.018	0.07	0.11	0.14	0.17

3.2. Process optimization results

After obtaining the maximum allowable CO_2 content, the process is optimized, and the results for a typical feed gas (15 mol% C_2H_6, 14 mol% CO_2) are presented in Table 4.

Table 4 Optimization results

Stream	t (°C)	p (Pa)	N (kmol/h)	Stream	t (°C)	p (kPa)	N (kmol/h)

101	40.0	120	1000	302	112.4	2500	247
105	40.0	3600	1000	303	40.0	2500	247
108	-72.0	3600	1000	401	34.8	183	2679
109	-92.0	3550	710	403	40.0	1000	2679
110	-75.7	4600	710	409	40.0	2688	2679
111	-95.0	4600	710	410	26.0	2688	2679
112	-161.3	4600	710	414	-96.0	2688	1485
113	-159.7	120	710	415	-96.1	183	1485
201	-2.1	3600	290	420	-160.0	2688	1194
202	-16.2	2500	290	421	-164.6	183	1194
203	-14.8	2500	290	426	-45.9	183	290
204	30.0	2500	290	427	22.0	183	290
205	54.1	2500	397	428	36.9	183	2389
206	-12.8	2400	140	501	500.0	120	1505
207	-0.2	2400	150	502	283.9	120	1505
209	-89.0	2400	150	601	120.2	200	1150
210	-88.1	120	150	602	122.0	200	1150
301	20.0	2500	247	603	120.2	200	1150

3.3. Process performance

Table 5 shows the performance indicators of the proposed process under different feed gas conditions (the CO_2 content is the maximum allowable). It can be seen that the proposed process can remove more than 99.3% of the CO_2 and recover 99.5% of high-purity ethane with very little methane loss, which indicates that this carbon removal method has obvious advantages over other methods like membrane separation. In addition, with increased contents of ethane and CO_2, the minimum theoretical work gradually decreases, while the actual work consumed by the system rises slightly, which leads to a slight decline in the exergy efficiency.

Table 5 System performance indexes of the optimal state

Ethan content	0.02	0.05	0.10	0.15	0.20
W_{min} (kW)	5137	5117	5059	4983	4883
W (kW)	9116	9120	9260	9250	9189
η	0.5635	0.5611	0.5463	0.5387	0.5314
w (kWh/Nm^3(NG))	0.4070	0.4071	0.4134	0.4129	0.4102
α (%)	0.28	0.11	0.41	0.48	0.45
β (%)	99.31	99.60	99.52	99.44	99.40
γ (%)	99.50	99.51	99.50	99.50	99.51

According to previous research results that only consider natural gas liquefaction and ethane recovery, the liquefaction power consumption is 0.38 - 0.42 kWh/Nm^3(NG) (He and Lin, 2020). Therefore, from the energy perspective, the CO_2 removal process proposed in this study increase the energy consumption of the liquefaction system only marginally through reasonable system integration. When adopting the most widely used chemical absorption method, although the additional power consumption is not very large, an additional heat load of 2.2 - 2.5 MJ/kg CO_2 for solvent regeneration is needed (Baccioli et al., 2018). As for equipment required, chemical absorption processes require at least two columns, one for the CO_2 absorption and the other for the absorbent regeneration. If ethane recovery is considered, one more cryogenic distillation column is also required. So, the required main equipment for the two methods are similar. Besides, the solvent required in this study is hydrocarbons, which can be directly obtained from natural gas, while the chemical absorption method requires a large amount of absorbent, thereby increasing its cost.

4. Conclusion

In this study, a novel integrated Single Mixed Refrigerant (SMR) natural gas liquefaction process is proposed, which combines cryogenic distillation and extractive distillation to realize ethane recovery and CO_2 removal. The proposed process is designed and optimized using Aspen HYSYS and Matlab. The results show that the process can handle a maximum allowable CO_2 content of 1.8 - 17 mol% when the ethane fraction is 2 - 20 mol %. More than 99.3% of the CO_2 can be removed with very little methane loss and over 99.5% of the ethane can be recovered as a high-purity product. The specific power consumption corresponding to the maximum allowable CO_2 content is about 0.41 kWh/Nm^3(NG), and the system exergy efficiency is in the range 53.1 - 56.4 %.

Acknowledgement

This publication has been funded by HighEFF - Centre for an Energy-Efficient and Competitive Industry for the Future. The authors gratefully acknowledge the financial support from the Research Council of Norway and user partners of HighEFF, an 8-years Research Centre under the FME-scheme (Centre for Environment-friendly Energy Research, 257632).

The authors gratefully acknowledge the financial support from China Scholarship Council (CSC) during Ting He's visit to Norwegian University of Science and Technology.

References

H. Ansarinasab, M. Mehrpooya. 2017. Evaluation of novel process configurations for coproduction of LNG and NGL using advanced exergoeconomic analysis, Applied Thermal Engineering, 115, 885-898.

M. Babar, M. A. Bustam, A. Ali, A. Shah Maulud, U. Shafiq, A. Mukhtar, S. N. Shah, K. Maqsood, N. Mellon, A. M. Shariff. 2019. Thermodynamic data for cryogenic carbon dioxide capture from natural gas: A review, Cryogenics, 102, 85-104.

A. Baccioli, M. Antonelli, S. Frigo, U. Desideri, G. Pasini. 2018. Small scale bio-LNG plant: Comparison of different biogas upgrading techniques, Applied Energy, 217, 328-335.

R. J. Gugnoni, J. W. Eldridge, V. C. Okay, T. J. Lee. 1974. Carbon dioxide-ethane phase equilibrium, AlChE Journal, 20, 357-362.

T. He, W. S. Lin. 2020. A novel propane pre-cooled mixed refrigerant process for coproduction of LNG and high purity ethane, Energy, 202, 117784.

E. A. Kort, M. L. Smith, L. T. Murray, A. Gvakharia, A. R. Brandt, J. Peischl, T. B. Ryerson, C. Sweeney, K. Travis. 2016. Fugitive emissions from the Bakken shale illustrate role of shale production in global ethane shift, Geophysical Research Letters, 43, 9, 4617-4623.

W. S. Lin, X. J. Xiong, A. Z. Gu. 2018. Optimization and thermodynamic analysis of a cascade PLNG (pressurized liquefied natural gas) process with CO_2 cryogenic removal, Energy, 161, 870-877.

J. Park, S. Yoon, S.Y. Oh, Y. Kim, J.K. Kim. 2021. Improving energy efficiency for a low-temperature CO_2 separation process in natural gas processing, Energy, 214, 118844.

V. A. Shcherba, A. P. Butolin, A. Zieliński. 2019. Current state and prospects of shale gas production, IOP Conference Series: Earth and Environmental Science, 272, 032020.

A. Vatani, M. Mehrpooya, B. Tirandazi. 2013. A novel process configuration for co-production of NGL and LNG with low energy requirement, Chemical Engineering and Processing: Process Intensification, 63, 16-24.

M. Yang, F. You. 2017. Process Modeling and analysis of manufacturing pathways for producing ethylene and propylene from wet shale gas and naphtha, Computer Aided Chemical Engineering, 40, 361-366.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50054-3

A new decomposition approach for synthesis of heat exchanger network with detailed heat exchanger sizing

Zekun Yang[a], Nan Zhang[a*] and Robin Smith[a]

[a]*Centre for Process Integration, Department of Chemical Engineering and Analytical Science, The University of Manchester, Manchester, M13 9PL, UK*
nan.zhang@manchester.ac.uk

Abstract

Due to the complexities arisen from the non-convexities in the mathematical models for the HEN synthesis incorporating detailed exchanger design, constant heat transfer coefficients and short-cut model for the calculation of exchanger capital cost are used for a majority of approaches to obtain a synthetic network topology, which causes inaccurate heat transfer areas and trade-offs between energy usage and capital investment. This paper presents an enhanced iterative-based decomposition algorithm to achieve realistic HEN synthesis with detailed heat exchanger sizing, which targets to overcome the drawbacks associated with the use of short-cut heat exchanger model in configuration synthesis, and further presents how these exchanger details can be employed to lead the HEN synthesis towards generating more cost effective solutions. Fouled individual stream heat transfer coefficients and corrected total process cost are updated iteratively between heat exchanger design (HED) and HEN superstructure (HENS) to guide HEN topology optimization. Global optimization for heat exchanger sizing is achieved in each iteration using a global solver BARON/GAMS.34 to overcome instability in the iteration process caused by local optimum issues. A case study shows that it can provide a better solution than the results in the literature with a lower total annual cost and computational time.

Keywords: Heat exchanger network synthesis, Detailed heat exchanger sizing, Mathematical programming, Optimization, Process synthesis

1. Introduction

The increased pressure of reducing carbon emissions in the worldwide chemical industries leads to rising awareness for incorporating cost-effective ways of saving energy. Heat exchanger networks (HEN) are essential in the process industries, since they can improve energy efficiency and reduce “greenhouse” gas emissions by heat integration of process heat sources and sinks to reduce utility consumptions. The approaches used in HEN synthesis can be categorized into sequential and simultaneous methods. Pinch Technology has been developed based on the sequential thermodynamic analysis, but it requires experienced designers and may lead to missing promising solutions.

Mathematical programming has been developed by many researchers for simultaneous HEN synthesis. The synthesis problem is commonly formulated as a superstructure, in which the HEN topologies, stage temperatures, utilities and heat duties can be optimized simultaneously. A widely used stage-wise superstructure (SWS) was proposed by Yee and Grossmann (1990). By using SWS, the HEN synthesis is formulated as an MINLP problem, targeting the minimum total annual cost. In addition, several other approaches

have been reported to solve the HEN synthesis problem by using different algorithms, such as stochastic algorithm (Rathjens and Fieg, 2020) and a deterministic approach (Yang et al., 2021).

Understandably, to achieve industrial applications, practical considerations related to heat exchanger details are significant for HEN synthesis, as short-cut heat exchanger calculations can lead to impractical design. More recently, simultaneous and iterative-based decomposition approaches have been developed to enhance the HEN synthesis towards realistic design by bringing heat exchanger details in the HEN optimization. Xiao et al. (2019) adopted a simultaneous approach based on a hybrid GA/SA algorithm. But their implementation coupled with stochastic algorithms inclines to generate a local HEN solution as requiring a relatively high utility usage.

Alternative to the simultaneous optimization, iterative-based decomposition strategies deal with detailed heat exchanger design as an individual block to avoid the massive nonlinearities in the MINLP HEN superstructure, which can helps to reduce the combinational difficulties for a large-scale MINLP problem. Ravagnani and Caballero (2007) presented a heuristic decomposition method to update stream heat transfer coefficients from heat exchanger design to HEN superstructure. Short et al. (2016) and Kazi et al. (2021) proposed a two-step optimization procedure, in which several correction factors were introduced to the iterative procedures to correct investments of heat exchangers in HENS. But their method brings two certain problems, including (1) tricky convergence of the proposed algorithm with many preliminary iterations; (2) difficulty to solve large-scale problems. This work addresses the existing problems identified from the literature, and proposes a novel iterative-based decomposition algorithm that integrates a heat exchanger network superstructure (HENS) (Yang et al., 2021) and heat exchanger design (Yang et al., 2020).

2. Mythologies

2.1. Global optimization for detailed heat exchanger sizing

In this work, we focus on Shell and Tube Heat Exchangers (STHEs) with plain tubes and single segmental baffles. The mathematical optimization model has been proposed in previous work (Yang et al., 2021). The geometrical variables include tube pinch Pt, tube number N_t, tube length L, tube inside and outside diameters D_i, D_O , shell inner diameter D_{SI}, tube outside bundle diameter D_{SB}, baffle spacing B_S, baffle cut B_c, baffle number N_b, the number of tube passes N_{TP}, the number of shell passes N_S. Some discrete decisions are formulated by generalized disjunction programming, including the selection of tube passes, tube sizes, tube angle arrangement. These geometries are optimized for their impacts on heat transfer coefficients, pressure drops and heat transfer area, which is guided by the constraints associated with Tubular Exchanger Manufactures Association (TEMA) standards. The objective function is to minimize the total exchanger cost.

Notably, global optimization plays an important role in the iterative method. Local optimum for individual heat exchanger design could mislead HEN synthesis solutions and cause convergence issues in the iterative algorithm. A case study (Yang et al., 2020) is tested in this work to investigate the need of global optimization. Fig.1 shows the result comparison of using the MINLP local solver DICOPT and global solver BARON, by GAMS, with 38.5 % of total exchanger cost savings from the global optimization and acceptable CPU time of 292 s. Consequently, the application of global optimization in

HED can lead to the cost of exchanger always being minimized and promoting the iterative process to reach convergence rapidly because of consistent corrections.

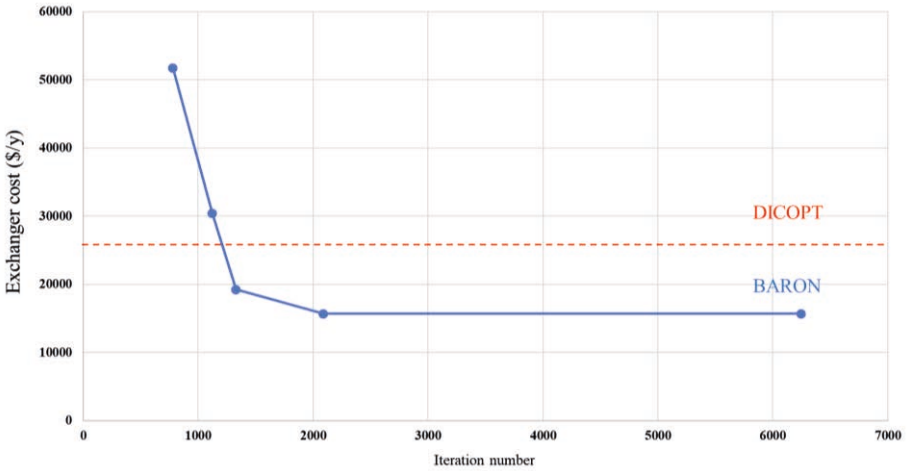


Figure 1. Comparison of problem-solving process between DICOPT and BARON

2.2. HEN superstructure optimization

The mathematical model for HEN topology optimization is based on the well-known stage-wise superstructure (SWS) proposed by Yee and Grossmann (1990), and stages are introduced, in which all possible matches between hot streams and cold streams are optimized simultaneously. By adopting the enhanced deterministic-based approach proposed in the earlier work (Yang et al., 2021), a cost effective solution can be targeted with low computational time. The iso-thermal mixing assumed in the original SWS method is removed, and additional constraints are employed to model non-isothermal mixing. This proposed model is formulated as a non-convex MINLP problem that is solved by the global solver BARON/GAMS.

2.3. Iterative algorithm

The proposed algorithm adopts a modification of the heuristic-based decomposition approach proposed by Ravagnani and Caballero (2007), which integrates heat exchanger design model (Section 2.1) for detailed exchanger geometries, heat transfer coefficients, pressure drops and cost details, and the HEN superstructure approach (Section 2.2) for optimal HEN configurations, heat duty for each exchanger, inlet and outlet temperatures of each exchanger and split ratios for mass flow rate calculation.

In this work, a correction factor for total process cost (F_{TPC}) is introduced to correct the deviation of total process cost derived by multiple shell passes, LMTD correction factor (F_T), geometries, tube and shell side pressure drops. Furthermore, convergence criteria are introduced to select optimal results, including the percentage differences of stream fouled stream heat transfer coefficient (Re_{HTC}) and total process cost (Re_{TPC}) between that are used in HENS and obtained by HED at an individual iteration. Re_{HTC} and Re_{TPC} are employed to reflect the stability of the iteration process. They are able to indicate how the corrections work in the iterative procedures and guide to generate realistic solutions, which is more accurate to represent the level of convergence. The tolerances of both Re_{HTC} and Re_{TPC} can be specified, for which 5% for each individual iteration seems to be

reasonable in our test cases. Fig.2 shows the scheme of this approach, including seven steps, as described:

Step-1: Define process parameters and constant heat transfer coefficients in HENS. The constant heat transfer coefficients can be supplied by the program or generally assumed at a range from 0.5 kW/m^2°C to 1 kW/m^2°C. Next, generate an initial HEN configuration considering stream splitting.

Step-2: According to heat duty allocations and temperatures from the initial HEN configuration, optimize each heat exchanger design by solving the MINLP HED model.

Step-3: By using HED, calculate the fouled HTC of each stream, using an average value of fouled tube side (shell side) HTC h_{TF} (h_{SF}) from exchangers that are installed for an individual stream. Meanwhile, calculate the correction factor for total annual process cost by $F_{TPC} = TPC_{HE}/TPC_{HEN}$.

Step-4: Update these calculated fouled HTC and F_{TPC} in HENS. Solve the HENS to generate optimal HEN configuration. In this step, the maximum acceptable computational time could be specified in HENS.

Step-5: Use MINLP HED mode to optimize each heat exchanger and calculate relatively errors Re_{HTC} and Re_{TPC}. Meanwhile, calculate the total annual cost TAC_{HE} of the HEN with detailed heat exchanger design.

Step-6: Check if the relatively error Re_{HTC}, Re_{TPC} are lower than the tolerance. If yes, go to Step-7. Otherwise, go to Step-3.

Step-7: Check if the TAC_{HE} is higher than the current one. If yes, stop and output the current HEN as the optimal result. Otherwise, replace the previous HEN result and return to Step 3.

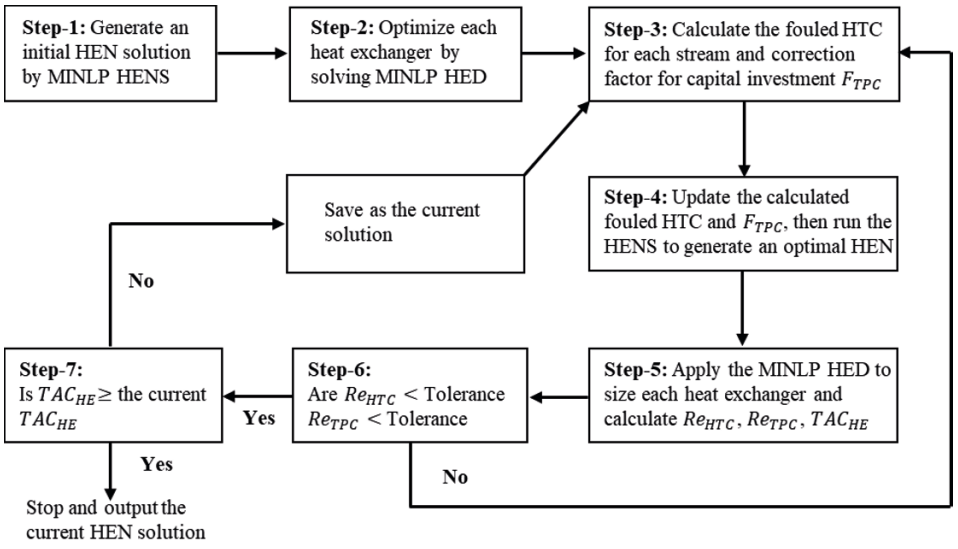


Figure 2. Iterative decomposition algorithm

3. Case study

This case study is taken from the literature (Ravagnani and Caballero, 2007; Xiao et al., 2019; Short et al., 2016; Kazi et al., 2021). It was solved on a computer resource Intel ® Core™, I7-8700 CPU, 3.20 GHz with 16 GB RAM, 6 cores, 12 processors. Global solver BARON in GAMS.34 was applied. Constant heat transfer coefficients of 0.888 kW/m^2°C are assumed for the initial synthesis.

At the initial design, the practical consideration of multiple shells and pump operating cost associated with detailed HED model leads to the total process cost (TPC) to be much higher than that from the HENS design with constant heat transfer coefficients. Large stream Re_{HTC} mean that the assumed heat transfer coefficients need to be corrected. Iteratively, Re_{HTC} reduces gradually from iteration-1 to iteration-4, when it is less than the specified tolerance of 5 %. F_{TPC} is first calculated at the iteration-1, as 2.829, then updated from iteration-2. From that point, the deviation of TPC between HED and HENS is decreased promptly under an allowed Re_{TPC} (within 5 %). The final optimal TPC-HED is 2.985 times the initial TPC-HENS with assumed coefficients. Using the proposed approach, the best solution was found at iteration-4 with 5760 s CPU time, which brings 77% of computational time saving compared with that demanded by Kazi et al. (2021).

Fig.3 illustrates the optimal HEN configuration of the case study. The comparison of results is presented in Table 1. Compared with the best solution so far (Kazi et al., 2021), the proposed approach led a lower TAC with 3,620,095 $/y. The TAC saving is not significant, because the cost is dominated by energy cost (Total operating cost/TAC > 0.95). But for the total process cost, 26.5 % of TPC is achieved, because considering the trade-off between area cost and pumps cost in heat exchanger optimization, the better utilization efficiency of exchanger geometries that are guided by the proposed global optimization, generates 58.4 % of total process pump cost saving even using the similar total process area.

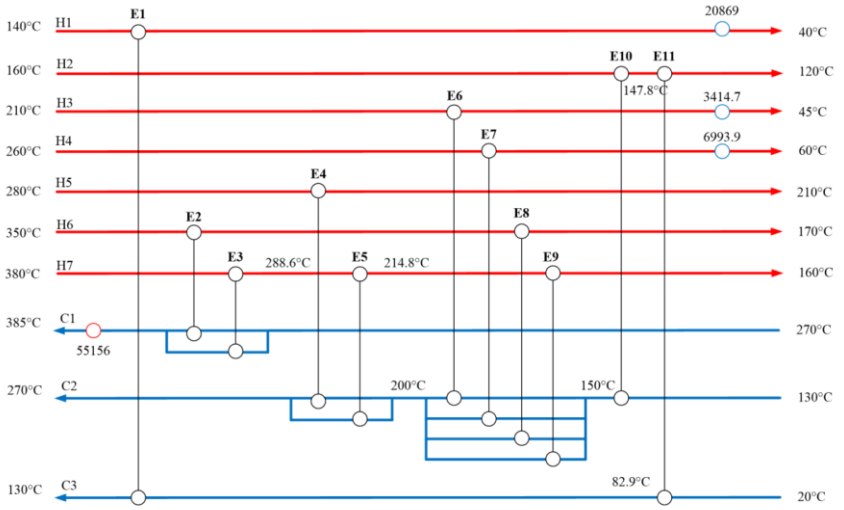


Figure 3. Optimal HEN solution of Example

Table 1: Results comparison

	Short et al.	Xiao et al.	Kazi et al.	This work
Hot utility usage (kW)	64,187	61,063	55,156	55,156
Cold utility usage (kW)	40,299	37,175	31,267	31,267
Total utility cost ($/y)	4,091,975	3,886,803	3,496,972	3,496,972
Total process area (m^2)	5451	10754	12,151	12,187
The number of exchangers	22	18	38	35
Process capital cost ($/y)	44,998	45,747	99,705	94,993
Process pump cost ($/y)	46,099	6907	67,692	28,129
Total process cost ($/y)	91,097	52,654	167,397	123,123
Total annual cost ($/y)	4,183,072	3,939,457	3,664,369	3,620,095

4. Conclusions

This work proposed an iterative approach for the heat exchanger network synthesis with detailed heat exchanger sizing. Corrected stream heat transfer coefficients and process investment were determined by global optimization of STHE-HED through an iterative procedure. The proposed method shows better performance than the existing approaches, with 1.2 % savings of total annual cost (TAC), 58.4 % savings of pump cost and 26.5 % savings of total process cost. This methodology offered a time-efficient way towards a cost effective and practical HEN design with quick convergence, further improving the feasibility for solving industrial-scale problems.

References

M.A.S.S. Ravagnani, J.A. Caballero, Optimal heat exchanger network synthesis with the detailed heat transfer equipment design, Computers & Chemical Engineering 31 (2007) 1432-1448.

M. Rathjens, G. Fieg, A novel hybrid strategy for cost-optimal heat exchanger network synthesis suited for large-scale problems, Applied Thermal Engineering 167 (2020) 114771.

M. Short, A.J. Isafiade, D.M. Fraser, Z. Kravanja, Synthesis of heat exchanger networks using mathematical programming and heuristics in a two-step optimisation procedure with detailed exchanger design, Chemical Engineering Science 144 (2016) 372-385.

S.R. Kazi, M. Short, A.J. Isafiade, L.T. Biegler, Heat exchanger network synthesis with detailed exchanger designs—2. Hybrid optimization strategy for synthesis of heat exchanger networks, AIChE Journal 67 (2021) e17057.

T.F. Yee, I.E. Grossmann, Simultaneous optimization models for heat integration—II. Heat exchanger network synthesis, Computers & Chemical Engineering 14 (1990) 1165-1184.

W. Xiao, K. Wang, X. Jiang, X. Li, X. Wu, Z. Hao, G. He, Simultaneous optimization strategies for heat exchanger network synthesis and detailed shell-and-tube heat-exchanger design involving phase changes using GA/SA, Energy 183 (2019) 1166-1177.

Z. Yang, Y. Ma, N. Zhang, R. Smith, Design optimization of shell and tube heat exchangers sizing with heat transfer enhancement, Computers & Chemical Engineering 137 (2020) 106821.

Z. Yang, N. Zhang, R. Smith, Enhanced deterministic approach for heat exchanger network synthesis, Computer Aided Chemical Engineering, (2021), vol. 50, pp. 833-838. Elsevier.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50055-5

A mathematical approach for the synthesis of a wastewater treatment plant using the concept of circular economy

Jo Yee Ho[a], Wai Teng Tee[a], Yoke Kin Wan[a*]

[a] *Department of Chemical and Environmental Engineering, University of Nottingham Malaysia, Broga Road, 43500 Semenyih, Selangor, Malaysia.*

yokekin.wan@nottingham.edu.my

Abstract

Huge generation of waste from industrial manufacturing processes has become a concern to many countries especially in the world with finite resources. Among these wastes, wastewater generation is one of the biggest issues faced by most industrial processes as the treatment of these wastewaters requires different treatment stages. Consequently, improper treatment and direct discharge of wastewater often occurs which had contaminated the world's waterways. Therefore, stricter environmental discharge regulations had been enforced by local government authorities. This becomes a challenge for new manufacturing plants in designing their wastewater treatment process to comply with the government regulations set while reducing environmental impacts. This paper presents the preliminary evaluation of a model-based decision making on wastewater treatment technologies selection based on the concept of circular economy. To prolong the natural water cycle, treated wastewater were recycled back to the manufacturing process which reduces freshwater consumption. A case study on semiconductor manufacturing process and its wastewater treatment plant is solved in this work. Based on the results, the synthesis of the wastewater treatment plant incorporating circular economy has obtained 55.83% circular economy efficiency of water being recycled back to the semiconductor manufacturing process.

Keywords: Wastewater treatment process, Circular Economy, Mathematical model.

1. Introduction

Circular economy (CE) is a well-established concept in encouraging sustainable development initiatives. The CE mainly focuses on a perfectly balanced operation by promoting and utilizing renewable energy resources such as biomass, water, and solar. It replaces the 'end-of-life' concept by restoration and turns the goods and services into alternative resources with minimum waste leakage and toxic chemicals. A circular economy-orientated business model prioritizes reusing, refurbishing, remanufacturing, recycling, and repairing the waste creation after the consumption stage instead of discarding them into landfills (Pires and Martinho, 2019). As a result, these materials and products can be productively used repeatedly, thereby increasing and retaining the value of the products. In this respect, this research aims to develop a preliminary decision-making tool integrating with the concept of circular economy to prolong and sustain the natural water cycle. Unlike linear manufacturing process, a transition towards a circular economy model will maximise the circularity of water in the system which reduces the discharge of wastewater while minimising the use of natural resources. The circularity metrics can be categorised into circularity measurement indices and circularity

assessment tools. In this work, circularity measurement indices will be applied to directly determine the circularity of water for a new manufacturing plant. This way, the selection of wastewater treatment technologies will ensure maximise recovery of treated wastewater back to the manufacturing process. This work is expected to benefit industry sectors, policy makers and local government authorities on future sustainable development of new manufacturing sectors.

2. Problem statement

Figure 1 illustrates the problem statement of a wastewater treatment process for new manufacturing plants. Wastewater feed $i \epsilon I$ from manufacturing process enters a series of wastewater treatment process beginning from pre-treatment stage $a \epsilon A$, chemical treatment $b \epsilon B$, biological treatment $c \epsilon C$ and tertiary treatment $d \epsilon D$ to ensure the treated wastewater produced complies with local discharge regulation. In light of this, recycle streams are designed in these four treatment stages to maximise water reuse in the manufacturing company. Removal of contaminants will generate sludge water during the chemical treatment and biological treatment. Eventually, these sludge water will enter sludge treatment $e \epsilon E$ for wastewater removal before disposal of sludgecake.

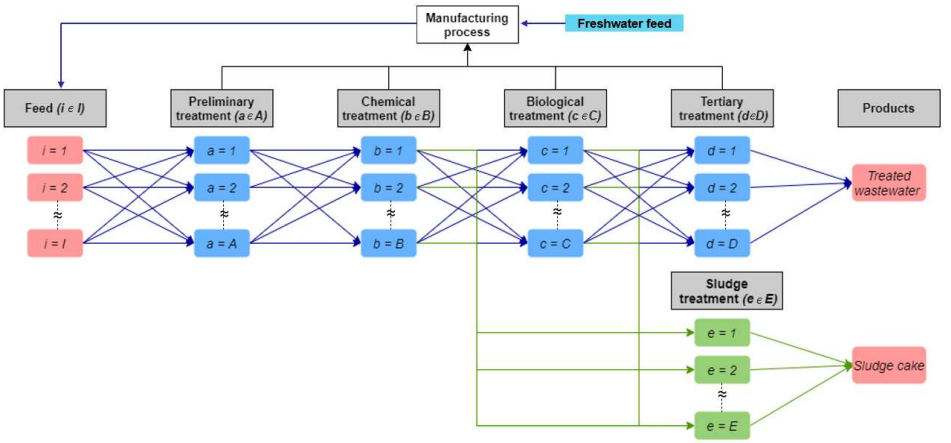


Figure 1: Generic superstructure of wastewater treatment process in a new manufacturing plant

3. Circular economy optimization model

The mathematical model developed in this model consist of flowrate balance, contaminant component balance and circular economy index formulations. By applying law of conservation of mass, the wastewater treatment flowrate balance and contaminant component balance are repetitive in each treatment stage as shown in Figure 1. Hence, Eq. (1) to Eq. (6) depicts a more generic set of equations to represent these formulations. The generic formulation appoints index t to represent preceding treatment stage, index u as current treatment stage, and index v as succeeding treatment stage respectively. For example, to formulate the equations for chemical stage b, the current index u will be chemical stage b ($u = b$), the previous index t will be preliminary treatment stage a ($t = a$) and subsequent index v will be biological treatment stage c ($v = c$). The same

formulation method is repeated for other stages. In this model, all the parameters and variables are represented as non-Italic and Italic, respectively.

3.1. Flowrate of wastewater treatment process

The flowrate balance of treatment stage *u* is summarised in Eq. (1) to Eq. (3).

$$F_u^{in} = \sum_{t=1}^{T} F_{t,u} \qquad \forall u \tag{1}$$

$$F_u^{in} = F_u^{out} \qquad \forall u \tag{2}$$

$$F_u^{out} = \sum_{v=1}^{V} F_{u,v} + F_u^{ww,recycle} \qquad \forall u \tag{3}$$

Where, F_u^{in} (m^3/day) and F_u^{out} (m^3/day) represents the inlet and outlet flowrate of wastewater at treatment stage *u*; $F_{t,u}$ (m^3/day) and $F_{u,v}$ (m^3/day) represents the flowrate transferred between the treatment stages; $F_u^{ww,recycle}$ represents the treated wastewater from treatment stage *u* that can be recycled back to the manufacturing process.

3.2. Contaminant component balance

The formulation of generic component balance for contaminant *g* at treatment stage *u* were summarized as shown in Eqs.(4) to (7). Contaminant *g* refers to any wastewater contaminant characteristics. As shown, the mass of contaminant *g* entering technology *u* $M_{g,u}^{in}$ (kg/day) depends on the concentration of contaminant *g* present in the inlet stream of technology *u*, $C_{g,u}^{in}$ (kg/m^3). At every stage of the wastewater treatment process, a certain mass of contaminant *g* will be removed from technology *u*, $M_{g,u}^{removed}$ (kg/day) based on the removal efficiency of technology *u*, $\mathrm{R}_{g,u}$ (kg contaminant/m^3 WWT). The mass of contaminant *g* discharging from technology *u*, $M_{g,u}^{out}$ (kg/day) can then be calculated. The constraint equation is shown in Eq. (8) where M_g^{std} is referring to the discharge limit set by government.

$$M_{g,u}^{in} = C_{g,u}^{in} F_u^{in} \qquad \forall g \forall u \tag{4}$$

$$M_{g,u}^{\mathrm{removed}} = M_{g,u}^{\mathrm{in}} \mathrm{R}_{g,u} \qquad \forall g \forall u \tag{5}$$

$$M_{g,u}^{in} = M_{g,u}^{removed} + \sum_{v=1}^{V} M_{g,u,v} \qquad \forall g \forall u \tag{6}$$

$$M_{g,u}^{out} = M_{g,u}^{in} - M_{g,u}^{removed} \qquad \forall g \forall u \tag{7}$$

$$M_{g,u}^{out} < \mathrm{M}_g^{std} \qquad \forall g \forall u \tag{8}$$

3.3. Circular economy efficiency index

In this research, the concept of circular economy efficiency by Molina-Moreno et al. (2017) is adapted to recover treated wastewater back into the manufacturing process. The total recycled wastewater, $F^{total,recycle}$ (m^3/day) comes from treated wastewater from pre-

treatment a, chemical treatment b, biological treatment c and tertiary treatment d as shown in Eq. (9). Water consumption required by the manufacturing process, $\mathrm{F}^{\mathrm{watercomsumption}}$ (m^3/day) can be obtained from the total recycled treated wastewater and freshwater feed as illustrated in Eq. (10). To reduce freshwater consumption, circular economy index, I^{ww} will be maximised as shown in Eqs. (11) and (12).

$$F^{total,recycle} = \sum_{a=1}^{A} F_a^{ww,recycle} + \sum_{b=1}^{B} F_b^{ww,recycle} + \sum_{c=1}^{C} F_c^{ww,recycle} + \sum_{d=1}^{D} F_d^{ww,recycle} \quad (9)$$

$$\mathrm{F}^{\mathrm{watercomsumption}} = F^{total,recycle} + F^{freshwater} \quad (10)$$

$$I^{ww} = \frac{F^{total,recycle}}{\mathrm{F}^{\mathrm{waterconsumption}}} \times 100\% \quad (11)$$

$$\text{Max } I^{ww} \quad (12)$$

4. Case study

Due to the increasing demand of electronic product and solar energy, the semiconductor industry has been expanding and increasing rapidly over the years. The manufacturing process of semiconductors involves large quantity of water which causes huge volume of wastewater being generated during the process (Huang et al., 2011). Thus, the proposed approach is illustrated in this research by using a local semiconductor manufacturing plant in Penang, Malaysia to synthesize a WWTP to maximize the recovery of treated wastewater from each treatment stage back to the manufacturing process. The wastewater discharged from the manufacturing process typically has a higher chemical oxygen demand (COD) (Lin and Kiang, 2003). Therefore, Figure 2 illustrates the case study superstructure of a wastewater treatment process in a semiconductor manufacturing plant consisting of high COD removal treatment technologies.

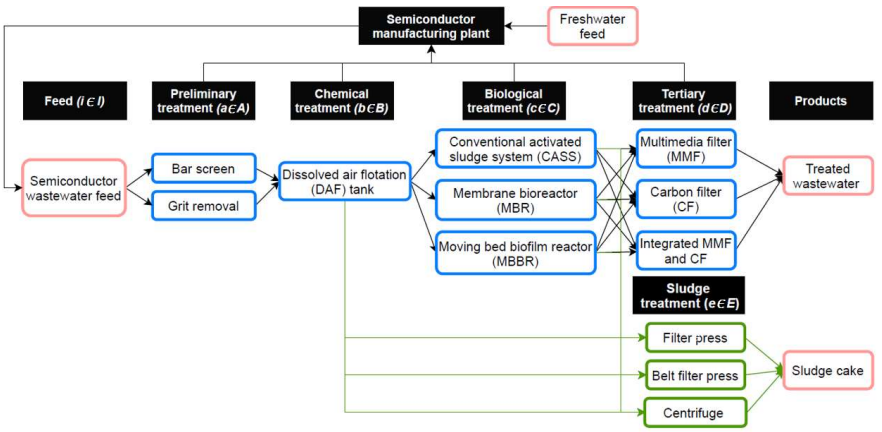


Figure 2: Case study superstructure of wastewater treatment process in a semiconductor manufacturing plant

Based on technologies selected in Figure 2, case study parameters were obtained. The average wastewater flowrate generated from the semiconductor manufacturing plant is 1,012 m^3/d. Table 1 summarise the case study wastewater characteristics from the partnered semiconductor manufacturing plant along with the local discharge regulation Standard A (Department of Environment Malaysia, 2010). Meanwhile, Table 2 and Table

3 summarises the COD removal efficiency for the case study wastewater treatment technologies.

Table 1: Case study wastewater contaminants characteristic and discharged regulations by Department of Environment (2013).

Concentration (ppm)	COD
Semiconductor wastewater	2,285
Discharge regulations (Standard A)	80

Table 2: COD removal efficiency of wastewater treatment technologies (Ho et al., 2019)

	Technologies	COD removal efficiency (%)
Preliminary treatment	Bar Screen	0
	Grit Removal	0
Chemical treatment	Dissolved air flotation (DAF)	65
Biological treatment	Conventional aerated filter (CAF)	85
	Moving bed biofilm reactor (MBBR)	90
	Membrane bioreactor (MBR)	90
Tertiary treatment	Multimedia filter (MMF)	0
	Carbon filter (CF)	50
	Integrated multimedia filter and carbon filter (MMF + CF)	65

Table 3: Dryness of sludge cake produced by each sludge treatment technology (Faure Equipments, 2018).

Technologies	Dryness (kg SS/m^3)
Filter press	25
Belt filter press	29.9
Centrifuge press	28.5

The case study is solved using the mathematical formulation developed in this research from Eqs. (1) to (12). These formulations and case study parameters from Table 1 to Table 3 were coded into a commercial optimization software, LINGO version 19 and solved using the global solver in 5 seconds. The specifications of computer used for this case study were Intel ® Core ™ i7-6500U with 8 GB RAM and x64-based processor. The developed mathematical model is a MINLP model, consisting of 194 variables, 198 constraints and 11 integers.

The global optimized results consist of bar screen, DAF, MBBR, integrated multimedia filter and carbon filter as well as belt press as shown in Figure 3. These technologies were selected due to their high COD removal efficiency to maximize the recovery of treated wastewater. Due to the compliance to wastewater discharge regulation Standard A (COD level < 80ppm), wastewater that can be recovered back to the manufacturing system must pass through all treatment stages to achieve a justifiable COD level (28 ppm). The results

have clearly shown that the total flowrate of treated wastewater recycled back is more than half of the freshwater input needed for the semiconductor manufacturing process. Therefore, this has proven that a circular economy oriented WWTP can achieve 55.58% of circular economy efficiency of treated wastewater recycled by minimizing the wastewater generation to the environment and reduce the demand of freshwater resources as well as lower the cost of freshwater input.

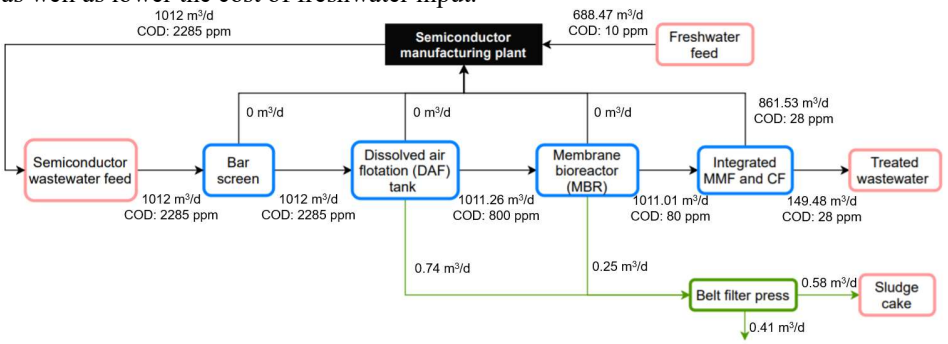


Figure 3: Synthesised WWTP for case study

5. Conclusions

Water scarcity is an important issue to be solved due to the rising number of water pollution around the world. Due to this reason, this research has incorporated the concept of circular economy in synthesizing a WWTP for industrial manufacturing process. This prominently reduces the discharge of untreated wastewater to the environment and decreases consumption of freshwater using the concept of circular economy. A case study on a local semiconductor manufacturing process is solved. By maximizing the circularity index of treated wastewater recycling, the results from the case study indicates an approximate of 45% reduction of freshwater consumption. As future work, other contaminant components such as BOD and TSS levels can be included to further enhance WWTP optimization.

References

Department of Environment Malaysia, 2010. Environmental Requirements : A guide for investors.

Equipments, F., n.d. Filter Presses Leading mechanical dewatering system.

Ho, J.Y., Wan, Y.K., Andiappan, V., Ng, D.K.S., 2019. Material Flow Cost Account-based Approach for Synthesis and Optimisation of Wastewater Treatment Plant. Chem. Eng. Trans. 76, 529–534.

Huang, C.J., Yang, B.M., Chen, K.S., Chang, C.C., Kao, C.M., 2011. Application of membrane technology on semiconductor wastewater reclamation: A pilot-scale study. Desalination 278, 203–210.

Lin, S.H., Kiang, C.D., 2003. Combined physical, chemical and biological treatments of wastewater containing organics from a semiconductor plant. J. Hazard. Mater. 97, 159–171.

Molina-Moreno, V., Carlos Leyva-Díaz, J., Llorens-Montes, F.J., Cortés-García, F.J., 2017. Design of Indicators of Circular Economy as Instruments for the Evaluation of Sustainability and Efficiency in Wastewater from Pig Farming Industry. Water 9, 1–13.

Pires, A., Martinho, G., 2019. Waste hierarchy index for circular economy in waste management. Waste Manag. 95, 298–305.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50056-7

Convex Q-learning: Reinforcement learning through convex programming

Sophie Sitter[a], Damien van de Berg[a], Max Mowbray[b], Antonio del Rio Chanona[a], Panagiotis Petsagkourakis[a*]

[a]*Sargent Center for Process Systems Engineering, Imperial College London, South Kensington Campus, SW72AZ, London, United Kingdom*
[b]*Centre for Process Integration, The Mill, University of Manchester, M1 3AL, United Kingdom*
p.petsagkourakis@ic.ac.uk

Abstract

Over the last decade, Reinforcement Learning (RL) has received significant attention as it promises novel and efficient solutions to complex control problems. This work builds on model-free RL, namely Q-learning, to determine optimal control policies for nonlinear, complex biochemical processes. We propose convex functions instead of deep neural networks as state-action value function approximators to reduce computational complexity. A convex Q-function surrogate is trained using semidefinite programming. The surrogate is then minimized to determine the optimal control action. This results in 75.3% lower computational time compared with deep Q-networks. By alleviating the computational burden of traditional RL approximation functions, this work addresses one of the major obstacles for the successful implementation of RL to real-world engineering applications.
Keywords: machine learning; convex Q-learning; semi-definite programming; data-driven batch optimization; dynamic process control;

1. Introduction

Biochemical production generally has a higher cost than its fossil fuel counterpart (del Rio-Chanona et al. 2015), for this to be sustainable, production costs must be lowered. An important aspect of this, is to lower cost by leveraging dynamic optimization with robust and effective control schemes. Yet, the dynamic optimization of highly nonlinear, complex real-world biochemical processes is often hindered by plant-model mismatch and computational intractability (del Rio-Chanona et al. 2016). Model-free Reinforcement Learning methods do not require an explicit model of the environment. They `learn' environment dynamics through data generated by environment interaction. Model-free RL is categorized into value-based, policy-based, and actor-critic optimization methods. Value-based methods seek to maximize the sum of future rewards for reaching a certain state or for taking a certain action given a state. Policy-based methods directly optimize the policy in accordance with the sampled reward values (Dong et al. 2020). The combination of the two, actor-critic, learns approximations to value functions and policy, overcoming the challenges associated with either method. Chemical processes generally provide little data. Since RL algorithms are usually regarded as "data-hungry", these applications call for the most data-efficient algorithms, namely value-based methods.

Pan et al. (2020) as well as Singh and Kodamana (2020) investigate the application of state-action value functions within Q-learning (a model-free RL method) to the dynamic optimization of batch processes. Pan et al. (2020) illustrate the superior performance of Q-learning in navigating process uncertainties through its closed-loop feedback; they address the challenge of safe reinforcement learning by proposing a chance-constrained Q-Learning algorithm based on deep Q-networks (DQN), extending the work of Mowbray et al. (2021a), and Petsagkourakis et al. (2020a). This approach mitigates the shortcomings of previously proposed ways of handling operational and safety constraints, which reintroduce model dependencies (safety filter or barrier function methods), or achieve constraint satisfaction only in expectation (penalty-adjusted rewared functions, constrained Markov Decision Processes, policy projection to safety layer) (S. Huh, I, Yang, 2020 & J. Cho et al. 2020).

However, DQNs are time-intensive to train, which severely hinders algorithm efficiency. Our work tackles this obstacle by substituting the DQNs with tractable convex functions. Finding safe Q-functions that are accurate and tractable is a considerable step towards the successful implementation of RL to real-world chemical engineering systems, and this work adds to this effort.

2. Methodology

At each iteration, data is generated which maps states x∈R^(n_x), controls u∈R^(n_u) and the respective time step t to their respective state-action value (Q-value). Q(·) denotes the empirical, cumulative cost required after enacting control u in state x, and is given by Eq. (1):

$$Q(\boldsymbol{x},\boldsymbol{u}) = \sum_{t}^{T_f} R_t \tag{1}$$

where R_t denotes the cost incurred at time step t. Next, a convex approximation $Q_{convex}(\boldsymbol{u},\boldsymbol{x};\ \boldsymbol{\theta})$ of $Q(\boldsymbol{x},\boldsymbol{u})$ is built using a combination of convex basis functions, such as those in Table 1.

To find the parameters $\boldsymbol{\theta}$ including weigh coefficients of the basis functions a nonlinear least squares optimization problem is formulated. Its objective function (Eq. 2) minimizes the error between the estimated and empirical Q-values:

$$\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum_{j} \left(Q_{convex}(\boldsymbol{u}_j,\boldsymbol{x}_j;\boldsymbol{\theta}) - Q_{data\,j}\right)^2 \tag{2}$$

where j is the iteration counter over all historic data points; and Q_{data} is an empirical estimate of the state-action value for the pair $(\boldsymbol{x}_j,\boldsymbol{u}_j)$ obtained in training via Monte Carlo simulation of the policy (discussed subsequently) under the process model. Taking advantage of the structure of the problem, to ensure positive definiteness of the norm matrices, and to maximize solution efficiency, a semidefinite program (SDP) is formulated (Vandenberghe and Boyd 1996) and solved using the Python-embedded modelling language CVXPY (S. Diamond and S. Boyd, 2016). This procedure produces the convex approximation $Q_{convex}(\boldsymbol{u},\boldsymbol{x};\boldsymbol{\theta})$ of $Q(\boldsymbol{x},\boldsymbol{u})$.

To find the optimal control action at every step, the convex Q-function approximation is optimized:

$$\pi_*(\cdot) = \begin{cases} \underset{\boldsymbol{u_t}}{\operatorname{argmin}} \quad Q_{convex}(\boldsymbol{u_t}, \boldsymbol{x_t};\, \boldsymbol{\theta}) \\ \qquad \boldsymbol{u_t} \in \mathbb{U} \subseteq \mathbb{R}^{n_u} \end{cases} \tag{3}$$

where subscript t denotes the current timestep, $\boldsymbol{x}$ and $\boldsymbol{u}$ are state and control action at t and $\mathbb{U} \subseteq \mathbb{R}^{n_u}$ defines the constraints on controls.

Table 1. Convex basis functions

Function type	Expression	Domain
Affine	$f_1(\boldsymbol{x}) = \boldsymbol{a}^T\boldsymbol{x} + b$	$\boldsymbol{x} \in \mathbb{R}^{n_x},\ \boldsymbol{a} \in \mathbb{R}^{n_x}, b \in \mathbb{R}$
Exponential	$f_2(\boldsymbol{x}) = e^{\boldsymbol{a}^T\boldsymbol{x}+b}$	$\boldsymbol{x} \in \mathbb{R}^{n_x},\ \boldsymbol{a} \in \mathbb{R}^{n_x}, b \in \mathbb{R}$
Powers	$f_3(x_i) = x_i^{\alpha_i}$	$\alpha_i \geq 1,\ x_i \in \mathbb{R}_+, \quad i = 1, \dots, n_x$
Negative entropy	$f_4(x_i) = x_i \log x_i$	$x_i \in \mathbb{R}_+, \quad i = 1, \dots, n_x$
Negative logarithms	$f_5(x_i) = -\log x_i$	$x_i \in \mathbb{R}_+, \quad i = 1, \dots, n_x$
P-norm	$f_6(\boldsymbol{x}) = \|\boldsymbol{x}\|_p$	$\boldsymbol{x} \in \mathbb{R}^{n_x}$
Quadratic over linear	$f_7(x_i, y_i) = \frac{x_i^2}{y_i}$	$x_i, \mathrm{y_i} \in \mathbb{R}_+, \quad i = 1, \dots, n_x$

Notes on implementation

The algorithm implementation can be found in Algorithm 1. We distinguish between pre-training (steps 1-3) and main training (steps 4-11). In pre-training, the Q-function approximation is fitted off-line using numerical simulations. This pre-fitted Q-function can then be leveraged on-line during main training to improve data collection. Our

Algorithm 1: Q-Learning through convex programming

Input: Environment simulation and reward function, number of pre-training iterations M, number of epochs N, Number of time steps per epoch n_T,Final time T_f, Batch size s_B, Initial conditions for optimal control $\mathbf{x}_{init}$

Initialise: Replay buffer $\mathcal{B}$ of size s_B

Pre-training:

1 Generate explorative samples using randomized control trajectories

2 Record generated states $\mathbf{x}_t$, controls $\mathbf{u}_t$, and Q-values in $\mathcal{B}$ as $\mathcal{B} = [X, \mathbf{y}]$ where: $X = [[\mathbf{x}_0, \mathbf{u}_0]_0, \dots, [\mathbf{x}_{T_f}, \mathbf{u}_{T_f}]_0, \dots, [\mathbf{x}_0, \mathbf{u}_0]_N, \dots, [\mathbf{x}_{T_f}, \mathbf{u}_{T_f}]_N]$ and $\mathbf{y} = [Q_{0,0}, \dots, Q_{t_f,0}, \dots, Q_{0,N}, \dots, Q_{t_f,N}]$

3 Find parameters θ that minimize $Q_{convex}(\cdot)$ in Eq. 5 using all entries j of X and y

Main training:

4 **for** *training iteration* $i = 1, \dots, M$ **do**

5 Initialise: $\mathbf{x} = \mathbf{x}_{init}$

6 **for** $j = 1, \dots, n_T$ **do**

7 Find $\mathbf{u}_{t_{j-1}}$ by minimising $Q_{convex}(\mathbf{u}; \mathbf{x}_{j-1}, \theta)$ using Eq. (3)

8 Sample the next state $\mathbf{x}_{t_j}$ by implementing control $\mathbf{u}_{t_{j-1}}$ in $\mathbf{x}_{t_{j-1}}$

9 Add $X_{new} = [[\mathbf{x}_{t_0}, \mathbf{u}_{t_0}], \dots, [\mathbf{x}_{n_T}, \mathbf{u}_{n_T}]]$ to X, and all associated Q-values in $\mathbf{y}_{new} = [Q_{t_0}, \dots, Q_{t_{n_T}}]$ to Y

10 Update $\mathcal{B}$ as $\mathcal{B} = [X, \mathbf{y}]$

11 Update optimal parameters θ by repeating step 3

approach essentially follows the typical RL pipeline, but rather than using DQNs for the Q-function approximation, we use convex surrogates that can be trained efficiently on fixed points using semidefinite programming, meaning that the controls identified are globally optimal given the learning approximation.

3. Results and Discussion

The selected case study in this paper simulates the photo-production of phycocyanin synthesised by *Arthrosporic platensis* which is a highly sought-after bioproduct. The dynamic system is assumed to take place in a semi-batch fixed volume fed-batch reactor and is set up in accordance with E.A. del Rio-Chanona et al. (2015). The two dependant states C_x and C_N represent *Arthrosporic platensis*'s biomass concentration in $g.L^{-1}$ and nitrate concentration within the batch in $mg.L^{-1}$. In order to control the process, light intensity I in $\mu mol.m^{-2}.s^{-1}$ and nitrate inflow rate F_N in $mg.L^{-1}.h^{-1}$ can be manipulated within their hard path constraints described by the continuous intervals $I \in [0, 300]$ and $F_N \in [0,7]$. To best reflect the process' economic viability, the objective function maximizes the biomass product while minimizing waste product nitrate concentration. It also considers initial conditions and the overall cost in form of controls expended:

$$R_{t_f} = -100 * \left(C_{X_{t_f}} - C_{X0}\right) + \left(C_{N_{t_f}} - C_{N0}\right) + \sum_{t=0}^{T_f} \left|\left|\boldsymbol{u_t}\right|\right|^2_{U_{max}} \quad (4)$$

$$with\ \boldsymbol{u}_t = [I_t, F_{N_t}]^T and\ \ U_{max} = \begin{bmatrix} {F_{N_{max}}}^{-2} & 0 \\ 0 & {I_{max}}^{-2} \end{bmatrix}$$

It is found that the most accurate predictions are given by the convex Q-function approximation as presented in Eq. (5).

$$\begin{aligned} Q_{convex}(\boldsymbol{u_t}, \boldsymbol{x_t}, \boldsymbol{\theta}) \\ &= {\boldsymbol{x}_t}^T P \boldsymbol{x}_t + {\boldsymbol{u}_t}^T \mathrm{W} \boldsymbol{u}_t + {\boldsymbol{x}_t}^T R \boldsymbol{u}_t - {\boldsymbol{x}_t}^T S \log(\boldsymbol{u}_t + 1) \\ &- {\boldsymbol{u}_t}^T T \log(\boldsymbol{x}_t + 1) + \boldsymbol{q}\boldsymbol{x}_t + \boldsymbol{s}\,\boldsymbol{u}_t - r \end{aligned} \quad (5)$$

where $\boldsymbol{x_t} \in \mathbb{R}^{3x1}$, $\mathbf{u_t} \in \mathbb{R}^{2x1}$, $\mathrm{P} \in \mathbb{R}^{3x3}$, $\mathrm{W} \in \mathbb{R}^{2x2}$, $\mathrm{R} \in \mathbb{R}^{3x2}$, $\mathrm{S} \in \mathbb{R}^{3x2}$, $\mathrm{T} \in \mathbb{R}^{2x3}$, $\mathrm{q} \in \mathbb{R}^{1x3}$, $\mathrm{r} \in \mathbb{R}$, $\mathrm{s} \in \mathbb{R}^{1x2}$, $\boldsymbol{\theta} = [P, Q, R, S, T, \boldsymbol{q}, \boldsymbol{s}, r]$.

The established Q-function approximation is next optimized to determine the optimal control profile resulting in maximum cumulative reward. The progress of convex optimization is tracked and can be seen converging after 80 iterations to a stable maximum of -173.6 with a standard deviation of 14.8 upon convergence as depicted in Figure 1a. The cumulative reward generated by the last training iteration is benchmarked against gPROMS' and Pyomo's optimization of the same case study, which, in both cases, yields -175.17. The slight difference of 0.91 % in cumulative cost (Figure 1a) might be attributable to differences in numerical solvers or rounding errors.

Lastly, the decrease in computational time by substituting the DQN utilized by Pan et al. (2020) with a convex Q-function is evaluated. Figure 1b tracks the computational time required to train the Q-function approximator versus the number of iterations. It must be noted that the DQN-based algorithm's computational time is cleared of the additional time that its explorative steps necessitate to ease plotting and comparing results. Consequently, Figure 1b depicts an underestimation of the DQN-based Q-learning's computational time. Still, it can be observed that our proposed algorithm consistently outperforms the DQN-based algorithm by a factor of around 3.6 with respect to time. When considering explorative behaviour, total computational time of the DQN-based Q-

learning takes 1680 seconds over the course of conducting 100 training iterations. In comparison, convex Q-learning only requires 415.53 seconds for the same amount of training, reducing computational time by 75.3%.

Our results suggest that convex function approximations can estimate Q-functions of highly nonlinear bioprocesses over continuous action spaces at similar solution quality to conventional dynamic optimization while requiring less computing power than DQN-based Q-learning.

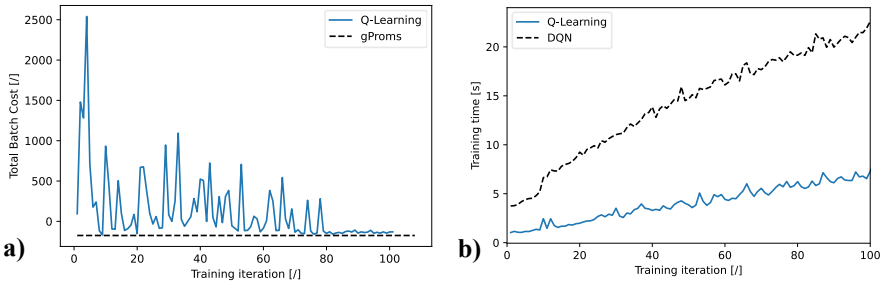


Figure 1. a) Training plot. Convergence to optimum of cumulative cost over training iterations. b) Training time. Computational time of convex Q-learning vs. DQN based Q-learning over training iteration

4. Conclusion and Future Work

In this work, an algorithm that utilizes convex function approximation for the Q-function in Q-learning is designed. Its generated results when applied to the photo-production of phycocyanin demonstrate high performance in precisely and efficiently approximating the Q-function as well as finding an optimal control policy even in a highly nonlinear environment with a continuous action space. Special focus is put on the algorithm's overall efficiency as it is benchmarked against a Q-learning algorithm using deep Q-networks optimized by evolutionary algorithms. This efficiency stems from utilizing convex functions as Q-function approximators which can be fitted efficiently on the whole replay buffer by using semidefinite programming.
In future work, the algorithm could be further improved by automating the finding of the convex Q-function approximation. Secondly, state constraints could be added to the model to better ensure process safety and demonstrate Q-learning's ability to learn uncertain systems as demonstrated by Pan et al. (2020). Thirdly, this algorithm could be implemented to conduct bi-level optimization to couple design and control problems or to couple scheduling and control problems as demonstrated by Sachio et al. (2021).

References

E. A. del Rio-Chanona, P. Dechatiwongse, D. Zhang, G. C. Maitland, K. Hellgardt, H. Arellano-Garcia, and V. S. Vassiliadis, 2015, Optimal Operation Strategy for Biohydrogen Production, Industrial & Engineering Chemistry Research, 52, 24, 6334-6343.

E.A. del Rio-Chanona, D. Zhang, V. S. Vassiliadis, 2016, Model-based real-time optimisation of a fed-batch cyanobacterial hydrogen production process using economic model predictive control strategy, Chemical Engineering Science, 142, 289-298, https://doi.org/10.1016/j.ces.2015.11.043

E.A. del Rio-Chanona, P. Dechatiwongse, D. Zhang, G. C. Maitland, K. Hellgardt, H. Arellano-Garcia, V. S. Vassiliadis, 2015, Optimal Operation Strategy for Biohydrogen Production, Industrial&Engineering Chemistry Research, 6334

H. Dong, Z. Ding, S. Zhang, 2020, Deep reinforcement learning : fundamentals, research and applications, Springer Singapore.

M. Mowbray, T. Savage, C. Wu, Z. Song, B. Anye Cho, E. A. Del Rio-Chanona, D. Zhang, 2021, Machine learning for biochemical engineering: A review, Biochemical Engineering Journal, 172, 108054.

E. Pan, P. Petsagkourakis, M. Mowbray, D. Zhang, E. A. del Rio-Chanona, 2020, Constrained Model-Free Reinforcement Learning for Process Optimization, Computers & Chemical Engineering, 154, 107462.

P. Petsagkourakis, I.O. Sandoval, E. Bradford, D. Zhang, E.A. del Rio-Chanona, 2020, Reinforcement learning for batch bioprocess optimization, Computers & Chemical Engineering, 133, 106649.

P. Petsagkourakis, I.O. Sandoval, E. Bradford, D. Zhang, E.A. del Rio-Chanona, 2020a, Constrained Reinforcement Learning for Dynamic Optimization under Uncertainty, IFAC-PapersOnLine 53 (2), 11264–11270.

S. Huh, I, Yang, 2020, Safe reinforcement learning for probabilistic reachability and safety specifications: A Lyapunov-based approach, arXiv preprint arXiv:2002.10126

J. Choi, F. Castañeda, C. J. Tomlin, K. Sreenath, 2020, Reinforcement Learning for Safety-Critical Control under Model Uncertainty, using Control Lyapunov Functions and Control Barrier Functions, arXiv preprint arXiv: 2004.07584

S. Diamond and S. Boyd, 2016, A python-embedded modeling language for convex optimization, Journal of Machine Learning Research, 17 (83)

S. Sachio, M. Mowbray, M. Papathanasiou, E. A. del Rio-Chanona, P. Petsagkourakis, 2021, Integrating process design and control using reinforcement learning, Submitted to Elsevier and available online at https://arxiv.org/pdf/2108.05242.pdf (accessed 9/26/2021).

V. Singh, H. Kodamana, 2020, Reinforcement learning based control of batch polymerisation processes, IFAC-PapersOnLine, 53, 1, 667-672.

L. Vandenberghe, S. Boyd, 1996, Semidefinite Programming, SIAM Review, 38(1), 49-95

J. Zhang, J. Kim, B. O'Donoghue, S. Boyd, 2020, Sample efficient reinforcement learning with REINFORCE. arXiv preprint arXiv:2010.11364.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50057-9

Differential Dynamic Programming Approach for Parameter Dependent System Control

Hyein JUNG[a], Jong Woo KIM[b], Jong Min LEE[a*]

[a]*School of Chemical and Biological Engineering, Seoul National University, 1, Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea.*
[b]*Technische Universität Berlin, Chair of Bioprocess Engineering, Straße des 17. Juni 135, 10623 Berlin, Germany*

[*]*jongmin@snu.ac.kr*

Abstract

This paper gives a differential dynamic programming (DDP) method for parameter-dependent system control. Parameter dependent system appears in the chemical and biological process engineering field, due to variable feed conditions, plant deterioration, etc. Model predictive control (MPC) has been applied to it in various forms, but its high online computation requirement makes practical application unrealistic. In contrast, DDP approach offers a simple state feedback control policy by approximating the value function based on the assumption of quadratic system dynamics and objectives. To handle parameter-dependent system without online re-calculation of the value function and control policy, parameter-dependent DDP (PDDP) method is proposed. PDDP method utilizes hyper-state, state and parameter augmented vector, and least square (LS) parameter estimator. Hyper-state enables PDDP method to retain the benefits of DDP method while incorporating parameter sensitivity information within its dynamics. The method was applied to a simple discrete-time linear system and outperformed its DDP counterpart.

Keywords: Process dynamic control, Adaptive control, Optimal control, Differential dynamic programming

1. Introduction

There has been great interest and challenge to control the systems with unknown parameters in chemical and biological process engineering. Common sources of parameter change include inaccurate estimates of model parameters and unknown aspect of the model itself - for example, variable feed condition or plant deterioration. Both data-driven models and first principle models require corrections using online data to resolve model plant mismatch. The problem of model plant mismatch becomes severe if a model structure is used more explicitly. This raised a need for an adaptive control method by [Anderson, 1985].

As a remedy, many researchers proposed model predictive control (MPC) approaches, as well-reviewed by [Heirung et al., 2018]. With MPC, parameter estimates can be easily applied after exploration, due to its recursive optimization structure. However, this suffers from a high online computation load when parameter uncertainty is taken into account.

Another approach is approximate dynamic programming (ADP). Its root is in dynamic programming (DP) which is an optimization methodology based on Bellman's principle

of optimality. It solves a multi-step decision-making problem by breaking it down into a one-step problem and encoding the information in a "reward-to-go" function, as known as value function. The resulting optimal control action is in state-feedback format, which is easy and fast to implement online. However, its offline computation burden suffers from "the curse of dimensionality," because of its backward sweep process.

ADP solves the bottleneck with value function approximation. Within a limited range of the state space, the original optimal control problem is approximated with a known function structure. However, in contrast to the MPC approach, the pre-computed optimal policy of ADP is useless when parameter changes. Accordingly, the offline computation should be implemented whenever there is a parameter change. Therefore, it is an important issue that how to implement the newly changed parameter value in the model when it comes to the ADP approach.

One solution is k-nearest neighbor (kNN) approximator [Lee and Lee, 2009]. This approach has proven its performance when applied to a batch bio-reactor [Byun et al., 2020]. The kNN approximation requires the Monte-Carlo search to approximate the value function. It can cover a wide range of parameter space, compensating its offline computation cost of the Monte-Carlo search.

One solution is using quadratic programming (QP), referred to as differential dynamic programming (DDP) method [Kobilarov et al., 2015]. DDP uses first and second-order derivative information of system dynamics and objective function to construct an approximate problem based on the Taylor expansion. This unconstrained QP problem has an analytical closed-form solution, which is state feedback. In this extension, the parameter-dependent differential dynamic programming (PDDP) method adopted hyper-state which is an augmented state of system state and parameter.

This concept has been proposed for robotic system control, where the unknown parameters are assumed to follow the Gaussian process [Kobilarov et al., 2015]. It provides the optimal control concerning the estimated parameter without online re-computation. This paper expands the application of PDDP to the step-change in parameter values with an online estimation of the least-squares (LS) method. Numerical simulation of the method is implemented in a parameter-dependent system of 2 by 2 linear system.

2. Background

2.1. Problem Formulation

The optimal control problem concerning parameter-dependency is formulated as below:

$$\begin{aligned} \min_{u_{[0:N-1]}} \quad & J = \sum_{k=0}^{N-1} (\|x_k - x_{ref,k}\|_Q^2 + \|u_k\|_R^2) + \|x_N - x_{ref,N}\|_{Q_f}^2 \\ \text{s.t.} \quad & x_{k+1} = f(x_k, u_k, p),\ y_k = h(x_k, p) + v_k,\ v_k \sim \Sigma_v \end{aligned} \tag{1}$$

Throughout this paper, x_k and u_k denote state and input respectively at time step k. A parameter, p, is also a variable, but it is assumed as an unknown constant for a finite time

horizon, because its dynamical behavior is much slower than that of the state. The operator $\|\cdot\|_X^2$ stands for the square of the ℓ_2-norm with a weight matrix X, i.e. $\|a\|_X^2 = a^T X a$. In PDDP method, a parameter variable is incorporated into a hyper-state, augmented with a state variable, z.

2.2. Differential Dynamic Programming

The DDP approximates nonlinear dynamics into a quadratic equation based on the Taylor expansion to utilize a quadratic programming structure. As the method only requires the local relationship between state, control input, and parameter, the second-order Hessian terms can be neglected in practice.

$$\delta x_{k+1} = f_{x,k}\delta x_k + f_{u,k}\delta u_k, \quad \delta y_k \quad = h_{x,k}\delta x_k \tag{2}$$

In the above equation, $f_{x,k}, f_{u,k}$, and $h_{x,k}$ refer to Jacobian matrices of the function $f(x_k, u_k, p_k)$ and $h(x_k, p_k)$ with regard to the subscript variables at their nominal values, $\bar{x}_k$ and $\bar{u}_k$, respectively, and $\delta x_k = x_k - \bar{x}_k$.

The control cost function, J, to be minimized is separated into two terms as a stage-wise cost, $l(x_k, u_k)$, and a terminal cost, $l_f(x_N)$. Based on this control cost function, a value function is defined as an expected sum of cost values beginning from the present time step given the state information:

$$V_k(x_k) = \min_{u_k, \ldots u_{N-1}} \sum_{i=k}^{N-1} l(x_i, u_i) + l_f(x_N). \tag{3}$$

Eq. (3) can be obtained recursively based on the Bellman optimality relation.

$$V_k(x_k) = \min_{u_k}[l(x_k, u_k) + V_{k+1}(f(x_k, u_k))] \tag{4}$$

To use deviation variables, let's set the deviation of Eq.(4) as $Q(\delta x, \delta u)$:

$$Q(\delta x, \delta u) = l(\bar{x} + \delta x, \bar{u} + \delta u) + V'(f(\bar{x} + \delta x, \bar{u} + \delta u)) - l(\bar{x}, \bar{u}) - V'(f(\bar{x}, \bar{u})), \tag{5}$$

where a subscript k is dropped to simplify a notation and $V' = V_{k+1}$. This notation applies to the equations appearing from now on. As the cost function can be exactly formulated as a quadratic form, the value function and Q function are also quadratic.

$$Q(\delta x, \delta u) = \frac{1}{2}\begin{pmatrix} 1 \\ \delta x \\ \delta u \end{pmatrix}^T \begin{bmatrix} \bar{Q} & Q_x^T & Q_u^T \\ Q_x & Q_{xx} & Q_{xu} \\ Q_u & Q_{ux} & Q_{uu} \end{bmatrix} \begin{pmatrix} 1 \\ \delta x \\ \delta u \end{pmatrix} \tag{6}$$

Now, the coefficients in Eq. (6) is recursively obtained through Eq. (5).

$$\begin{aligned} Q_x &= l_x + f_x^T V_x', & Q_u \quad &= l_u + f_u^T V_x', \quad Q_{ux} = l_{ux} + f_u^T V_{xx}' f_x \\ Q_{xx} &= l_{xx} + f_x^T V_{xx}' f_x, & Q_{uu} \quad &= l_u + f_u^T V_{xx}' f_u \end{aligned} \tag{7}$$

As a result, the optimal control input is analytically given as a minimizer of Q-function and can be expressed as below.

$$u^* = \bar{u} - Q_{uu}^{-1}(Q_u + Q_{ux}\delta x) \tag{8}$$

3. Methodology

3.1. Least Squares Parameter Estimation

Before implementing parameter-dependent control, a parameter estimation is required. In this paper, parameter estimation is formulated as a LS parameter estimation [Englezos and Kalogerakis, 2000].

$$\begin{aligned} \min_{\hat{p}} J &= \sum_{i=0}^{k} \|v_i\|^2_{Q_v} + \|\hat{p} - p_0\|^2_{Q_p} \\ \text{s.t. } \hat{x}_{i+1} &= f(\hat{x}_i, u, \hat{p}), \quad i = 1, ..., k \\ \hat{y}_i &= h(\hat{x}_i, \hat{p}) + v_i, \quad i = 0, ..., k \end{aligned} \tag{9}$$

where the sequence of y_i and u_i is an accumulated data from the initial time ($i = 0$) to the current time ($i = k$). The variables $\hat{x}_i, \hat{y}_i$ and $\hat{p}$ are the estimated variables, when initial value of state, x_0, and parameter, p_0, is given. A weighting parameter $Q_v = \Sigma_v^{-1}$, where Σ_v is given from the system of interest. The second term in the objective function, a parameter arrival cost, gives a smoothing effect for parameter estimation.

3.2. Parameter Dependent Differential Dynamic Programming

PDDP is formulated upon the hyper-state, z, instead of the state, x, extending the DDP method.

$$\delta z_{k+1} = F_{z,k}\delta z_k + F_{u,k}\delta u_k, \quad \delta y_k = H_{z,k}\delta z_k \tag{10}$$

Here, $F_{z,k}, F_{u,k}$, and $H_{z,k}$ refer to Jacobian matrices of the function $F(z_k, u_k)$ and $H(z_k)$ with regard to $\bar{z}_k$ and $\bar{u}_k$, respectively. Since there is no assumption for parameter dynamics or its uncertainty, it is assumed that the parameter value stays the same as the previous value. Hence, it can be said:

$$F_{z,k} = \begin{bmatrix} f_{x,k} & f_{p,k} \\ 0 & I_p \end{bmatrix}, F_{u,k} = \begin{bmatrix} f_{u,k} \\ 0 \end{bmatrix}, \tag{11}$$

where I_p stands for an identity matrix with the dimension of parameter vectors.

Then the cost function, J, from Eq. (1) is separated into two terms, stage-wise cost, $L(z_k)$ and the terminal cost, $L_f(z_N)$.

$$L(z_k) = \|x_k - x_{ref,k}\|^2_Q + \|u_k\|^2_R, \quad L_f(z_N) = \|x_N - x_{ref,N}\|^2_{Q_f} \tag{12}$$

The Q function is acquired equivalently as in the DDP method, and the optimal control is given as:

$$u^* = \bar{u} - Q_{uu}^{-1}(Q_u + Q_{uz}\delta\hat{z}). \tag{13}$$

where $\hat{z}$ is an estimated hyper-state from the estimator. This optimal state feedback control policy should be iteratively trained beforehand, saving Jacobian matrices and state-feedback gains $K_{u,k} = Q_{uu}^{-1}Q_u$ and $K_{z,k} = Q_{uu}^{-1}Q_{uz}$ at nominal states according to a model. For linear time-invariant systems, however, the optimal solution can be found directly from the given dynamics.

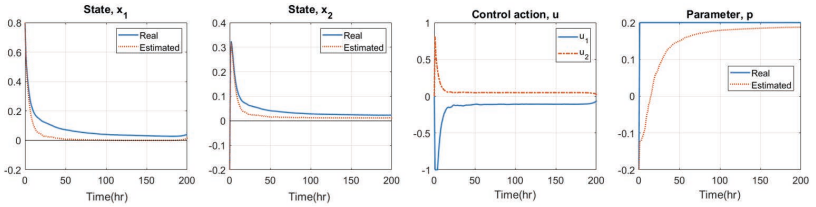


Figure 1: Linear system simulation result for PDDP with LS estimation

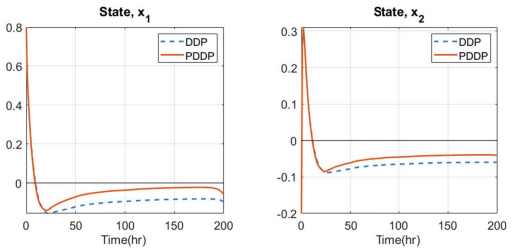


Figure 2: PDDP and DDP simulation result compared

4. Simulation Results

In this section, the efficacy of the proposed PDDP algorithm is shown with a following simple discrete time linear system:

$$\begin{aligned} x_{k+1} &= \begin{bmatrix} 0.9146 & 0.1665 \\ 0.2665 & 0.3353 \end{bmatrix} x_k + \begin{bmatrix} 0.0544 & -0.0757 \\ 0.0053 & 0.1477 \end{bmatrix} u_k + \begin{bmatrix} 0.0405 \\ 0.0058 \end{bmatrix} p_k \\ y_k &= x_k + v_k, \quad v_k \sim \begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix}, \end{aligned} \tag{14}$$

where v_k is uncorrelated measurement noise. This kind of parameter deviation may occur, for example, when process inlet condition (flow rate, composition, or temperature) changes.

With the change of parameter, PDDP with LS parameter estimation was able to successfully control the system as shown in Fig. 1. Also, it was compared with its DDP counterpart which uses the same hyper-state LS estimator in Fig. 2. As a result, PDDP was able to draw the states near the desired origin, while DDP was not.

5. Conclusions

In this paper, the solution of parameter-dependent system control was considered using an adaptive dynamic programming approach. With the use of hyper-state, PDDP is derived from the common DDP approach. For its application with unknown parameters, LS parameter estimation is combined. The test on a simple linear system showed that PDDP can

utilize the estimated parameter information through a state feedback format with nearly zero computation burden online. The works presented in this paper can be extended to nonlinear process by linearizing the process dynamics. Also, application combined with any other popular estimation methods such as Kalman filter is left as future work, which will guarantee the control performance even with unknown or unmeasurable disturbances.

ACKNOWLEDGEMENT

This paper was supported by Korea Institute for Advancement of Technology(KIAT) grant funded by the Korea Government (MOTIE) (P0008475, The Competency Development Program for Industry Specialist).

References

[Anderson, 1985] Anderson, B. D. (1985). Adaptive systems, lack of persistency of excitation and bursting phenomena. *Automatica*, 21(3):247–258.

[Byun et al., 2020] Byun, H.-E., Kim, B., and Lee, J. H. (2020). Robust adaptive control with active learning for fed-batch process based on approximate dynamic programming. *IFAC-PapersOnLine*, 53(2):5201–5206.

[Englezos and Kalogerakis, 2000] Englezos, P. and Kalogerakis, N. (2000). *Applied parameter estimation for chemical engineers*. CRC Press.

[Heirung et al., 2018] Heirung, T. A. N., Paulson, J. A., Lee, S., and Mesbah, A. (2018). Model predictive control with active learning under model uncertainty: Why, when, and how. *AIChE Journal*, 64(8):3071–3081.

[Kobilarov et al., 2015] Kobilarov, M., Ta, D.-N., and Dellaert, F. (2015). Differential dynamic programming for optimal estimation. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 863–869. IEEE.

[Lee and Lee, 2009] Lee, J. M. and Lee, J. H. (2009). An approximate dynamic programming based approach to dual adaptive control. *Journal of process control*, 19(5):859–864.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50058-0

Optimization of an air-cooler operation in an industrial distillation column

Masaharu Daiguji[a,b*] and Yoshiyuki Yamashita[b]

[a]*ENEOS Corporation, Otemachi 1-1-2, Chiyoda-ku, Tokyo 100-8162, Japan*
[b]*Department of Chemical Engineering, Tokyo University of Agriculture and Technology, Naka-cho 2-24-16, Koganei, Tokyo 184-8588, Japan*
daiguji.masaharu@eneos.com

Abstract

In the process industry, air coolers are some of the main cooling equipment. The air cooler of the atmospheric distillation column has fixed and variable fans, and the cooling duty is controlled. When the controller output exceeds the operating limits, the number of fixed fans in operation is changed to return it within the limits. However, the power consumption of the air cooler is not minimized. Daiguji and Yamashita (2022) proposed a method for optimizing the number of fixed fans in operation to minimize the power consumption, while reducing the frequency of fixed-fan starts and stops. Unfortunately, when this optimization method is applied to an air cooler with several variable fans, multiple fixed fans start and stop simultaneously, excessively disturbing the process. This paper proposes a modified optimization method in which the fixed fans start or stop one at a time. The modified optimization method was applied to industrial process, and the results showed that the power consumption was reduced, compared to the actual operation.

Keywords: Mixed-integer non-linear optimization, air cooler, optimal operation.

1. Introduction

Refineries and petrochemical plants have many facilities for cooling process streams. However, these cooling systems typically cannot be operated at the lowest cost. Recently, several studies have been conducted on minimizing the operating cost of water-cooling systems (Rubio-Castro et al., 2013; Muller and Craig, 2015; Viljoen et al., 2018; Viljoen et al., 2020). These studies also attempted to minimize the operating costs of air coolers and showed good results with hybrid nonlinear model-predictive control (HNMPC). Zhang et al. developed a dynamic model of the cold side of the cooling system of a power-generation boiler and proposed a model-predictive control with the model-based feed-forward compensation (Zhang et al., 2019). It has been shown that the back pressure of the unit can be controlled to the desired setpoint, while suppressing the disturbance of the air temperature, by properly manipulating the rotation speed.

The atmospheric-distillation columns are equipped with air coolers to cool the column-overhead gas and pump-around liquids. Tower overhead air coolers are often configured with a combination of multiple fixed and variable-pitch fans, because the cooling duty must be changed, according to the annual change in air temperature, while suppressing the equipment cost. Some air coolers have variable-speed fans instead of variable-pitch fans. The air coolers are not only disturbed by air temperature changes, but also by process-side disturbances. To suppress these disturbances, variable-pitch fans usually

control the process outlet temperature or column-top pressure. When the controller output exceeds the variable-pitch operating range, it is necessary to start or stop the fixed fan to return to the operating range. This causes problems.

One problem is that the power consumption of the fans is not minimized. Conventionally, a fixed fan is started or stopped, only when the controller output exceeds the operating range. At that time, the fact that it is designed to return to the operating range means that there is a choice in the number of fixed fans in operation required to obtain any given cooling duty. Therefore, it is possible to change the number of operating fixed fans to minimize the power consumption. However, frequently starting and stopping a fixed fan should be avoided because it causes the cooling duty to fluctuate and loads the process. Minimizing the power consumption of the fans, while considering the frequent starts and stops, is a challenge.

Daiguji and Yamashita (2022) attempted to stabilize the control and minimize the power consumption of an existing air cooler of a distillation column, without process changes. To optimize the number of fixed fans in operation, the paper proposed a method for minimizing power consumption while reducing the number of fixed-fan operation changes, and described the results of applying the proposed method to simulation data.

However, when the above optimization method is applied to an air cooler with several variable fans, which is often seen in industrial processes, another problem was found, where multiple fixed fans started or stopped simultaneously. In this paper, we propose a modified optimization method in which the fixed fans start or stop one at a time (Section 2). Then, the modified optimization method is applied to industrial data, and the results are compared with actual operation and other methods (Section 3). Finally, the conclusions are presented in Section 4.

2. Method for optimizing the number of fixed fans in operation

Daiguji and Yamashita (2022) described the results of a basic study on optimizing the number of fixed fans in operation to minimize the power consumption, while changing the number of fans less frequently under equal air-flow rates. Their study investigated a method for an air cooler with two fixed fans and two variable fans; however, this paper describes a method that can be applied to an air cooler with more variable fans.

2.1. Fans power consumption

According to the proportional law of basic fan characteristics, the relationship between the fan speed ω and power consumption W is as follows:

$$\frac{W_2}{W_1} = \left(\frac{\omega_2}{\omega_1}\right)^3. \tag{1}$$

Therefore, the power consumption W of the air cooler is given by

$$W = L_1 N \omega^3 + L_2 m, \tag{2}$$

where N is the number of variable-speed fans, m is the number of operating fixed fans, ω is the speed of a variable-speed fan, L_1 is the power-consumption coefficient of the variable-speed fan, and L_2 is the power consumption of the fixed fan. The above study is based on the assumption of a variable-speed fan; however, even in the case of a

variable-pitch fan, the power consumption increases at an accelerating rate when the air-flow rate is increased (Johnson, 1988). Therefore, the same explanation can apply.

2.2. Optimizing the number of fixed fans in operation

As described in Section 1, variable fans are typically used to control the process outlet temperature or the column-top pressure. If this control consists of a conventional single-loop PID controller, it fluctuates significantly when the fixed fan starts or stops. In addition, Sen (2012) states that induction motors draw three to eight times their rated value during startup. This means that starting a fixed fan increases the power consumption. Therefore, it is necessary to consider a method to reduce the frequency of changing the number of fixed fans. The following penalty function is defined, using the elapsed time t_a after the change in the number of fans as a variable:

$$\mathrm{f}(t_a) = \begin{cases} 1/t_a & \text{if changing the number of fixed fans.} \\ 0 & \text{if not changing the number of fixed fans.} \end{cases} \tag{3}$$

By multiplying this penalty function by the weight λ and adding it to Eq. (2), the following evaluation equation J is obtained:

$$J = L_1 N\omega^3 + L_2 m + \lambda \cdot \mathrm{f}(t_a). \tag{4}$$

Minimizing the value of this evaluation equation J minimizes the power consumption. Moreover, if the number of fixed fans is repeatedly changed in a short period of time, the penalty function becomes large, which reduces the frequency of the changes. Even if the number of fixed fans changes, the cooling duty must remain constant. Assuming that the independent variables, other than the number of fans (e.g., air-inlet temperature), do not change, the cooling duty can be considered to be constant when the air-flow rate Q is constant. Therefore, the following constraint conditions are obtained:

$$Q = K_1 N\omega + K_2 m = \text{const.}, \tag{5}$$

where K_1 is the air-flow coefficient of the variable fan and K_2 is the air-flow rate of the fixed fan. Another constraint is that the speed of the variable fan ω must be within the allowable upper and lower limits, which is expressed by the following equation:

$$\omega_{\min} \leq \frac{1}{K_1 N}(Q - K_2 m) \leq \omega_{\max}. \tag{6}$$

Furthermore, in the case of air coolers with several variable fans, which are often used in industrial processes, the number of fixed fans with minimum power consumption easily changes for disturbances of the same magnitude. Therefore, it is necessary to add the following constraint condition so that the fixed fans start and stop one by one:

$$\max(0, m_{old} - 1) \leq m \leq \min(M, m_{old} + 1), \tag{7}$$

where m_{old} is the number of fixed fans in operation before optimization, and M is the number of fixed-fan facilities. Under the constraints in Eqs. (5), (6), and (7), a modified optimization method is proposed to find the number of fixed fans in operation that minimizes the result of Eq. (4).

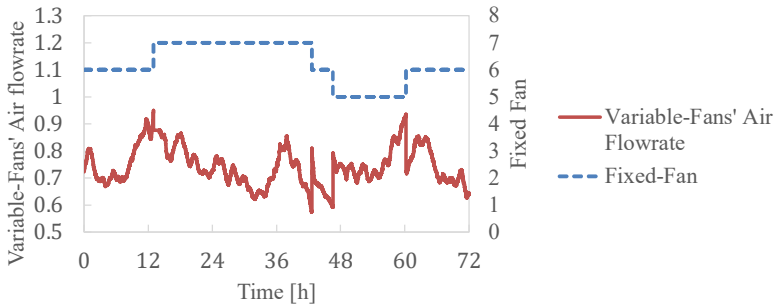


Fig.1 Time-series data of the air-flow rate of the variable fans and the number of fixed fans in operation, collected from an industrial plant.

3. Application example using operation data from the process industry

In this section, we describe the results of applying the proposed method to industrial process data, and confirm its effectiveness by comparing it with actual operations.

3.1. Identification of air-flow rate and power-consumption equations

The example air cooler has eight fixed fans and eight variable-pitch fans. The fans' equipment specifications state that the column-top pressure is controlled in the range of 57.4–95.2% of the variable-fan air-flow capacity. The following equation for the air-flow rate Q of the air cooler and the air-flow ratio q of the variable fan was obtained:

$$Q \propto 8q + m, 0.574 \leq q \leq 0.952, m \in \mathbb{Z}, 0 \leq m \leq 8. \tag{8}$$

Next, the following equation for the relationship between the power consumption W and the air-flow capacity ratio q of the variable fan was obtained:

$$W \propto 8(8.15q^3 + 18.3q) + 26.45m. \tag{9}$$

3.2. Industrial-data collection

Five-second cycle operation data were collected for three days from the output of the column-top pressure controller and the ON/OFF status of each fan. Based on these historical data and the number of fixed fans in operation, Eq. (8) was used to obtain the air-flow rate data of the variable fans (Fig. 1).

3.3. Application of conventional method

Figure 2(a) shows the results of applying the conventional method to industrial data. In the conventional method, when the air-flow rate of the variable fans reaches the lower or upper limit, the number of fixed fans is changed, such that it returns within the range. The results showed that the number of fixed fans changed four times in three days, which is the same as the actual operation. Regarding the average power consumption, the difference from the actual operation was less than 0.01%.

3.4. Application of proposed method

The optimization method proposed in Section 2 was coded in the MATLAB® environment using the genetic-algorithm function of the Global Optimization Toolbox

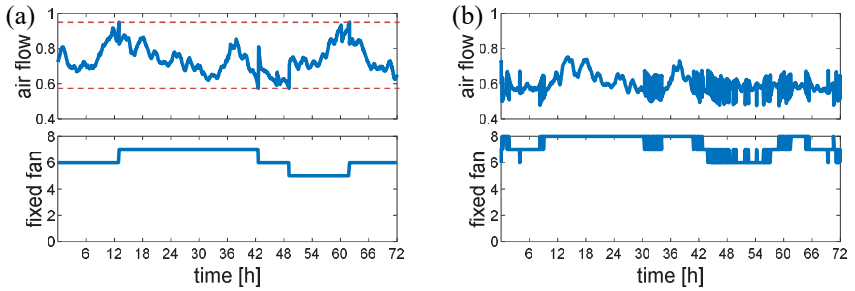


Fig.2 (a) Control response for the minimizing the frequency of changing the number of fixed fans (conventional method). (b) Control response of the optimization without the penalty (power-consumption minimization).

(Deep et al., 2009). Figure 2(b) shows the results of applying the proposed method, without a penalty, to industrial data. Optimization without penalty results in the minimum power-consumption solution. The results showed that the air-flow rate of the variable fans operated to stay in the range of 47–75%, and the power consumption was reduced by 1.3%, on average, compared to the conventional method. However, the number of fixed fans changed 166 times in three days.

Figure 3(a) shows the results of applying the proposed method with a penalty to the industrial data. By considering the elapsed time after starting or stopping a fixed fan, it was found that the air-flow rate of the variable fan was extended over a range of 38–81% and the reduction in power consumption was 1.1%. In contrast, the change in the number of fixed fans was reduced to 4, and the fixed fans no longer started and stopped continuously within a short period of time. However, it was found that two fixed fans started and stopped simultaneously, causing excessive disturbances to the process.

Figure 3(b) shows the results of applying the modified proposed method—which limits the number of simultaneous fixed-fan starts and stops to one—to industrial data. By considering the number of simultaneous starts and stops, the air-flow rate of the variable fan was further extended to the range of 38–88%, and the reduction in power consumption was 0.9%. On the other hand, the change in the number of fixed fans was maintained at four times, with one fan per change.

4. Conclusions

In this paper, a modified method for minimizing the power consumption of controlled

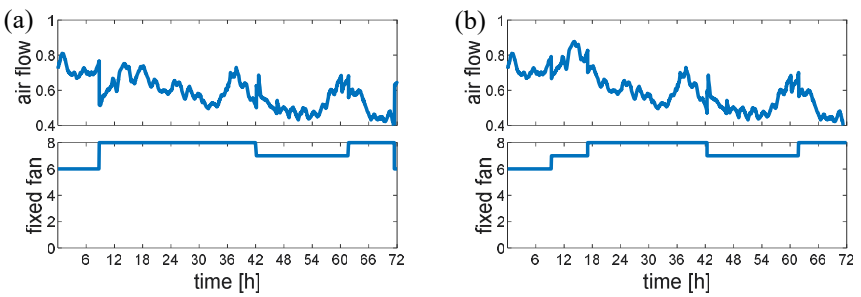


Fig.3 Control response of the optimization with a penalty, (a) Original method. (b) Limiting the number of simultaneous fixed-fan starts and stops to one (proposed method).

Table 1. Comparison of controller performances

	Max. number of simultaneous fixed-fan starts or stops	Power-consumption reduction ratio [%]	Number of fixed-fan starts or stops
Actual data	1	-	4
Conventional method	1	0.0	4
Proposed method (modified)	1	0.9	4
Proposed method (original)	2	1.1	4
Power-consumption minimization	1	1.3	166

air coolers was proposed. When the using the method to minimize power consumption, while reducing the frequency of changing the number of fixed fans, it is possible to start and stop multiple fixed fans simultaneously. Therefore, a modified method was proposed that limited the number of simultaneous fixed-fans changes to one.

Next, the proposed method was applied to industrial data, and the actual operation was compared with the method that minimized the frequency of changing the number of fans (conventional method), the power-consumption minimization method, and the proposed method (Table 1). The results showed that the proposed method reduced the power consumption by approximately 1%, compared to the actual operation, although it was slightly inferior to the power-consumption minimization method. However, the number of fixed-fan starts or stops was approximately 40 times greater in the power-consumption minimization method than in the actual operation. It was the same in the proposed method as in the actual operation, with a minimum of four times. Furthermore, the fixed fans started or stopped one at a time using the modified proposed method.

References

C. J. Muller and I. K. Craig, 2015, Modelling of a dual circuit induced draft cooling water system for control and optimization purposes, Journal of Process Control, 25, 105-114

C. M. Johnson, 1988, Comparison of variable pitch fans and variable speed fans in a variable air volume system, *Building Serv. Eng. Res. Technol*, 9, 89-98

E. Rubio-Castro and J. M. Ponce-Ortega, 2013, Synthesis of cooling water systems with multiple cooling towers, *Applied Thermal Engineering*, 50, 957-974

J. H. Viljoen, C. J. Muller and I. K. Craig, 2018, Dynamic modelling of induced draft cooling towers with parallel heat exchangers, pumps and cooling water network, Journal of Process Control, 68, 34-51

J. H. Vilijoen, C. J. Muller and I. K. Craig, 2020, Hybrid nonlinear model predictive control of a cooling water network, *Control Engineering Practice*, 97 104319

K. Deep, K. P. Singh, M. L. Kansal and C. Mohan, 2009, A real coded genetic algorithm for solving integer and mixed integer optimization problems, *Applied Mathematics and Computation*, 212, 505-518

M. Daiguji and Y. Yamashita, 2022, Optimization of multiple-fan air-cooler control in a distillation column, *Kagaku Kogaku Ronbunshu*, accepted

P. C. Sen, 2012, 2012, Chapter 9 Transients and Dynamics, *Principles of Electric Machines and Power Electronics*, John Wiley and Sons, Inc., USA

Y. Zhang, F. Zhang and J. Shen, 2019, On the dynamic modeling and control of the cold-end system in a direct air-cooling generation unit, *Applied Thermal Engineering*, 151, 373-384

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50059-2

Dynamic Operability Analysis for the Calculation of Transient Output Constraints of Linear Time-Invariant Systems

San Dinh[a], Fernando V. Lima[a*]

[a]*West Virginia University, Morgantown, West Virginia, 26506, U.S.A*
**Fernando.Lima@mail.wvu.edu*

Abstract

In this work, a dynamic operability mapping is developed to find an operable funnel for a linear time-invariant dynamic system. The existing operability mapping method to find this funnel is computationally expensive, which makes it unsuitable for online control applications. A novel two-step calculation procedure is proposed, which includes an offline computation of the nominal funnel by constructing a convex hull of the manipulated variable projections, followed by an online update that adjusts the funnel to an operable region based on the current state information. As a result, a dynamic funnel that contains all achievable outputs regardless of the process disturbances and measurement noises is obtained in the form of transient output constraints for model predictive control implementation.

Keywords: Dynamic Operability, Linear Control, Output Constraints

1. Introduction

Process operability is defined as the design and control ability to achieve desired performance from the given available inputs regardless of the realization of the disturbances (Gazzaneo et al., 2020). If the operability analysis is able to be carried out along with the operation of a process, not only the achievable portions of the desired outputs are known, but also the feasible output constraints can be provided for model predictive control to guarantee feasibility (Lima and Georgakis, 2009). However, the currently available operability analysis involves an exhaustive generation of the input combinations, and thus this approach may quickly become intractable.

In this paper, the achievable output sets at all values of the disturbances are formulated as a set of time-dependent polyhedra, which is referred to as the dynamic operable funnel. To avoid confusion between control theory and process operability concepts, external output constraints are defined here as the constraints on the output variables that are given by the physical nature of a process, such as thermodynamic and equipment's safety limits. In the application to online model predictive control, the dynamic funnel provides the transient output constraints to keep the process from moving toward an inoperable region, and the online calculation must be done efficiently to assure a sufficient time for the controller to solve for an optimal path.

In particular, the dynamic operable funnel of a linear time-invariant dynamic process is proven to be defined as a polyhedron. Also, the ability of the current process to move toward its stable operating region is quickly verified following Phase I of the simplex algorithm for linear programming, and the operable region in the presence of external output constraints can be obtained via the convex hull of suitable geometric duals with respect to a feasible solution (Muller and Preparata, 1978). Therefore, the remaining

challenge is constructing the dynamic operable funnel in a tractable manner. In the proposed framework, the funnel calculation is divided into two steps: the first step is computing the funnel offline before the full state information arrives; and the second step is updating the funnel online according to the full state information that becomes available. The preliminaries and concepts necessary to define the proposed approach are detailed next.

2. Dynamic Operability Problem Background

2.1. Preliminaries

Consider the following discrete-time linear time-invariant dynamic system:

$$x(k+1) = Ax(k) + Bu(k) + Gw(k);\ x(0) = x_0 \tag{1}$$

$$y(k) = Cx(k) + Du(k) + v(k) \tag{2}$$

in which $x(k) \in \mathbb{R}^{n_x}, u(k) \in \mathbb{R}^{n_u}\ and\ y(k) \in \mathbb{R}^{n_y}$ are the vectors of state variables, input/manipulated variables and output/controlled variables, respectively; $w(k) \in \mathbb{R}^{n_w}$ and $v(k) \in \mathbb{R}^{n_v}$ are the zero-mean multivariate Gaussian distributed vectors with the respective positive definite covariance matrices, $\Sigma_w \in \mathbb{R}^{n_w \times n_w}$ and $\Sigma_v \in \mathbb{R}^{n_v \times n_v}$. The initial time $k = 0$ is defined to be the current time instead of the time in which the process begins, and the initial state variables, x_0, are assumed to be given by a state observer.

Since $w(k)$ and $v(k)$ are assumed to be zero-mean with Gaussian distributions, the states and the outputs are also multivariate Gaussian random variables with the respective means $\bar{x}(k)$ and $\bar{y}(k)$. The sequences of covariance matrices for the states, $\Sigma_x(k)$, and the outputs, $\Sigma_y(k)$, are:

$$\Sigma_x(k+1) = A\Sigma_x(k)A^T + G\Sigma_w G^T; \Sigma_x(0) = 0_{n_x \times n_x} \tag{3}$$

$$\Sigma_y(k) = C\Sigma_x(k)C^T + \Sigma_{\mathrm{v}} \tag{4}$$

When a random vector $p \in \mathbb{R}^{n_p}$ is a Gaussian random vector with a mean $\bar{p}$ and a covariance matrix Σ_p, its 95% highest density region, $HDR(p)$, is the following ellipsoid with the scale l_p^2 equals to the inverse cumulative distribution function of the chi-squared distribution with n_p degrees of freedom:

$$HDR(p) = \left\{p | (p-\bar{p})^T \Sigma_p^{-1} (p-\bar{p}) \le l_p^2; l_p^2 = Inv_{\chi^2}(95\%; n_p)\right\} \tag{5}$$

2.2. Dynamic operability sets

The Available Input Set at the discretized time k (AIS_k) is defined as the set of all feasible sequences of manipulated variables from the initial time 0 to time k.

$$AIS_k = \left\{u_k = [u(0)^T, u(1)^T, \dots, u(k-1)^T]^T | u_{k,min} \le u_k \le u_{k,max}\right\} \tag{6}$$

The Expected Disturbance Set (EDS_k^d) is the set of all realizations of the disturbances, d, at the time k. The two sources of disturbances assumed here are the $w(k)$ and $v(k)$, which can take any real values due to their Gaussian distributions. Their values are constrained to their respective 95% highest density regions as follows:

$$EDS_k^d = \left\{ d(k) = [w(k)\ v(k)]^T \middle| \begin{matrix} d(k)^T diag(\Sigma_w^{-1}, \Sigma_v^{-1}) d(k) \le l_d^2 \\ l_d^2 = Inv_{\chi^2}(95\%; n_w + n_v) \end{matrix} \right\} \tag{7}$$

The Achievable Output Set at a fixed disturbance d ($AOS(d)$) is the set of all possible outputs at the discretized time k given the linear system (1), (2) and the range of manipulated variables. A necessary condition for a process to be operable is that the set of achievable outputs regardless of the realizations of the process disturbances, AOS_k, has to be nonempty. The AOS_k is defined as the intersection of all achievable output sets at fixed realizations of the disturbance:

$$AOS_k = \bigcap_{d \in EDS_k^d} AOS_k^d(d) = \left\{ y(k) \middle| \begin{matrix} (1), (2)\ are\ satisfied; \\ u_k \in AIS_k; d(k) \in EDS_k^d; \end{matrix} \right\} \tag{8}$$

3. Calculation of Transient Output Constraints

3.1. Offline computation of transient state funnel at nominal-valued disturbances

The following assumptions are considered for the offline calculation of the dynamic funnel that can be later addressed in the online update: $x_0 = 0_{n_x}; C = I_{n_x \times n_x}; D = 0_{n_x \times n_u};\ w(i) = 0_{n_w}; v(i) = 0_{n_x}\ \forall\ i \le k$. The considered outputs are the predicted state variables before $w(k)$ and $v(k)$ are accounted for, and the AOS_k has the form:

$$AOS_k = \{x(k) | x(k) = \bar{B}_k u_k; u_k \in AIS_k\} \tag{9}$$

where $\bar{B}_k = [A^{k-1}B\ A^{k-2}B\ \dots AB\ B]$. From the definition (6), the AIS_k is a bounded convex polyhedron. From the formulation of (9), $\bar{B}_k: \mathbb{R}^{k \times n_{nu}} \to \mathbb{R}^{n_x}$ is a linear transformation of the AIS_k into the AOS_k, so that the AOS_k is exactly the smallest convex hull that contains all the projections of the available input sequences on the state vector space. Additionally, for an achievable state $x(k)$ to be a vertex of the AOS_k, its preimage, u_k, must be a vertex of the AIS_k. Then the AOS_k can be computed by taking the convex hull of the $2^{k \times n_u}$ vertices of the AIS_k, which is the vector of the input sequence in which each element is either taken from the value of the lower bound $u_{k,min}$ or the upper bound $u_{k,max}$:

$$AOS_k = convexhull\left(\bar{B} u_k | u_k^T e_i \in \left\{u_{k,min}^T e_i, u_{k,max}^T e_i\right\}, \forall i \le k \times n_u\right) \tag{10}$$

where $e_i = [0,0,\dots,0,1,0,\dots,0]^T \in \mathbb{R}^{k \times n_u}$ is a standard basis for which only the i^{th} location has the value of 1. An efficient approach to find the convex hull in high-dimensional spaces is the Quickhull Algorithm (Barber et al., 1996). According to the Minkowski-Weyl's Theorem, every polyhedron is identically described by its vertex representation and its hyperplane representation, and thus the formulation of AOS_k in (10) can be equivalently converted to a set of linear constraints using the Double Description Method (Fukuda and Prodon, 1996):

$$AOS_k = \{x(k) | \bar{H}_k x(k) \le \bar{l}_k\} \tag{11}$$

In the simplest case of the online calculation, if the process disturbances and the measurement noises are not considered, the online update of the dynamic funnel can be established by the substitution of (11) into the state-space model (1), and the dynamic funnel at the current state x_0 is simply:

$$AOS_k = \{x(k) | \bar{H}_k x(k) \leq \bar{\bar{l}}_k; \bar{\bar{l}}_k = \bar{l}_k + \bar{H}_k A^k x_0\} \tag{12}$$

3.2. Transient state funnel with process disturbances

In this subsection, the interested outputs are the state variables, and the following assumptions are considered: $C = I_{n_x \times n_x}; D = 0_{n_x \times n_u};\ v(i) = 0_{n_x}, \forall\, i \leq k$. The process disturbance sequence can be redefined as the deviation, $w_x(k)$, from the mean value of the state vector, and the EDS is chosen as the HDR with respect to $x(k)$:

$$EDS_k^x = \{w_x(k) | w_x(k)^T \Sigma_x^{-1}(k) w_x(k) \leq l_x^2; l_x^2 = Inv_{\chi^2}(95\%; n_x)\} \tag{13}$$

The formulation of AOS_k in this subsection is

$$AOS_k = \bigcap_{w_x(k) \in EDS_k^x} AOS_k^w(w_x(k)) = \left\{x(k) \middle| \begin{matrix} x(k) = A^k x_0 + \bar{B}_k u_k + w_x(k) \\ u_k \in AIS_k; w_x(k) \in EDS_k^x; \end{matrix} \right\} \tag{14}$$

Let $\Sigma_x(k) = V_x(k) S_x(k) V_x^{-1}(k)$ be the eigenvalue decomposition of the covariance matrix $\Sigma_x(k)$. Since a basic property of any covariance matrix is positive definiteness, $V_x^{-1}(k) = V_x^T(k)$ is an orthogonal matrix, and $S_x(k)$ is a diagonal matrix with positive elements. Denoting $S_x^{-0.5}(k)$ to be an inverse of the square root of $S_x(k)$, a bijective mapping $L = S_x^{-0.5}(k) V_x^T$ that transforms the state vector $\hat{x}(k) = Lx(k)$ is introduced. The covariance matrix of the transformed vector $\hat{x}(k)$ is:

$$\Sigma_{\hat{x}}(k) = L \Sigma_x(k) L^T = S_x^{-0.5}(k) V_x^T V_x(k) S_x(k) V_x^T(k) V_x(k) S_x^{-0.5}(k) = I_{n_x \times n_x} \tag{15}$$

Because the covariance $\Sigma_{\hat{x}}$ is an identity matrix, the proposed linear mapping L corresponds to a change of coordinates to transform the state vector into a standard Gaussian random vector, and the ellipsoid EDS_k^x is transformed into an n-sphere $EDS_k^{\hat{x}}$ with radius l_x. This provides an advantage when finding the intersection AOS_k of all achievable output sets for the disturbance realizations based on the following theorem:

Theorem 1: Let $[H]_i$ denote the i^{th} row of a matrix $H: \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$. Given a bounded polyhedron in the form of $P_x = \{x \in \mathbb{R}^{n_x} | Hx \leq l\}$ and its image under a bounded translation according to an n-sphere $P_x(d) = \{\hat{x} | \hat{x} = x + d; Hx \leq l; d^T d \leq l_d^2\}$, the intersection of all $P_x(d)$ is given by:

$$P = \bigcap_{d^T d \leq l_d^2} P_x(d) = \left\{x \middle| Hx \leq \hat{l}; \left[\hat{l}\right]_i = [l]_i - l_d \sqrt{[H]_i^T [H]_i}\ \forall\, i \leq n_2\right\} \tag{16}$$

Proof: For each hyperplane $[H]_i x \leq [l]_i$, the hyperplane $[H]_i x \leq [l]_i - l_d \sqrt{[H]_i^T [H]_i}$ is the parallel hyperplane shifted toward the feasible half-space by a distance of l_d. Thus, a translation of all feasible points in $[H]_i x \leq [l]_i$ by a distance d can only violate $[H]_i x \leq \left[\hat{l}\right]_i$ if $d > l_d$. Therefore, $Hx \leq \hat{l}$ is the intersection of all hyperplanes $[H]_i x \leq [l]_i$ when the translation distance is less than or equal to l_d.

Note that the disturbance effects on the state vector are the same as translating the achievable output set in (12) by a translation vector in EDS_k^x, and the linear mapping L puts the AOS_k^w in the form that is applicable for Theorem 1. Finally, since L is a bijective mapping, the final form of the transient state funnel with process disturbances in the original state vector $x(k)$ is given by:

$$AOS_k = \left\{ x(k) \middle| \bar{H}_k x(k) \le \hat{l}_k; \left[\hat{l}_k\right]_i = \left[\bar{l}_k + \bar{H}_k A^k x_0\right]_i - l_x \sqrt{\left[\bar{H}_k L^{-1}\right]_i^T \left[\bar{H}_k L^{-1}\right]_i} \right\} \quad (17)$$

3.3. Transient output funnel with process disturbances and measurement noises

The output vector can be interpreted as a projection of the state variables and the manipulated variables at the same time step. Similarly to the previous subsection, using the Double Description Method, all the vertices of AOS_k in the state vector space can be found. Following the same procedure from (10) to (12), one can arrive at the achievable output set with process disturbances before considering the measurement noises:

$$AOS_k(v(k) = 0) = \left\{ y(k) | H_k y \le \bar{b}_k \right\} \quad (18)$$

Since the effects of the measurement noises on the outputs are the same as the disturbances on the state variables, a similar procedure from (13) to (17) can be followed with the linear mapping $L_y = S_y^{-0.5}(k) V_y^T(k)$ defined according to the eigenvalue decomposition of $\Sigma_y(k) = V_y(k) S_y(k) V_y^{-1}(k)$. The final form of the achievable output set is:

$$AOS_k = \left\{ y(k) \middle| H_k y \le b_k; [b_k]_i = \left[\bar{b}_k\right]_i - l_y \sqrt{\left[H_k L_y^{-1}\right]_i^T \left[H_k L_y^{-1}\right]_i} \right\} \quad (19)$$

4. Numerical Example

Consider the system given in (1), (2) with the following matrices:

$$\begin{gathered} A = \begin{bmatrix} 0.59 & -0.43 \\ -0.06 & 0.39 \end{bmatrix}; B = \begin{bmatrix} 0.42 & 1.82 \\ 2.48 & -0.71 \end{bmatrix}; G = \begin{bmatrix} 0.52 & -0.47 \\ 1.22 & 0.47 \end{bmatrix}; \\ C = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}; D = 0_{2\times 2}; \Sigma_w = \begin{bmatrix} 0.04 & 0 \\ 0 & 0.02 \end{bmatrix}; \Sigma_v = 10^{-5} \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}; x_0 = \begin{bmatrix} 20 \\ -30 \end{bmatrix} \end{gathered} \quad (20)$$

The prediction horizon is chosen to be 6 for illustrative purposes, and the objective is constructing the six achievable output sets AOS_k for $k = 1, \ldots, 6$. The input ranges of the considered AIS_k are $-1 \le u_1(k) \le 1$ and $-2 \le u_2(k) \le 2$. In the offline computation, the vertices of the AIS_k, which are all combinations of $u(k) \in \{[-1\ -2]^T, [-1\ 2]^T, [1\ -2]^T, [1\ 2]^T\}$ for all $0 \le k \le 5$, are applied to the linear state-space model to calculate the associated basis state vectors. The set of convex hulls of these basis state vectors at each time k is the nominal AOS_k, and the funnel of nominal state vectors obtained for this case is shown in Figure 1(a).

In the online update of the dynamic funnel, at each value of k, the AOS_k is adjusted according to (12), and the new dynamic funnel at $w(k) = 0$ and $v(k) = 0$ is shown in Figure 1(b). To find the intersection of all AOS_k at different values of $w(k)$ in the 95% highest density region, (17) is applied, and the new AOS_k that takes into account process disturbances, $w(k)$, is shown as the dashed-edge empty polytopes in Figures 1(c) and (d). In the next step, AOS_k of state vectors are projected into the space of the output vectors, and the convex hulls of the images at every time k is the AOS_k of output vectors, which is represented as the dashed-edge empty polytopes in Figures 1(e) and (f). Finally, to address the measurement noise, the hyperplanes of every AOS_k are shifted inward according to (19). The result is a funnel of output vectors that can always be achieved regardless of the realization of the process disturbances and the measurement noises by varying the constrained manipulated variables. This funnel is plotted with dotted-edge grey-filled polytopes in Figures 1(e) and (f).

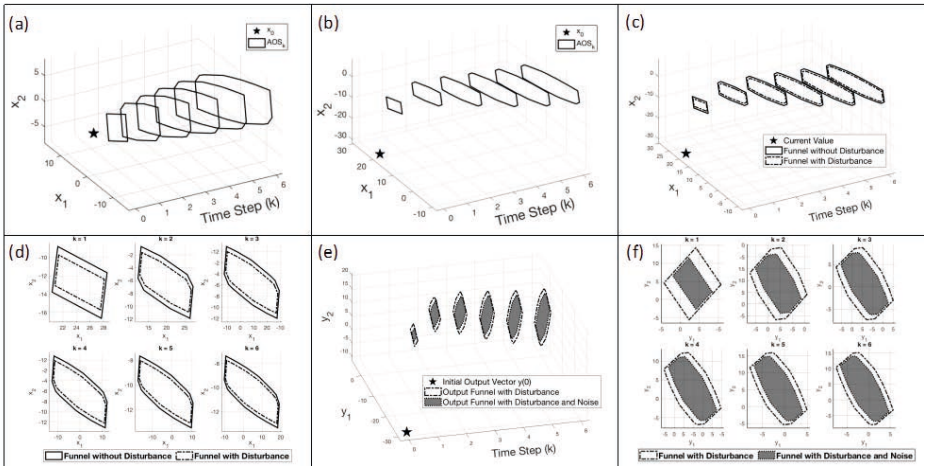


Figure 1: Dynamic operable funnels. (a): Funnel of state vector considering nominal initial state; (b): Funnel of state vector considering actual initial state without disturbances; (c), (d): Adjustment of funnel of state vectors with process disturbances; (e), (f): Funnels of output vectors with and without measurement noises.

5. Conclusions

Dynamic operability corresponds to an output controllability measure that can be used to assist with the formulation of online constrained control problems (Gazzaneo et al., 2020). However, in dynamic operability mapping, exhaustive input discretization methods in the reported literature quickly become intractable with the increase in predictive horizon length. In this work, a novel dynamic operability mapping was proposed in a two-step framework that allows the majority of the computational effort being performed offline. The achievable output sets at different predictive times were formulated as set of inequality constraints that are updated online according to the current full state information and uncertainty propagation. Even though the current framework is limited to a linear time-invariant dynamic process, the proposed theory is a valid basis for future work on linear time-varying and nonlinear dynamic processes.

References

Barber, C.B., Dobkin, D.P., Huhdanpaa, H., 1996. The Quickhull Algorithm for Convex Hulls. ACM Transactions on Mathematical Software (TOMS) 22, 469–483.

Fukuda, K., Prodon, A., 1996. Double Description Method Revisited, in: Deza, M., Euler, R., Manoussakis, I. (Eds.), Combinatorics and Computer Science, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 91–111.

Gazzaneo, V., Carrasco, J.C., Vinson, D.R., Lima, F.V., 2020. Process Operability Algorithms: Past, Present, and Future Developments. Ind. Eng. Chem. Res. 59(6), 2457–2470.

Lima, F.V., Georgakis, C., 2009. Dynamic Operability for the Calculation of Transient Output Constraints for Non-Square Linear Model Predictive Controllers. IFAC Proceedings Volumes 42(11), 231–236.

Muller, D.E., Preparata, F.P., 1978. Finding the Intersection of Two Convex Polyhedra. Theoretical Computer Science 7, 217–236.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50060-9

Effective Re-identification of a Multivariate Process under Model Predictive Control Using Information from Plant-Model Mismatch Detection

Masanori Oshima[a], Sanghong Kim[b*], Yuri A. W. Shardt[c], Ken-Ichiro Sotowa[a]

[a]*Department of Chemical Engineering, Kyoto University, Nishikyo-ku, Kyoto 615-8510, Japan*
[b]*Department of Applied Physics and Chemical Engineering, Tokyo University of Agriculture and Technology, Naka-cho, Koganei-city, Tokyo 184-8588, Japan*
[c]*Department of Automation Engineering, Technical University of Ilmenau, P.O. Box 10 05 65, Ilmenau D-98684, Germany*
sanghong@go.tuat.ac.jp

Abstract

A process under model predictive control is required to be re-identified when plant-model mismatch (PMM) occurs. During data acquisition for re-identification, the process is excited to enable accurate re-identification. However, the excitation of the process worsens control performance. This research proposes a new method for re-identification that can deal with the problem. In the proposed method, only the inputs of the transfer functions that have significant PMM are excited, and, at the same time, the other inputs are manipulated to suppress the variations of the controlled variables. The usefulness of the proposed method was confirmed through a simulation case study of a 3-input, 3-output process. As a result, it was shown that the proposed method can reduce the mean absolute control error during data acquisition to 87% of that of an existing method without compromising model accuracy after re-identification.

Keywords: Dual control; Model predictive control; Multivariate process; Plant-model mismatch; Re-identification

1. Introduction

Due to the spread of the Internet and the concept of mass customization, external demands on industrial processes are changing more rapidly than they were previously. Also, the characteristics of industrial processes change over time due to various factors such as degradation of catalysts or fouling of pipes. Therefore, to improve process productivity, a control system that adapts to the internal and external changes of the process and achieves optimal operation is required. One possible solution is the use of model predictive control (MPC). MPC can adaptively achieve optimal operation for various processes including nonlinear processes, time-varying processes, and processes with constraints. Since MPC is a model-based control method, the control performance depends on the prediction accuracy of the model. Therefore, it is important to maintain the high prediction accuracy of the model used for MPC.

In order to prevent degradation of the prediction accuracy due to plant-model mismatch (PMM), re-identification of the process is required when significant PMM occurs. During re-identification, data containing useful information can be obtained by applying persistent excitation signals to the process. However, at this time, the variations of the

controlled variables usually increase, and thus, the control performance will decrease. Therefore, there is a trade-off between the excitation level of the process and the control performance during data acquisition.

In previous studies, various dual control methods using MPC have been proposed to find the optimal operating condition, considering this trade-off (Shouche et al., 1998; Aggeligiannaki and Sarimveis, 2006; Sotomayor et al., 2009; Zacekova et al., 2013; Gonzalez et al., 2014; Marafioti et al., 2014; Patwardhan et al., 2014; Larsson et al., 2015; Zheng et al.; 2018; Thangavel et al., 2018). These approaches modify the optimization problem of the MPC so that the excitation of the process can be achieved without considerable loss of control performance. Yet, these approaches still have room for improvement. In the modified optimization problem, all the inputs of the process are excited even though only a few elements of the transfer function matrix of the process have large PMM. The excitation of the inputs of transfer functions with small PMM will lead to excessive loss of control performance. Therefore, more efficient re-identification can be realized by exciting only the inputs of the transfer functions with large PMM. As far as the authors know, such a re-identification method has not yet been proposed.

In this research, the re-identification method for multivariate processes using PMM information is proposed. In the proposed method, the excitation signals are applied only to the inputs of the transfer functions with large PMM, to avoid excessive excitation of the process. Furthermore, the other input variables are used to suppress the variations of the controlled variables. The transfer functions with large PMM can be detected using existing methods, such as those proposed by Badwe et al. (2009) and Kano et al. (2010). This allows us to obtain informative data for re-identification while high control performance is maintained. The validity of the proposed method is examined by a simulation case study of a 3-input, 3-output process.

2. Problem setting

In this paper, the multivariate process of interest is given as

$$\boldsymbol{y}(s) = \boldsymbol{G}(s)\boldsymbol{u}(s) + \boldsymbol{v}(s), \tag{1}$$

where $\boldsymbol{y}(s) = [y_1(s), \cdots, y_N(s)]^\top$, $\boldsymbol{u}(s) = [u_1(s), \cdots, u_M(s)]^\top$ and $\boldsymbol{v}(s) = [v_1(s), \cdots, v_N(s)]^\top$ are respectively the output, input, and noise vectors; N and M are respectively the numbers of output variables and input variables; and $\boldsymbol{G}(s)$ is the transfer function matrix of the true process given as

$$\boldsymbol{G}(s) = \begin{bmatrix} G_{1,1}(s) & \cdots & G_{1,M}(s) \\ \vdots & & \vdots \\ G_{N,1}(s) & \cdots & G_{N,M}(s) \end{bmatrix}, \tag{2}$$

where $G_{n,m}(s)$ is a transfer function from the m-th input u_m to the n-th output y_n. The model of the process is

$$\hat{\boldsymbol{y}}(s) = \hat{\boldsymbol{G}}(s)\boldsymbol{u}(s). \tag{3}$$

The process is controlled by MPC using Eq. (3) as the prediction model. The objective function in the optimization problem solved at each time step in the MPC is

$$J = \sum_{t=t_0+1}^{t_0+N_\mathrm{P}} \|\boldsymbol{y}_\mathrm{ref}(t) - \hat{\boldsymbol{y}}(t)\|^2_{\boldsymbol{W}_y} + \sum_{t=t_0}^{t_0+N_\mathrm{C}-1} \|\boldsymbol{u}(t) - \boldsymbol{u}(t-1)\|^2_{\boldsymbol{W}_u}, \tag{4}$$

where t_0 is the current time, $N_\mathrm{P} \in \mathbb{N}$ is the length of the prediction horizon, $N_\mathrm{C} \in \mathbb{N}$ is the length of the control horizon, $\|\boldsymbol{x}\|^2_{\boldsymbol{W}} = \boldsymbol{x}^\mathsf{T}\boldsymbol{W}\boldsymbol{x}$, $\boldsymbol{W}_y$ and $\boldsymbol{W}_u$ are the weighting matrices, and $\boldsymbol{y}_\mathrm{ref}$ is the reference trajectory vector defined as:

$$\boldsymbol{y}_\mathrm{ref}(t) = \gamma^{t-t_0}\boldsymbol{y}(t_0) + (1-\gamma^{t-t_0})\boldsymbol{y}_\mathrm{set}, \tag{5}$$

where $\boldsymbol{y}_\mathrm{set} = \left[y_{\mathrm{set},1}, \cdots, y_{\mathrm{set},N}\right]^\mathsf{T}$ is a setpoint vector, and $\gamma \in [0,1]$ is a parameter.

In this paper, it is assumed that at least one of the transfer functions has significant PMM. As well, the locations of the elements with large PMM in the transfer function matrix are known before data acquisition for re-identification. The following sets are used to define the location of the PMM:

$$\mathbb{I}_{1:M} = \{1,2,\cdots,M\}, \tag{6}$$

$$\mathbb{M}_n = \left\{m \mid \exists\omega > 0 \text{ s.t. } \left|\, G_{n,m}(j\omega) - \hat{G}_{n,m}(j\omega)\right| > \epsilon\right\}, \tag{7}$$

$$\bar{\mathbb{M}}_n = \mathbb{I}_{1:M} \setminus \mathbb{M}_n, \tag{8}$$

$$\mathbb{N}_\mathrm{PMM} = \{n \mid \mathbb{M}_n \neq \phi\}, \tag{9}$$

where $G_{n,m}(j\omega)$ and $\hat{G}_{n,m}(j\omega)$ are the frequency transfer functions of respectively the process and the model, ϵ is a tolerance, and ϕ is the empty set.

3. Proposed Method

In the proposed method, the data acquisition and re-identification of the process are performed as follows:

1. Set $\mathbb{N}^*_\mathrm{MPM} = \mathbb{N}_\mathrm{MPM}$.
2. Set $n^* = \min \mathbb{N}^*_\mathrm{MPM}$ and remove n^* from $\mathbb{N}^*_\mathrm{MPM}$.
3. Operate the target process using MPC which solves the optimization problem defined by Eqs. (10) to (13) at each time step, to obtain the input-output data $\mathbb{D}_{n^*} = \{u_m(t), y_{n^*}(t) \mid m \in \mathbb{M}_{n^*}, t = 1,\cdots,T\}$, where T is sample size, that is,

$$\min_{\boldsymbol{u}(t_0),\cdots,\boldsymbol{u}(t_0+N_\mathrm{C}-1)} J \tag{10}$$

subject to

$$\hat{\boldsymbol{y}}(s) = \hat{\boldsymbol{G}}(s)\boldsymbol{u}(s), \tag{11}$$

$$u_m(t) = \tilde{u}_m(t), \qquad m \in \mathbb{M}_{n^*}, t_0 \le t \le t_0 + N_\mathrm{C} - 1, \tag{12}$$

$$\boldsymbol{u}(t) = \boldsymbol{u}(t_0 + N_\mathrm{C} - 1), \qquad t_0 + N_\mathrm{C} \le t \le t_0 + N_\mathrm{P} - 1, \tag{13}$$

where $\tilde{u}_m$ is an excitation signal applied to u_m.

4. Using $\mathbb{D}_{n^*}$, calculate the estimate $\hat{s}_{\mathbb{M}_{n^*}}$ of the sum of the outputs from the transfer functions with significant PMM in the n^*-th row using

$$\hat{s}_{\mathbb{M}_{n^*}}(t) = y_{n^*}(t) - \sum_{m \in \overline{\mathbb{M}}_{n^*}} \hat{y}_{n^*,m}(t), \qquad y_{n^*}(t) \in \mathbb{D}_{n^*}, \tag{14}$$

where $\hat{y}_{n^*,m}$ is the estimate of the output from $G_{n^*,m}$ and is calculated as follows:

$$\hat{y}_{n^*,m}(s) = \hat{G}_{n^*,m}(s) u_m(s), \qquad u_m(t) \in \mathbb{D}_{n^*}, \tag{15}$$

where $\hat{G}_{n^*,m}$ is a transfer function model from the m-th input u_m to the n^*-th output y_{n^*}.

5. Using the data $\mathbb{D}'_{n^*} = \left\{ u_m(t) \in \mathbb{D}_{n^*}, \hat{s}_{\mathbb{M}_{n^*}}(t) \mid m \in \mathbb{M}_{n^*}, t = 1, \cdots, T \right\}$ as input-output data, re-identify the transfer functions $\left\{ G_{n^*,m} \mid m \in \mathbb{M}_{n^*} \right\}$.

6. If $\mathbb{N}^*_{\mathrm{PMM}} = \phi$, then stop the procedure; otherwise, go back to step 2.

The data acquisition for re-identification is performed in Step 3. Here, the inputs to the transfer functions with significant PMM are excited by the constraint in Eq. (12), while the remaining other inputs are optimized to improve control performance. Note that the degree of improvement in the control performance will be limited when the number of input variables is much smaller than that of the output variables. Steps 4 and 5 are the re-identification steps. Note that, in Step 4, the sum of the outputs from the transfer functions with large PMM in the n^*-th row is estimated without using the models with large PMM.

4. Case study

4.1. Settings

The target process in this paper is a 3-input, 3-output process with first-order transfer functions given as

$$\boldsymbol{G}(s) = \begin{bmatrix} \frac{4.7}{55s+1} & \frac{4.5}{49s+1} & \frac{5.0}{52s+1} \\ \frac{5.2}{48s+1} & \frac{4.6}{52s+1} & \frac{5.5}{50s+1} \\ \frac{4.5}{47s+1} & \frac{5.2}{46s+1} & \frac{4.7}{53s+1} \end{bmatrix}. \tag{16}$$

In this case study, the transfer function model given by Eq. (16) is used as the true process, and the sampling period is set to 1 s. Only the (3, 2)-entry of the transfer function matrix has significant PMM with a steady-state gain 50% larger in the prediction model than in the true process. The output noise vector is defined as

$$\boldsymbol{v}(t) = \left[\frac{1-0.1551q^{-1}}{1+0.8648q^{-1}} w_1(t) \quad \frac{1+0.0464q^{-1}}{1+0.6807q^{-1}} w_2(t) \quad \frac{1+0.9650q^{-1}}{1+0.5256q^{-1}} w_3(t) \right]^{\mathrm{T}}, \tag{17}$$

where $w_n(t) \sim \mathcal{N}(0, 0.1)$ $(n = 1,2,3)$, and q^{-1} is the backward shift operator.

4.2. Procedure

In this case study, the proposed method and the existing method by Shouche et al. (1998) were used for excitation of the process. For each excitation method, the following procedure was conducted 100 times with different seed values for generating w_n. Note that the M-series signals with the clock period of 5 s and amplitudes of 0.2, 0.3, 0.4, and 0.5 were used as $\tilde{u}_2$ in the proposed method; 9 patterns of the parameters, which affect excitation level during data acquisition, were used in the existing method.

Table 1: Parameters of the MPC

Parameters	N_P	N_C	γ	$\boldsymbol{W}_y$	$\boldsymbol{W}_u$
Values	20 s	5 s	0.8	diag(1,1,1)	diag(1,1,1)

I. Five-hundred data samples were acquired during operation using MPC with either the proposed method or the existing method. Here, the MPC parameters were set as in Table 1, and the setpoint vector was set to $[0,0,0]^\mathrm{T}$. The control performance in this step was validated using MAE, which is defined as

$$\mathrm{MAE} = \sum_{n=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\left|y_n(t) - y_{\mathrm{set},n}\right|\right). \tag{18}$$

II. Using the data acquired in Step I, re-identification of the transfer function with a considerable PMM was performed using Steps 4 and 5 of the proposed method. Here, the prediction error method was used as the system identification method.

III. Five-hundred steps of the control simulation were performed to assess the performance of the MPC system after re-identification. Here, the setpoint vector was changed from $[0,0,0]^\mathrm{T}$ to $[2,2,2]^\mathrm{T}$ at $t = 1\,\mathrm{s}$, and MAE was used as the performance index.

4.3. Results

Fig. 1 shows the relationship between the mean MAE in step I (MAE_1) and step III (MAE_2). MAE_1 is smaller when the control performance during data acquisition is high, and MAE_2 is smaller when the model accuracy after re-identification is high. Therefore, we can achieve both higher control performance during data acquisition and higher model accuracy after re-identification as we move in Fig. 1 to the lower left corner. In the existing method, MAE_2 tends to increase steeply as MAE_1 decreases because the sufficiently information-rich data cannot be obtained at the smaller MAE_1 due to the significant reduction of the excitation level. Therefore, the plots for the existing method in Fig. 1 cannot approach the lower left corner. In the proposed method, on the other hand, MAE_2 tends to increase more gently as MAE_1 decreases than in the existing method, and the data points of the proposed method are more closely located to the lower left corner of the figure than those of the existing method. Specifically, MAE_1 of the proposed method can be reduced to 87% of the minimum MAE_1 of the existing method, while keeping MAE_2 smaller than the minimum MAE_2 of the existing method. Thus, the proposed method can realize both higher control performance during data acquisition and higher model accuracy after re-identification than the existing method.

5. Conclusions

In this paper, a new re-identification method for multivariate processes using MPC was proposed. In the proposed method, only the inputs of the transfer functions with large PMMs are excited, and the other inputs are used for suppressing the variations of the controlled variables. The usefulness of the proposed method was validated using a simulation case study of a 3-input, 3-output process with first-order transfer functions. As a result and compared with the existing method, the proposed method improves the control performance during data acquisition by 13% while maintaining a high model accuracy after re-identification. As well, it is expected that the proposed method will be useful when it is applied to more complicated processes, such as unstable processes and non-minimum-phase processes. This will be confirmed in future work.

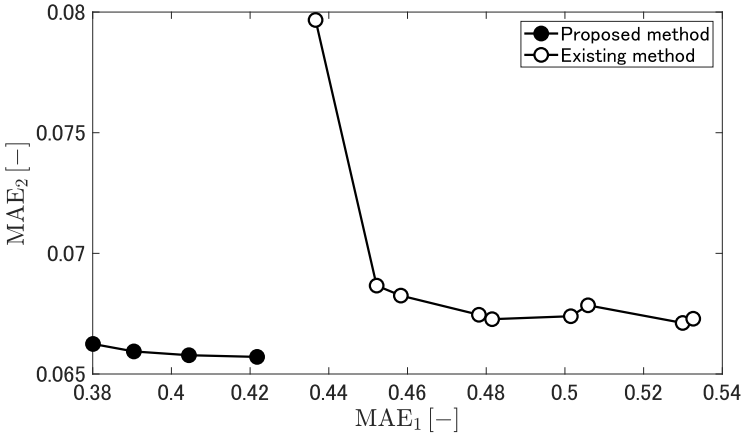


Fig. 1: The relationship between MAE_1 and MAE_2.

References

E. Aggelogiannaki, H. Sarimveis, 2006, Multiobjective constrained MPC with simultaneous closed-loop identification, International Journal of Adaptive Control and Signal Processing, 20, 4, pp. 145-173.

A. S. Badwe, R. D. Gudi, R. S. Patwardhan, S. L. Shah, S. C. Patwardhan, 2009, Detection of model-plant mismatch in MPC applications, Journal of Process Control, 19, 8, pp. 1305-1313.

A. H. González, A. Ferramosca, G. A. Bustos, J.L. Marchetti, M. Fiacchini, D. Odloak, 2014, Model predictive control suitable for closed-loop re-identification, Systems and Control Letters, 69, 1, pp. 23-33.

M. Kano, Y. Shigi, S. Hasebe, S. Ooyama, 2010, Detection of significant model-plant mismatch from routine operation data of model predictive control system, IFAC Proceedings Volumes (IFAC-PapersOnline), 43, 5, pp. 685-690.

C. A. Larsson, C. R. Rojas, X. Bombois, H. Hjalmarsson, 2015, Experimental evaluation of model predictive control with excitation (MPC-X) on an industrial depropanizer, Journal of Process Control, 31, pp. 1-16.

G. Marafioti, R. R. Bitmead, M. Hovd, 2014, Persistently exciting model predictive control, International Journal of Adaptive Control and Signal Processing, 28, 6, pp. 536-552.

R. S. Patwardhan, R. B. Goapluni, 2014, A moving horizon approach to input design for closed loop identification, Journal of Process Control, 24, 3, pp. 188-202.

M. Shouche, H. Genceli, P. Vuthandam, and M. Nikolaou., 1998, Simultaneous constrained model predictive control and identification of DARX processes. Automatica, 34, 12, pp. 1521-1530.

O. A. Z. Sotomayor, D. Odloak, L. F. L. Moro, 2009, Closed-loop model re-identification of processes under MPC with zone control, Control Engineering Practice, 17, 5, pp. 551-563.

S. Thangavel, S. Lucia, R. Paulen, S. Engell, 2018, Dual robust nonlinear model predictive control: A multi-stage approach, Journal of Process Control, 72, pp. 39-51.

E. Žáčeková, S. Prívara, M. Pčolka, 2013, Persistent excitation condition within the dual control framework, Journal of Process Control, 23, 9, pp. 1270-1280.

H. Zheng, T. Zou, J. Hu, H. Yu, 2018, A framework for adaptive predictive control system based on zone control, IEEE Access, 6, pp. 49513-49522.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50061-0

Model Predictive Control of Grade Transition with Attention Base Sequence-to-Sequence Model

Zhen-Feng Jiang[a], Xi-Zhan Wei[a], David Shan-Hill Wong[a*], Yuan Yao[a*], Jia-Lin Kang[b*], Yao-Chen Chuang[c], Shi-Shang Jang[a], John Di-Yi Ou[d]

[a]*Department of Chemical Engineering, National Tsing Hua University, Hsinchu 30043, Taiwan*
[b]*Department of Chemical and Materials Engineering, National Yunlin University of Science and Technology, Yunlin 64002, Taiwan*
[c]*Tsing Hua Intelligent Chemical Engineering Ltd. Co., Hsinchu 30043, Taiwan*
[d]*Center for Energy and Environmental Research, National Tsing Hua University, Hsinchu 30043, Taiwan*
**jlkang@yuntech.edu.tw, dshwong@che.nthu.edu.tw, yyao@mx.nthu.edu.tw*

Abstract

In industry, process input-output data exhibit complex nonlinear dynamics. Such behavior must be modeled by nonlinear time series for use in model-based control, optimization, and monitoring. In this work, a sequence-to-sequence (StS) model was developed for the ASPEN Polymer Plus simulator of an industrial high-density polyethylene (HDPE) slurry reactor. Inclusion of attention mechanism and elastic net (EN) training was found to substantially improve the gain consistency and time dynamics of the model. The resulting model was utilized as a non-linear model predictive control (NLMPC) to control the hydrogen to ethylene ratio (HER) and pressure. The NLMPC can navigate the grade transition of the reactor as well as maintaining the steady state.
Keywords: Sequence-to-Sequence; Attention mechanism; HDPE reactor; Grade transition.

1. Sequence-to-Sequence Model

Chemical processes or unit operations can be described by nonlinear state-space models with x and d being the unknown state and disturbance variables, and y being the observed output variables and u being the observed input variables

$$\frac{dx}{dt} = f(x,u,d), y = g(x) \tag{1}$$

In order to utilize the above model for model predictive control, the unknown x and d must be identified using past observation of u and y in a past window known functions of f and g. For a complex chemical process such as a polymerization reactor, both the identification procedure and the development of physics-based models f and g are nontrivial tasks.

It is desirable that the model development and variable identification procedure can be done in a purely data-driven approach. In such an approach it is necessary to ensure that predictions $\tilde{y}$ of action response in a future horizon, is consistent with the actual process. Previously, Chou et al. (2019) employed a sequence-to-sequence (StS) model developed for natural language processing to establish a data-driven dynamic model for a distillation column. An StS (as shown in Figure 1) consists of an encoder-observer of gated recurrent units (GRU) with input $\hat{u}_{t-i}, \hat{y}_{t-i}, \hat{h}_{t-i-1}$, and a hidden state output $\hat{h}_{t-i}$,

for $i = 1 \cdots W$; and a decoder-predictor GRUs with input $\tilde{u}_{t+i}, \tilde{h}_{t+i-1}$, and output $\tilde{y}_{t+i}$, $\tilde{h}_{t+i}$, for $i = 1 \cdots H$.

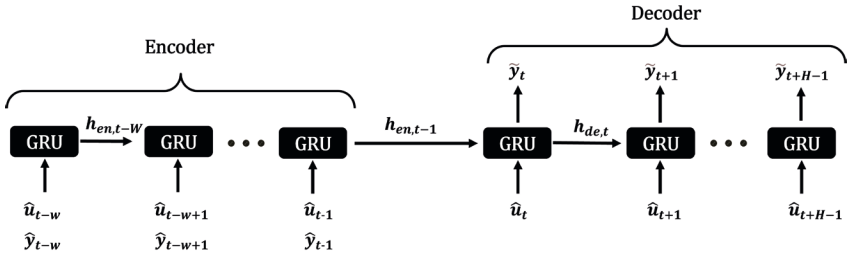


Figure 1. Structure of a StS model

Various forms of similar models have been used for soft sensors (Yuan et al., 2019) and key variable identification (Zhou et al., 2021) to identify latent variables for a time series.

2. Attention Mechanism

The StS structure is shown in Figure 1 leveraged information of the current hidden state identified at the current time t. In a physical state-space model, the knowledge of the current unobserved variables x and d is sufficient for prediction of the future. However, in a data-driven model, there is no guarantee that the observed hidden state h has successfully captured such information, Hence the attention mechanism (Figure 2, Bahdanau et al., 2014) can be included to improve the prediction.

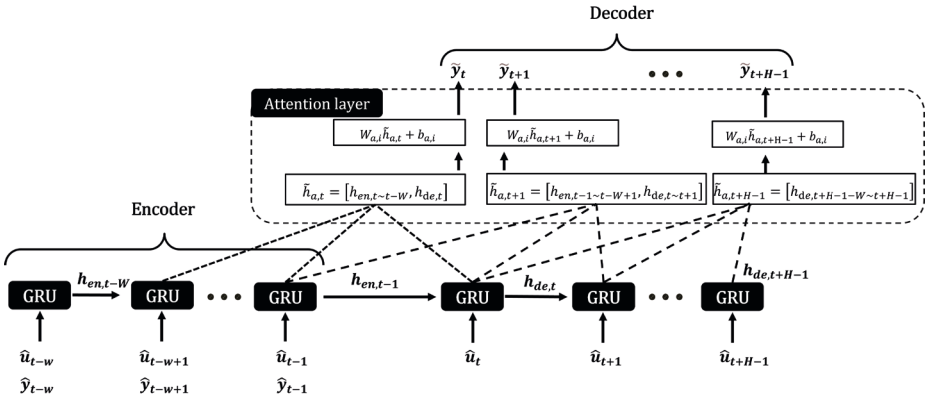


Figure 2. Structure of a StS model with attention

3. Elastic Net and Systematic Model Improvement

Normally the prediction loss for the training model can be given by mean absolute error:

$$\mathcal{L}^{pred} = \frac{\sum_{D \in \mathcal{TR}} \sum_{p=1}^{M} \sum_{t=1}^{H} \|\hat{y}_{p,t,D} - \tilde{y}_{p,t,D}\|}{M \times H} \tag{2}$$

where $\hat{y}_{p,t,D}$ and $\tilde{y}_{p,t,D}$ are observed and predicted values of output respectively, H is the length of decoder, M is the number of sensor variables and the data set D is the training set $\mathcal{TR}$.

To avoid overfitting, elastic net (EN, Zhou and Hastie, 2005) included the l_1 (Lasso) and l_2 (ridge regression) norms of the network weights in the loss function with weight parameters of λ_1 and λ_2

$$\mathcal{L}^{total} = \mathcal{L}^{pred} + \lambda_1 \sum_i |w_i| + \lambda_2 \sum_i w_i^2 \tag{3}$$

The training of such a model is done by minimizing the loss function via adjusting parameters λ_1 and λ_2 of weight w_i. The performance of the trained model is rated by the following combined objective function:

$$\mathcal{J} = R_{\mathcal{V}}^2 + \mathcal{G}_{\mathcal{V}} \tag{4}$$

where $R_{\mathcal{V}}^2$ is the coefficient of determination of the validation set $\mathcal{V}$ and $\bar{y}$ is the average of observed value $\hat{y}_D$.

$$R_{\mathcal{V}}^2 = 1 - \frac{\sum_{\mathrm{D}\in\mathcal{V}} \Sigma_{p=1}^{M} \Sigma_{t=1}^{H} \left\| \hat{y}_{p,t,D} - \tilde{y}_{p,t,D} \right\|^2}{\sum_{D\in\mathcal{V}} \Sigma_{p=1}^{M} \Sigma_{t=1}^{H} \left\| \hat{y}_{p,t,D} - \bar{y}_D \right\|^2} \tag{5}$$

and the gain consistency $\mathcal{G}_{\mathcal{V}}$ of the validation set $\mathcal{V}$.

$$\mathcal{G}_{\mathcal{V}} = \frac{\sum_{i\in\mathcal{C}, j\in\mathcal{M}} \left(\sum_{D\in\mathcal{V}} \Sigma_{t=1}^{T} Con_{t,ij,D} \right)}{T_{\mathcal{V}}} \tag{6}$$

where $\mathcal{C}$ is the set of all controlled variables (CV) and $\mathcal{M}$ is the set of all manipulated variables (MV). A dynamic gain $K_{t,ij}$ is obtained by changing a manipulated variable $u_j, j \in \mathcal{M}$ with a value $\Delta u_{t,j}$ at time instant t and finding the change in the predicted output $\Delta \tilde{y}_{t,j}$ in the future horizon of the controlled variable $\tilde{y}_i, i \in \mathcal{C}$.

$$K_{t,ij} = \frac{\Delta \tilde{y}_{t,i}}{\Delta u_{t,j}} \tag{7}$$

The consistency $Con_{t,ij}$, a binary variable is given by

$$Con_{t,ij} = Heaviside\left(K_{t,ij} \langle {K^{ss}}_{ij} \rangle \right) \tag{8}$$

where ${K^{ss}}_{ij}$ is the expected steady-state gain provided by physics-based simulation or experience of the operator. The $Heavside$ function indicated that we are only concerned with the sign of the gain rather than the absolute value. This is because the sign of the gain is usually known based on understanding the physics of the process while predicting the absolute value required a lot of modeling efforts.

Since both $R_{\mathcal{V}}^2$ and $\mathcal{G}_{\mathcal{T}\cup\mathcal{V}}$ are between 0 and 1, they can be summed and optimized without any weighting factors. The following optimization problem can be solved by any global optimization procedure such as the differential evolution algorithm (DEA, Qin et al., 2008) to find the best solution $\max_{\lambda_1,\lambda_2} \mathcal{J}$ in terms of prediction accuracy and gain consistency.

Such a model should serve as an adequate model to navigate and control a process. In our previous work (Chou et al., 2019), we found that $R_{\mathcal{V}}^2$ and $\mathcal{G}_{\mathcal{T}\cup\mathcal{V}}$ are not correlated when a simple deep neural network. The good values of $R_{\mathcal{V}}^2$ and $\mathcal{G}_{\mathcal{T}\cup\mathcal{V}}$ can be obtained by using an StS model. In this work, the aforementioned procedure of using EN and optimization with respect to parameters of EN constitutes a systematic improvement of our process model.

4. A High-Density Polyethylene Reactor

An ASPEN Polymer Plus dynamic model of an HDPE reactor was developed using the kinetic model provided by Khare et al. (2002). Sensors and controllers data of daily steady-state and grade transition operations were generated based on the operation of a local plant with varying catalyst activities. MV and CV in the HDPE polymerization process as shown in Table 1. The operator employed catalysts flow and hydrogen flow as daily MV to control the HER and pressure, The directionality of MV/CV pairs utilized for determining gain consistency is shown in Table 2.

Table 1. List of variables

Type	Tag
MV	Flow rate of ethylene
MV	Flow rate of hydrogen
MV	Flow rate of 1-butene
MV	Flow rate of catalyst
MV	Flow rate of cocatalyst
MV	Flow rate of hexane
MV	Temperature of reactor
CV	Pressure of reactor
CV	HER of purge gas

Table 2. Directionality of gain

	HER	Pressure
Flow rate of catalyst	+	−
Flow rate of hydrogen	+	+

5. Dynamic Modelling

The generated data were sampled at 10 minutes intervals. The window length of encoder and decoder horizon consists of 24, and 18 samples, or 240 and 180 minutes respectively; to ensure that the model can consider time delay and slow dynamic of the system. The data of transition between different grades are distributed to the training, validation, and testing data set according to the ratio of 6:2:2. There are six kinds of grade transition in total, five of which are assigned to the training, validation data set. The other is assigned to the testing data set and utilized as the indicator of grade transition navigation.

Comparison of prediction accuracy and gain consistency results of the StS model and StS with attention mechanism (StS-ATT) model for the testing data is given in Table 3. Substantial improvement can be achieved by including the attention mechanism.

Table 3. The prediction RMSE and R^2 result of StS and StS-ATT model in the testing dataset.

Model	Pressure		HER		$\mathcal{G}_{\mathcal{T}e}$
	$RMSE_{\mathcal{T}e}$	$R^2_{\mathcal{T}e}$	$RMSE_{\mathcal{T}e}$	$R^2_{\mathcal{T}e}$	
StS	0.596	0.898	0.064	0.957	0.988
StS-ATT	0.138	0.995	0.047	0.977	0.999

Figure 3 (a) shows the HER contribution plot of the hidden states at various time points in the past window. The contribution peaks at around 80 minutes in the past showing a long time characteristic time and possible time delay of the system. Figure 3 (b) demonstrated substantial improvement in gain consistency during the optimization of EN parameters due to the fact that gain consistency was not included in the neural network parameter training.

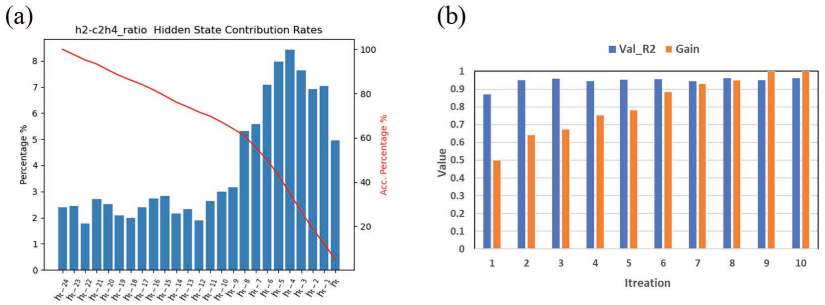


Figure 3. (a)HER contribution plot of attention state in the past (b)The result of gain consistency and R squared at various iteration points.

6. Grade Transition Navigation

To demonstrate the usefulness of the dynamic input-output model developed, it is applied to operation navigation of grade transition of the HDPE reactor. An artificial intelligence model predictive control (AIMPC) algorithm was developed based on our StS-ATT model using the following bounded optimization:

$$\min_{u_{j,t+k},k=1\cdots H,j\in\mathcal{M}} \sum_{k=1}^{H}\left\{\sum_{i\in\mathcal{C}}\left|\tilde{y}_{i.t+k}-y_i^{sp}\right|^2+\alpha\sum_{j\in\mathcal{M}}\left|u_{j.t+k}-u_{j.t+k-1}\right|^2\right\} \quad s.t. \quad u_{j.lb}\le u_{j.t+k}\le u_{j.ub} \tag{9}$$

The DEA calculates the minimized change of manipulated variable $u_{j.t+k}-u_{j.t+k-1}$ at each point of time to make the prediction $\tilde{y}_i$ closer to setpoint y_i^{sp}. Simultaneously, the manipulated variable $u_{j.t+k}$ at each point of time will be within the operating upper and lower limits, $u_{j.ub}$ and $u_{j.lb}$. Again the DEA is used to solve optimization and the solution is implemented on the ASPEN Polymer Plus dynamic model to simulate the grade transition. The operation changes in HER, pressure, catalyst flow, and hydrogen flow are shown in Figure 4(a) to (d) respectively. The blue line is the result of AIMPC and the red line is a manual control procedure suggested by the operator. It is found that the pressure is more stable than the manual control as shown in Figure 4 (b), and the operation navigation adjusts more quickly and amplitude is relatively stable than the manual control to make the HER achieve and stay on the setpoint 1 as shown in Figure 4(a).

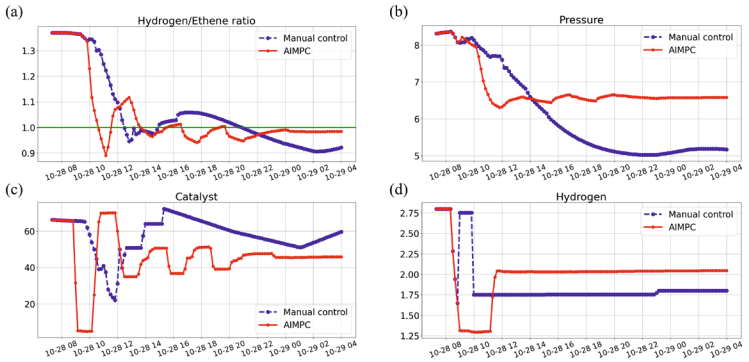


Figure 4. AIMPC with the StS-ATT (a) the H_2/C_2H_4 ratio (b) pressure of reactor (c) the flow rate of catalyst (d) the flow rate of hydrogen.

7. Conclusion

In this work, the StS-ATT model is used to model the dynamics of a high-density polyethylene reactor without the knowledge of catalyst activity, polymerization kinetics, or other first principle knowledge. Furthermore, the physical consistency of the gain relation, which is between critical manipulated variables and sensor variables, and differential evolution optimization of weight parameters of L1 and L2 norm. Such improvement allows the application of model predictive control of grade transitions. The results showed that the model predictive control of grade transition using the dynamic model is more efficient and stable compared to manual transition based on operator advice. The optimized operation guidelines can be found that a much quicker transition can be achieved by using the model predictive control of our data-driven model.

Acknowledgment

This work was supported by the Ministry of Science and Technology, ROC (Grant number: MOST 110-2221-E-007 -014 -).

References

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.

Chou, C. H., Wu, H., Kang, J. L., Wong, D. S. H., Yao, Y., Chuang, Y. C., ... & Ou, J. D. Y. (2019). Physically consistent soft-sensor development using sequence-to-sequence neural networks. IEEE Transactions on Industrial Informatics, 16(4), 2829-2838.

Khare, N. P., Seavey, K. C., Liu, Y. A., Ramanathan, S., Lingard, S., & Chen, C. C. (2002). Steady-state and dynamic modeling of commercial slurry high-density polyethylene (HDPE) processes. *Industrial & engineering chemistry research*, *41*(23), 5601-5618.

Qin, A. K., Huang, V. L., & Suganthan, P. N. (2008). Differential evolution algorithm with strategy adaptation for global numerical optimization. IEEE transactions on Evolutionary Computation, 13(2), 398-417.

Yuan, X., Gu, Y., Wang, Y., Yang, C., & Gui, W. (2019). A deep supervised learning framework for data-driven soft sensor modeling of industrial processes. IEEE transactions on neural networks and learning systems, 31(11), 4737-4746.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the royal statistical society: series B (statistical methodology), 67(2), 301-320.

Zhou, J., Wang, X., Yang, C., & Xiong, W. (2021). A novel soft sensor modeling approach based on Difference-LSTM for complex industrial process. IEEE Transactions on Industrial Informatics.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50062-2

Real Time Optimization of series of fixed bed catalytic reactors

Naganjaneyulu Suruvu [a*] , Kazuya Ijichi [a], Satoru Hashizume [a]

[a]*Sumitomo Chemical Co., Ltd , 5-1, Anesakkaigan, Ichihara 299-0195, Japan*
suruvun@sc.sumitomo-chem.co.jp

Abstract

Real time optimization has become an increasingly important subject in the chemical industry due to high competition. The combination of process modelling and computer simulation provides clear understanding to know about the improvement potential of the plant.

This paper introduces the real time optimization, and its online implementation in series of fixed-bed catalytic reactors process. In general, catalyst deactivation occurs in most of the fixed bed reactors, and optimal operation of reactor systems undergoing catalyst deactivation is an important economic issue. In the process of controlling the chemical reaction, it is necessary to change the operating conditions according to the catalyst activity. A system is developed to deal with the depreciating catalyst activity by using the concept of mathematical optimization methods that allow to calculate the optimal operating conditions of multiple variables simultaneously. The system consists of 3 modules, (i) Analyser: This module consists of steady state one-dimensional plug flow reactor model, which automatically receives the plant historian data as well as lab analysis data such as flow rate, temperature, concentration etc. to estimate the catalyst activity. (ii) Predictor: This module develops the empirical equation using the activity data from the analyser with the help of Recursive Least Square method. This empirical model estimates the real time catalyst activity where no process data is available. In general the catalyst activity is assumed to decrease at a constant rate with respect to the age as given by the vendor information. (iii)Optimizer: In this module the objective function and its constraints are defined. Minimization of objective function is carried out using appropriate algorithm while ensuring the product quality constraints.

Online implementation is carried out by connecting this system with plant historian data and displaying the calculated optimum conditions on the dashboard. The operators adjust the process conditions based on the dashboard. As a result it is determined that the production cost has significant reduction.

Keywords: Control; Optimization; Operation; Digital Twin.

1. Introduction

Catalytic fixed bed reactors are widely used in the chemical industries, from refinery to the fine chemicals. In a general chemical plant life cycle, once the plant has designed, constructed and started operations, the duty of the process engineer is to optimise the operating conditions in order to realize the maximum production with minimum production cost and high profits. In order to minimize the cost, it is necessary to decrease the by-products formation rate and improve the yield of desired product. The performance of the reactor system is adversely affected by the catalyst decay. As the

catalyst deactivates, the performance of the reactor decreases which means conversion rate of raw materials decreases. A common operation strategy to increase the conversion is to adjust the operating condition such as temperature of the reactor to compensate the decrease in activity. Change in the temperature leads to change in conversion rate as well as the by-products formation rate. In order to optimize the process with minimum by-products and maximum conversion, a real time optimization model is necessary to simulate, understand the potential and to achieve the better economic efficiency of the plant. A number of variables are involved in the optimization of the fixed bed reactors, such as feed composition, catalyst activity, flowrates, bed temperature etc. Feed composition, flowrates and temperature are online measurable variables and catalyst activity is difficult to measure online. There has been a lot of works (Biscarri et al., 2012 & Fuada et al., 2012) carried out on optimization of fixed bed reactors undergoing catalyst deactivation. But in most of the cases the catalyst activity is assumed to decrease at a constant rate with respect to the age as given by the vendor information. The vendor information is based on the ideal conditions, where as in real case, overall catalyst bed activity depends so many factors like porosity, historical operation conditions in catalyst life such as feed flow rate, temperatures etc.

In this article, a Digital Twin technology is proposed to deal with the optimization problem. A Digital Twin is a simulator that reflects the real system behaviour with maximum possible precession. The proposed Digital Twin estimates the overall catalyst bed activity in real time and based on that activity the system calculates the optimum conditions. This Digital Twin consist of a steady state plug flow reactor model to simulate the series of fixed bed catalytic reactors. The catalyst activity is estimated using lab analysis data such as inlet and outlet concentrations of the reactant and process historian data such as temperature, flowrate etc. Along with the reactor model a regression equation is developed to estimate the catalyst activity in the absence of lab data.

2. Digital Twin Development

In this section, detailed procedure of Digital Twin development is described. Before going into the Digital Twin development, brief description about the process is presented here. The raw materials (A & B in below equations) are passed through the series of fixed bed reactors and the reactants convert into the products over the catalyst. Along with the desired products C, undesired products (D, E, and F) are also formed from the reaction. Products and unreacted fluids absorbs the heat generated from the reaction and carry on to the next process. As shown in Figure 1, the inlet temperature of the reactor is controlled by manipulating the utility of heat exchanger.

$$A + B \rightarrow C\ (\text{desired product}) + \Delta H\ (\text{heat}) \quad \text{(i)}$$

$$A + B \rightarrow D\ (\text{undesired product}) + \Delta H1\ (\text{heat}) \quad \text{(ii)}$$

$$C + B \rightarrow E\ (\text{undesired product}) + \Delta H2\ (\text{heat}) \quad \text{(iii)}$$

$$D + B \rightarrow F\ (\text{undesired product}) + \Delta H3\ (\text{heat}) \quad \text{(iv)}$$

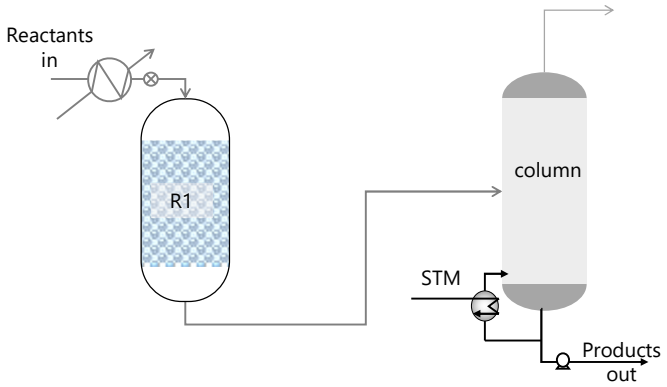


Figure 1. Simplified process flow

This system is configured with following 3 modules. All these modules are developed using python. The detailed explanation of each module are as follows.

I. Analyser

This module consists of steady state one-dimensional plug flow reactor model, which automatically receives the plant historian data as well as lab analysis data like concentration etc. to estimate the catalyst activity. Algorithm for the analyser is shown in the Figure 2.

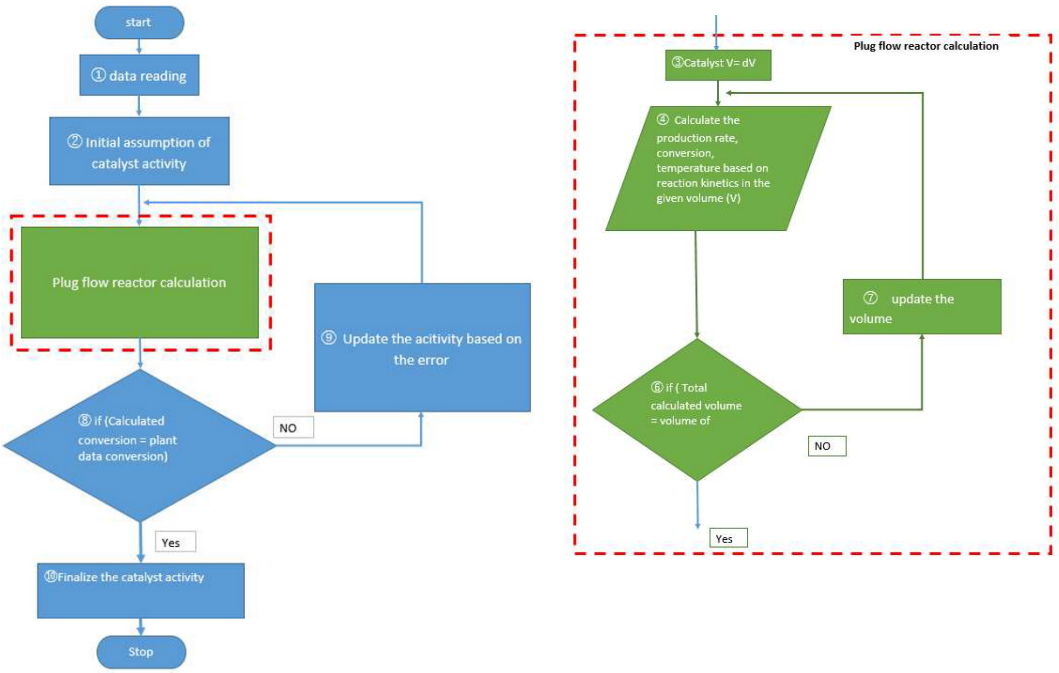


Figure 2. Calculation flow chart for the Analyser

In step 1, Analyser reads the data from the Plant historian, and does all the pre-processing like data cleaning, etc.

In step 2, initial value for the catalyst activity is assumed in order to carry out the plug flow reactor calculations.

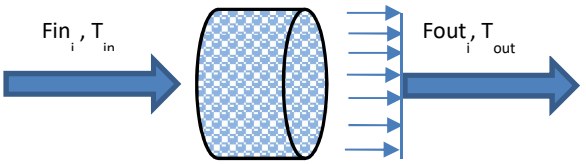


Figure.3 Mass and Energy balance around tiny part of reactor

In step 3 to step 7, plug flow reactor calculation is carried out by dividing the reactor into small parts and in each part the material & energy balance equation (v) to (vii) reported below are used to calculate the temperature and concentration of each component.

$$Fin_i - Fout_i = r_i \cdot \Delta V \quad \text{(v)}$$

$$r_i = K_i \cdot e^{-\left(\frac{Ea_i}{RT}\right)} \quad \text{(vi)}$$

$$\rho \cdot C_P(T_{in} - T_{out}) = Q_{rxn} \quad \text{(vii)}$$

Where i refers to component and K_i is pre-exponential factor which is directly proportional to the catalyst activity and the relation is derived from the experimental data for each component. Q_{rxn}, ρ, C_P, T are the reaction heat, density, specific heat and temperature respectively.

In step 8, once the entire plug flow reactor calculation is done, the overall reactor conversion rate is calculated using concentration result obtained. This conversion rate is compared with conversion rate calculated from the lab analysis data. If both conversion rates are not equal then catalyst activity is updated based on the error and the entire calculation procedure from step 3 is repeated. If both the conversion rates are equal then catalyst activity is finalized and updated into the plant historian from the module.

II. Predictor

In this module, the catalyst activity data and catalyst run time is obtained from the plant historian. Using this data, regression equation is developed to predict the catalyst activity in real time when there is no lab analysis data. Coefficients of regression equations are estimated using recursive least square method. This coefficients are updated to the plant historian.

III. Optimizer

In this module, the objective function as well as its constraints are defined. Minimization of objective function is carried out using appropriate algorithm while ensuring the product quality constraints.

The objective function is the value of loss in production cost which includes raw material loss due to undesired products and utility loss. The price of raw material and utility are included in the function. Process constraints such as concentration of by-product (F) which effects quality of products are also included in the function.

Optimization problem can be represented in the following equations

Minimization of Objective function $f = (\text{cost of Material loss} + \text{cost of utility loss})$ (1)

Subject to (impurities quantity) $x_F < z$ (2)

Decision variable are temperature and flowrate for the reactor.

Where, x_F is the concentration of component F, z is the product specification (Target value of impurities)

By-products formation rate is calculated from the process data such as catalyst activity, flowrates, temperature etc. and plug flow reactor model. Raw material loss is calculated from the by-product formation rate. Utility loss is calculated based on the reactor temperature. With help of both heat and material loss optimum operating conditions are derived using the optimization algorithm.

When coming to multivariable optimization problem, searching for global optimization is crucial. However, popular methods like Gradient, Newton etc. does not guarantee the global optima, in most of the times these methods struck in the local optima. And the same time it is difficult to do deterministically as it is non-linear problem. Recently meta-heuristic algorithms have become a topic of interest in multivariable optimization. In this system, PSO (Particle Swarm Optimization) (Deng et al., 2013), one of the meta-heuristic algorithms is used to calculate the optimum operation conditions.

3. Online Implementation

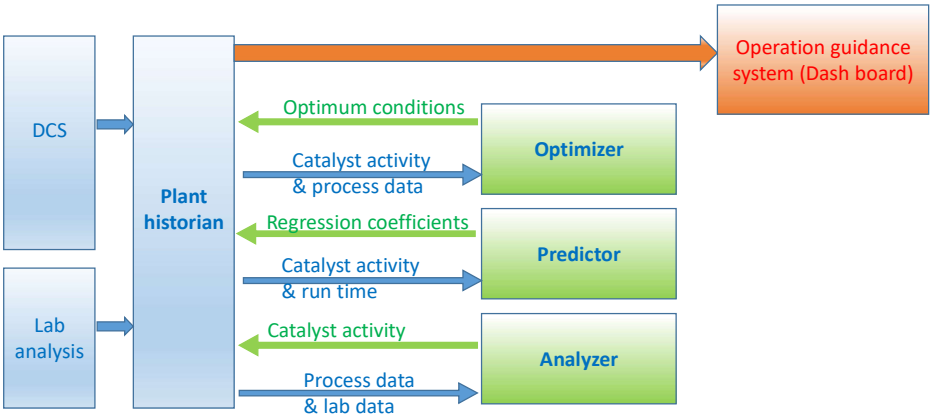


Figure 4. Online implementation Framework

As shown in the Figure 4, all the modules are connected to the plant historian to receive the plant process data as well as lab analysis data. Each module perform their task as per the schedule and write back the result to the plant historian. With help of commercial package, dashboard (operation guidance system) is developed and the plant optimum operating conditions are displayed on operation guidance system. The plant operator adjusts the operating conditions according to the display.

4. Results

In this section, optimum operating condition are compared with the conventional operation. In conventional operation, temperature of the reactors are determined by monitoring the raw materials in the exit of reactor from the lab analysis. When the raw material is detected in the exit of the reactor, the temperature of the reactor is increased. This increase in temperature leads to increase in conversion rate but at the same time it might decrease the selectivity. So, it is necessary to calculate the optimum temperature to balance conversion rate & selectivity.

In order to quantify the changes in the production cost with respect to catalyst activity, comparison of conventional operation and optimum operation is carried out. Figure 5 shows the difference between optimum operation and conventional operation. As shown in the figure, in the first half of catalyst life, the production cost is same (assuming 100% for fresh catalyst) in both the conditions. But coming to the later half of the catalyst life, production cost is less in optimum operating conditions. Based on the results, optimum operating conditions have good impact on the economic point. This system also considers the real time market changes of utility and material cost.

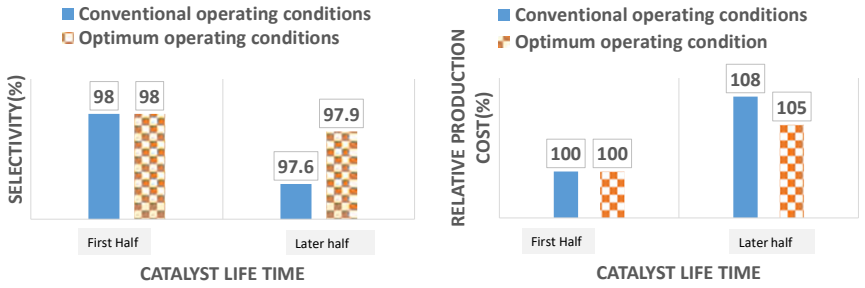


Figure 5. Comparison of conventional operating conditions and optimum operating conditions

5. Conclusions

Developed Digital Twin model allows to estimate the current activity of the catalyst and from that activity it also estimates the optimum operating conditions while ensuring the product quality constraints. A comparison study between conventional operation and optimum operation is conducted under relevant reaction conditions such as same catalyst volume and same inlet flowrates and concentrations. Decrease in production cost and increase in selectivity has achieved using this Digital Twin. These Digital Twins are one of the effective method for the processes with deactivating catalyst.

References

F.Biscarri, I.Monedero, C.León, J.I.Guerrero, R.González, L.Lombard, 2012, A Decision Support System for consumption optimization in a naphtha reforming plant, Computers and Chemical Engineering, 44, 1-10

M.Fuada, M.Hussaina, A.Zakaria, 2012, Optimization strategy for long-term catalyst deactivation in a fixed-bed reactor for methanol synthesis process, Computers and Chemical Engineering, 44, 104– 126

Y.Deng, Q.Jiang, Z.Cao, 2013, Online Optimization of Industrial FCC Unit Based on PSO Algorithm nad RBF Neural Network、(AASRI-WIET 2013)

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50063-4

Self-triggered MPC for Perturbed Continuous-time Non-linear Systems

Yueyang Luo[a], Xinmin Zhang[a]*, Zhihuan Song[a]

[a]*State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China*
xinminzhang@zju.edu.cn

Abstract

Recently model predictive control has made great progress both in theory and in practical. To further reduce the computational load, this paper introduces a self-triggered mechanism in model prediction control with a decreasing prediction horizon for continuous-time non-linear systems subject to bounded disturbances and certain constraints. Under this strategy, the next updating point of the optimal control sequence is determined according to the current system behaviours instead of the fixed sampling period. Besides, a dual-mode scheme is implemented based on the terminal region. Both the feasibility of the algorithm and the convergence of the controlled system are proved. The application results on a numerical example and a practical system demonstrated the superiority of the proposed strategy.

Keywords: model predictive control, self-triggered mechanism, continuous-time nonlinear systems, dual-mode scheme

1. Introduction

Model predictive control (MPC), also known as receding horizon control (Morari, 1999), mainly consists of prediction model, rolling optimization, and feedback correction (Ding, 2010). Prediction model aims to show the dynamic behaviour of the controlled system in the future. Previous information of the system and the upcoming inputs can be used to obtain the prediction of the nominal system, which therefore provides the prior messages for the optimal algorithm. Since it is hard for the model-based prediction to match the actual controlled process precisely, rolling optimization is adopted.

So far, MPC has also been extensively studied and developed in theory (Mayne, 2000) (Mayne, 2014). Some scholars proposed a self-triggering control strategy that adaptively selects the sampling interval for input affine nonlinear systems (Hashimoto, 2016). Different from the normal pattern of MPC, the next update instant can be pre-calculated based on the current information under the self-triggered mechanism (Sun, 2019). Therefore, the computational burden of carrying out self-triggered MPC is decreased both on frequency and one-shot complexity of solving the OCP with the adaptive prediction horizon.

A predictive control strategy based on self-triggered mechanism is proposed for nonlinear continuous-time systems subject to external additive disturbance and constraints of states and inputs. In the framework of dual-mode model predictive control (Michalska, 1993), sufficient conditions are derived to guarantee the recursive feasibility in MPC control mode, the convergence to the terminal region within finite time, and the stability

of the system after entering the terminal region. Furthermore, a simulation example is presented to verify the feasibility and effectiveness of the proposed strategy.

The remainder of this paper is organized as follows. Section 2 gives the description of system model and preliminary. In Section 3, the main problems to be solved are introduced, and the associated strategy in MPC control mode is described. In Section 4, simulation results are provided. Section 5 draws the conclusions.

2. Preliminary

2.1. System Model

The perturbed continuous-time nonlinear system and its nominal form are considered as:

$$\dot{x}(t) = f\big(x(t), u(t)\big) + w(t), \quad x(0) = x_0 \tag{1}$$

$$\dot{\hat{x}}(t) = f\big(\hat{x}(t), \hat{u}(t)\big), \quad \hat{x}(0) = x_0 \tag{2}$$

where $x(t) \in R^n$, $u(t) \in R^m$, and $w(t) \in R^n$ are the state, control input, and bounded disturbance, respectively. Required constraints are presumed as follows:

$$x(t) \in \mathcal{X}, \quad u(t) \in \mathcal{U}, \quad w(t) \in \mathcal{W} \tag{3}$$

Specifically, both $\mathcal{X} \subseteq \mathcal{R}^n$ and $\mathcal{U} \subseteq \mathcal{R}^m$ are compact sets containing the origin as an interior point. $\mathcal{W} \triangleq \{w \in \mathcal{R}^n : \|w\| \leq \rho\}$ represents the boundary of the disturbance. The system function $f(x,u)$ is a twice continuously differentiable nonlinear function satisfying $f(0,0) = 0$, and is stabilizable in linearization form (Dunbar, 2007). Besides, it is assumed that $f(x,u)$ is Lipschitz continuous respect to $x \in \mathcal{X}$ with Lipschitz constant L_f, i.e.,

$$\|f(x_1, u) - f(x_2, u))\|_P \leq L_f \|x_1 - x_2\|_P \tag{4}$$

2.2. Lemma 1

The Gronwall-Bellman inequality in continuous-time form is introduced. If function $\beta(\cdot): \mathcal{R} \to \mathcal{R}_{>0}$ satisfies the following inequality (Rawlings, 2017):

$$\mu(t) \leq \alpha(t) + \int_a^t \beta(s)\mu(s)ds\,, t \in [a,b]$$

Then, for $t \in [a,b]$, we have:

$$\mu(t) \leq \alpha(t) + \int_a^t \alpha(s)\beta(s) \exp\left(\int_s^t \beta(r)dr\right) ds\,, t \in [a,b]$$

Moreover, if $\alpha(\cdot): \mathcal{R} \to \mathcal{R}$ is a nondecreasing function, then a more explicit conclusion can be drawn:

$$\mu(t) \leq \alpha(t) \exp\left(\int_a^t \beta(s)ds\right), t \in [a,b] \tag{5}$$

2.3. Lemma 2

For the system (2)(1), within the scope that $X_r = \{\hat{x}: \|\hat{x}\|_P^2 \leq r^2\}$, there exists a state feedback $Kx \in U$ such that:

$$\frac{d(\|\hat{x}(t)\|_P^2)}{dt} \leq -\|\hat{x}(t)\|_\Phi^2 \tag{6}$$

Suppose A and B are the linearization matrices. Given a $Q > 0$, $R > 0$ and K satisfying that $\lambda(A + BK) < 0$, there is a Lyapunov equation to comply with:

$$P(A + BK + \kappa I) + (A + BK + \kappa I)^T P = -\Phi$$

where $\Phi = Q + K^T RK$.The value of κ is selected to ensure $\lambda(A + BK + \kappa I) < 0$.

3. Description for OCP and Triggering Mechanism

3.1. OCP

At the triggering instant, the optimal control trajectory can be obtained by solving an optimal control problem as follows:

$$\begin{aligned}
\hat{u}^*(s;t_k) &= \arg\min_{\hat{u}(s;t_k)} J_N(\hat{x}(s;t_k), \hat{u}(s;t_k), T_k) \\
&= \arg\min_{\hat{u}(s;t_k)} \int_{t_k}^{t_k+T_k} \|\hat{x}(s;t_k)\|_Q^2 + \|\hat{u}(s;t_k)\|_R^2 \, ds \\
&\quad + \|\hat{x}(t_k + T_k; t_k)\|_P^2 \\
s.t. \dot{\hat{x}}(s;t_k) &= f\big(\hat{x}(s;t_k), \hat{u}(s;t_k)\big), \quad \hat{x}(t_k;t_k) = x(t_k) \\
&\hat{u}(s;t_k) \in \mathcal{U} \\
&\hat{x}(s;t_k) \in \mathcal{X}_{s-t_k} \triangleq \mathcal{X} \ominus \mathcal{T}_{s-t_k} \\
&\hat{x}(t_k + T_k; t_k) \in \mathcal{X}_\varepsilon \triangleq \{x: \|x\|_P^2 \le \varepsilon^2\}
\end{aligned} \tag{7}$$

$\mathcal{X}_{s-t_k}$ represents the contracted constraints at the s instant from t_k. Moreover, $\mathcal{T}_{s-t_k}$ denotes the deviation between the predictive state from the actual state at s. The detailed definition of $\mathcal{X}_{s-t_k}$ and $\mathcal{T}_{s-t_k}$ will be discussed below.

3.2. Self-triggered Strategy

At the current time t_k, the interval between t_k and t_{k+1} and the reduction of prediction horizon are defined as follows:

$$m_{t_k} \triangleq t_{k+1} - t_k$$

$$n_{t_k} = T_k - T_{k+1}$$

$$m_{t_k} = min\{\hat{m}_{t_k}, \breve{m}_{t_k}, T_k\} \tag{8}$$

$$\hat{m}_{t_k} = sup\{m_{t_k} > 0: \bar{\lambda}(\sqrt{P}) \cdot \rho \cdot m_{t_k} \cdot e^{L_f T_k} \le r - \varepsilon\} \tag{9}$$

$$\begin{aligned}
\breve{m}_{t_k} = sup\{m_{t_k} > 0: &\bar{\lambda}(\sqrt{P}) \cdot \rho \cdot m_{t_k} \cdot e^{L_f T_k} \cdot (r + \varepsilon) \\
&+ \int_{t_{k+1}}^{t_k+T_k} \Big[\big(\bar{\lambda}(Q) \cdot \rho \cdot m_{t_k} \cdot e^{L_f(s-t_k)}\big)^2 \\
&2\bar{\lambda}(Q)\rho m_{t_k} e^{L_f(s-t_k)} + \|\hat{x}^*(s;t_k)\|_Q^2\Big] ds \\
&\le \sigma \int_{t_k}^{t_{k+1}} \big(\|\hat{x}^*(s;t_k)\|_Q^2 + \|\hat{u}^*(s;t_k)\|_R^2\big)\, ds\}
\end{aligned} \tag{10}$$

$$T_{k+1} = min\{m_{t_k}, \widetilde{T_k}\} \tag{11}$$

where σ ∈ (0,1), and $\widetilde{T_k}$ is defined as:

$$\widetilde{T_k} = inf\{0 \le h < T_k : \hat{x}^*(t_k + h; t_k) \in \mathcal{X}_\varepsilon\} \tag{12}$$

Algorithm 1 Dual-mode Self-triggered MPC

while at time t_k **do**
  if $x(t_k) \in \mathbb{X}_\varepsilon$ **then**
    Switch to $\kappa(x)$ to stabilize the system;
  else
    Update the initial state of OCP by $\hat{x}(t_k) = x(t_k)$;
    Solve the OCP in (7) to obtain the optimal predicted control sequence $\hat{u}^*(t_k)$ as well as the state trajectory $\hat{x}^*(t_k)$;
    Calculate the next triggering instant t_{k+1} by (8) and the next prediction horizon T_{k+1} by (11);
    Apply the first m_{t_k} control input in $\hat{x}^*(t_k)$ to (1);
  end if
  Assign $t_k = t_{k+1}$;
end while

4. Simulation

In this section, a cart-damper-spring system are given to verify the effectiveness of the proposed control scheme. The application results are compared with the conventional time-triggered MPC. The specific application of a cart-damper-spring system is presented by the following dynamics (Li, 2014):

$$\begin{cases} \dot{x}_1(t) = x_2(t) \\ \dot{x}_2(t) = -\dfrac{\zeta}{M_c} e^{-x_1(t)} x_1(t) - \dfrac{h_d}{M_c} x_2(t) + \dfrac{u(t)}{M_c} + \dfrac{v(t)}{M_c} \end{cases} \tag{13}$$

where $x_1(t)$ and $x_2(t)$ denote the location and velocity of the cart, respectively. $\omega(t)$ is the external disturbance bounded by $\|v(t)\| \le 0.0025$. The weight of the cart is $M_c = 1.25kg$, and the stiffness of the spring is $\zeta = 0.9N/m$. The damper factor is $h_d = 0.42N/m$. The input $u(t)$ is constrained as $-0.9 \le u(t) \le 0.9$. The state constraint set is given by $\mathcal{X} = x: \|x_1\| \le 0.35, \|x_2\| \le 1$. The stage weighting matrix Q is $Q = \begin{pmatrix} 0.06 & 0 \\ 0 & 0.06 \end{pmatrix}$, the input weighting matrix R is 0.001, and then the terminal weighting matrix can be calculated as $P = \begin{pmatrix} 0.1248 & 0.0260 \\ 0.0260 & 0.0358 \end{pmatrix}$. The Lipschitz constant of the system in (13) is $L_f = 0.1703$, and the feedback matrix K is $K = (-1.6000 \quad -2.3300)$. The terminal parameters are presented as $\varepsilon = 0.072$ and $r = 0.076$. The compromising factor σ is 0.12, and the initial state is $x_0 = (0.3 \quad -1)^T$. Fig.1 presents the state and input trajectory. Fig.2 denotes the triggered instants and prediction horizons at each sampling instant t_k. Similar to the former numerical case, the validity of the strategy has been shown from a drastic descension in triggering interval and a moderate falling in prediction horizon.

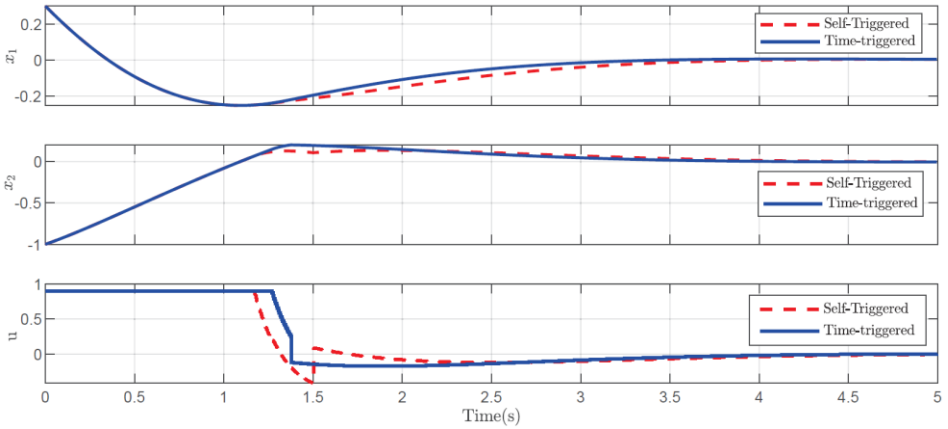


Figure 1: State and input trajectory for (13).

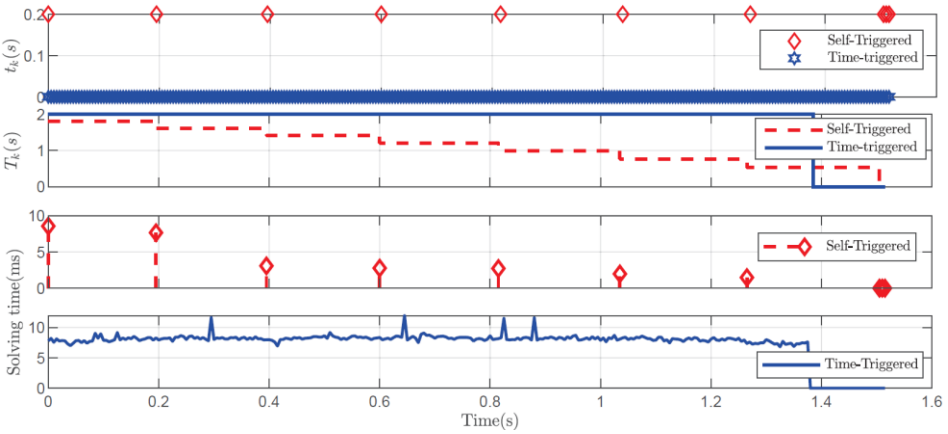


Figure 2: Triggering instants and prediction horizons for (13).

5. Conclusions

In this paper, we proposed a dual-mode MPC with a self-triggered strategy and declining prediction horizons. First, a self-triggered algorithm was proposed. Second, the feasibility and stability in and out of the terminal region were analysed. In addition, within the terminal region, the ultimate boundary of the system states related to perturbs was estimated. Moreover, a sufficient condition is proposed to prevent Zeno behavior. Finally, simulation results were presented to show the effectiveness of the proposed algorithm.

References

Baocang Ding, 2017, Theory and Method of Predictive Control, Machinery Industry Press.

William B Dunbar, 2007, Distributed receding horizon control of dynamically coupled nonlinear systems, IEEE Transactions on Automatic Control, 52, 7, 1249-1263.

Kazumune Hashimoto, Shuichi Adachi, and Dimos V Dimarogonas, 2016, Self-triggered model predictive control for nonlinear input-affine dynamical systems via adaptivecontrol samples selection, IEEE Transactions on Automatic Control, 62, 1, 177- 189.

Huiping Li and Yang Shi, 2014, Event-triggered robust model predictive control of continuous-time nonlinear systems, Automatica, 50, 5, 1507-1513.

David Q Mayne, 2014, Model predictive control: Recent developments and future promise, Automatica, 50, 12, 2967–2986.

David Q Mayne, James B Rawlings, Christopher V Rao, and Pierre OM Scokaert, 2000, Constrained model predictive control: Stability and optimality, Automatica, 36, 6, 789–814.

Hanna Michalska and David Q Mayne, 1993, Robust receding horizon control of con- strained nonlinear systems, IEEE transactions on automatic control, 38, 11, 1623–1633.

Manfred Morari and Jay H Lee, 1999, Model predictive control: past, present and future, Computers & Chemical Engineering, 23, 4-5, 667–682

James Blake Rawlings, David Q Mayne, and Moritz Diehl, 2017, Model predictive control: theory, computation, and design, WI: Nob Hill Publishing.

Zhongqi Sun, Li Dai, Kun Liu, Dimos V Dimarogonas, and Yuanqing Xia,2019, Robustself-triggered mpc with adaptive prediction horizon for perturbed nonlinear sys- tems, IEEE Transactions on Automatic Control, 64, 11, 4780–4787.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50064-6

A comparative study between MPC and selector-based PID control for an industrial heat exchanger

Anikesh Kumar[a], S. Lakshminarayanan[a*], I. A. Karimi[a], Rajagopalan Srinivasan[b]

[a]*Department of Chemical & Biomolecular Engineering, National University of Singapore, Singapore 117585, Singapore*

[b]*Department of Chemical Engineering, Indian Institute of Technology Madras, Chennai 600036, India*

laksh@nus.edu.sg

Abstract

This study compares the performance of two widely used control strategies for industrial systems with constraints, namely MPC and selector-based PID control. The simulation studies are carried out using models built from historical data provided to us for an industrial heat exchanger employed at a South-East Asian processing facility. The comparative studies specifically address the deterioration of performance for the two aforementioned strategies in the presence of varying degrees of valve stiction. Aspects of performance recovery via addition of derivative action and retuning of PI parameters for selector-based control are also discussed. The studies suggest that in addition to being a more formal and robust method to handle constraints, MPC is also able to retain its performance to a larger degree in the presence of stiction as compared to selector-based control.

Keywords: Selector-based PID control; Heat exchanger control; Model predictive control; Valve stiction.

1. Introduction

Heat exchangers are an indispensable part of process industries owing to their major role in the energy efficient operation of plants. Since processes operate in the vicinity of an operating point (or a few such points), a heat exchanger control loop is primarily designed for the regulatory control of the temperatures of relevant exit streams. Proportional-Integral-Derivative (PID) controllers are the most widely deployed controllers in the process industry because of their simplicity and an abundance of pre-existing tuning rules. However, tuning PID controllers for a heat exchanger control loop is not a trivial task when heat exchangers display non-linearity and time changing behavior. Additionally, some processes have inherent economic and safety constraints which impose bounds on the controlled and manipulated variables (CVs and MVs). For instance, direct manipulation of the hot/cold stream may be sufficient for control if it is an utility, however if there are constraints on both streams, bypassing is widely used in the industry (Luyben, 2011). Despite the popularity of PID controllers, one major drawback in their implementation is their inability to innately accommodate the aforementioned constraints. In order to circumvent this problem, many industrial practitioners resort to the usage of ad-hoc measures like selector-based PID control. In a selector-based control framework, a single MV is linked with multiple CVs through different PID controllers. At a given

instant, all these controller outputs are fed to a selector block and the value of the MV is chosen based on a safety logic. Although, selector-based control is widely prevalent, tuning the multiple controllers acting on the same MV is not straighforward and is usually carried out based on the empirical knowledge of plant operators. Furthermore, this approach becomes very cumbersome as the number of variables increase and an attractive alternate is the use of multivariable controllers such as Model Predictive Control (MPC) (Krishnamoorthy & Skogestad, 2020). MPC formally encapsulates the objectives and constraints of a system irrespective of its size (Camacho & Bordons, 2007) and has served as an effective control tool across numerous disciplines over the last few decades (Lee, 2011). Given the contrasting natures of the two aforementioned control strategies, it is interesting to assess their performance on industrial systems with constraints. Additionally, to the best of our knowledge, such a comparative assessment is scarce in the literature. To this end, we have conducted simulation studies comparing the performance and robustness of the two methods on an industrial heat exchanger in our forthcoming work (Kumar et al., 2021). However, it does not consider the effect of control valve faults on the two methodologies. Since, valve stiction is one of the most commonly occurring valve faults (Choudhury et al., 2008), comparative studies specifically addressing the deterioration of performance for the two aforementioned strategies in the presence of varying degrees of valve stiction is considered in this work.

The rest of this paper is organized as follows. Section 2 contains the description of the industrial heat exchanger considered in this study. Section 3 comprises the simulation studies for selector-based control and MPC in the presence of stiction for two different modes of operation of the exchanger. Finally, conclusions and prospective future works are discussed in the last section.

2. Heat exchanger system

The heat exchanger system considered in this study (see Figure 1) is currently employed at a South-East Asian processing facility. The exchanger cools a gas stream to a pre-specified temperature by exchanging heat with a liquid stream. Under nominal operating conditions, the gas outlet temperature can be readily controlled via the manipulation of the valve at the liquid outlet. However, because of the nature of process operation, there is a substantial variation in the gas flow rate throughout the day. This variability leads to the possibility of two-phase flow on the outlets of both streams. At low gas flows, the gas outlet may reach its due point; whereas for high gas flows, the liquid outlet might reach its bubble point. As the system is not designed for two-phase flow, the employed control architecture needs to take these constraints into account. To this end, there is a provision to bypass the gas so as to avoid two-phase flow on either side. Hence, in the bypass mode, the system has three CVs: Gas outlet temperature (CV_1), Liquid outlet temperature (CV_2), and Combined gas outlet temperature (CV_3); two MVs: Liquid valve (MV_1) and Gas main valve[1] (MV_2); and three disturbance variables (DVs): Gas flow rate (DV_1), Gas inlet temperature (DV_2) and Liquid inlet temperature (DV_3).

In order to handle the constraints on the outlet stream temperatures, the currently employed control architecture uses four controllers (TICs 1-4) in unison with two selectors. TIC-1 controls CV_1 through MV_1 under nominal conditions, whereas under the

[1] The main and bypass valves for the gas side are manipulated simultaneously by the same magnitude in opposite directions. Hence, they are together considered as one MV in the current control architecture.

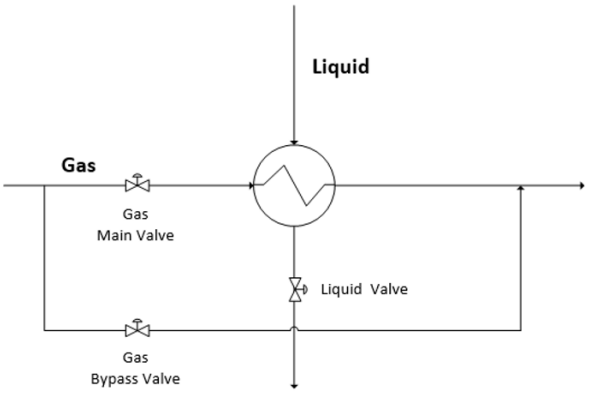


Figure 1: Schematic of the heat exchanger under study

bypass mode, TIC-4 controls CV_3 through MV_2. TIC-2 and TIC-3 both control CV_2 with set-points 10 degrees apart through MV_1 and MV_2 respectively. This accounts for large abrupt changes in gas flow, where bypassing may not be sufficient to prevent liquid overheating and hence MV_1 needs to be adjusted proactively. A max selector on the output of TICs 1 and 2 controls MV_1, whereas a min selector on TICs 3 and 4, controls MV_2. All the aforementioned information is captured in the Simulink model shown in Figure 2.

3. Simulation studies

The simulations conducted in this section are carried out using models identified, PID controller tuning, and MPC parameters used in Kumar et al. (2021). The models used are discrete state-space models (see Eqs. (1-2)) identified using an in-house developed multivariable identification algorithm (Schaper et al., 1994) on the historical plant data available to us[2].

$$X_{t+1} = \Phi X_t + G u_t + w_t \tag{1}$$

$$y_t = H X_t + A u_t + B w_t + v_t \tag{2}$$

For the purpose of this work, we have considered stiction only in the liquid valve (MV_1) as it is the primary manipulated variable and is subject to more wear and tear as opposed to the gas valve (MV_2). Additionally, to quantify stiction, a single parameter stiction model as proposed by Srinivasan et al. (2005) has been used.

$$x(t) = \begin{cases} x(t-1) & if\ |x(t-1) - u(t)| \le d \\ u(t) & otherwise \end{cases} \tag{3}$$

where $x(t)$ and $x(t-1)$ are the present and past MV movements, $u(t)$ is the present controller output (OP), and d is the stiction band which is usually represented as a fraction/percentage of the controller output (OP) range. A non-zero d value suggests the presence of stiction and its severity is directly proportional to the value of d.

[2] Please contact the authors for detailed information on the models used in the simulations.

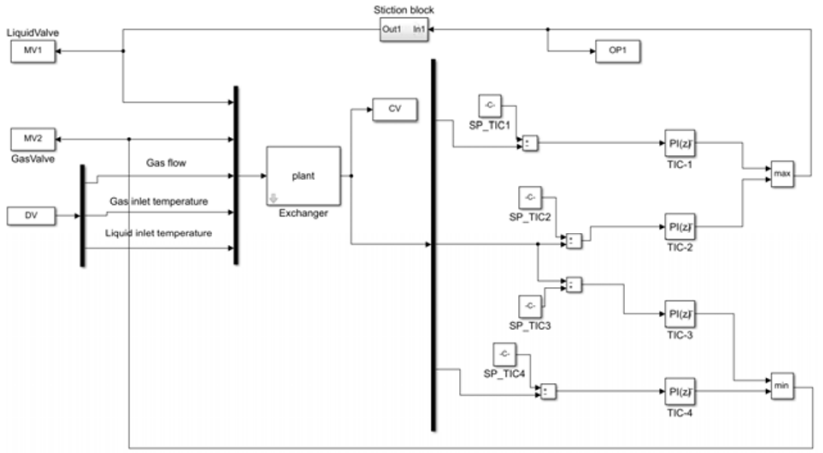


Figure 2: Exchanger model with selector-based control and stiction in liquid valve

3.1. Simulation results for nominal mode

We first consider the nominal mode of operation wherein only MV_1 is active. For disturbance trends observed during a day of operation, we simulate the closed-loop performance of the two control strategies under the presence of 1 and 2 percent stiction. Additionally, since the current controllers in the selector-based strategy are PI and judicious addition of derivative action is recommended as one of the methods of stiction compensation (Patwardhan, 2014), simulation studies with PID control are also included. Figures 3 and 4 show the closed-loop responses[3] for all the control strategies for $d = 1\%$ and $d = 2\%$ respectively. For all the strategies, both the outlet temperatures are at a safe margin from the two-phase regimes, despite the presence of stiction. The integral of squared errors (ISE) for CV_1 using PI, PID, and MPC for $d = 1\%$ and $d = 2\%$ are $[432.62 \;\; 339.02 \;\; 158.07]$ and $[1116.65 \;\; 1082.08 \;\; 699.05]$ respectively. It can be seen that MPC provides the best performance whereas addition of derivative action does lead to improvement in the performance of the selector-based methodology. However, since there is only one MV at the disposal of MPC, the performance gap with respect to selector-based control in the presence of stiction is not as stark as the stiction free case.

3.2. Simulation results for bypass model

Similar to the nominal mode, simulation studies were carried out for the bypass mode for two different levels of stiction in the liquid valve. MPC is able to keep CV_3 close to the set-point with ISE values of 35.18 and 39.43 for $d = 1\%$ and $d = 2\%$ respectively, while satisfying the safety constraints on other variables (see Figure 5). It is also able to keep the secondary variable (MV_2) at the desired opening of 100%. Selector-based control provides inferior performance in terms of squared error and shows a significant deviation from the set-point for CV_3 as opposed to MPC with an ISE value of 240.67 for $d = 1\%$ as evident from Figure 5. For $d = 2\%$, MV_1 saturates leading to two-phase flow on the gas side, and the responses have been omitted for the sake of brevity. It is apparent from the

[3] The numerical values of the controller set-points and stream temperatures have been omitted in all simulation studies for proprietary reasons.

simulation results that MPC can handle stiction without retuning any of the parameters whereas the selector-based methodology can be severely limited in its presence.

4. Conclusions and future work

In this work, we carry out comparative studies between selector-based control and MPC for an industrial heat exchanger with a sticky valve. Simulation studies show that judicious inclusion of derivative action can improve the performance of selector-based control in the presence of stiction; however, MPC is able to preserve its performance better, and one can expect its higher performance to hold, especially for more complex systems. Future work includes inclusion of stiction in the gas valve, exploration of strategies for retuning of PID controllers for multivariable systems using selectors, and optimization of MPC parameters for different levels of stiction.

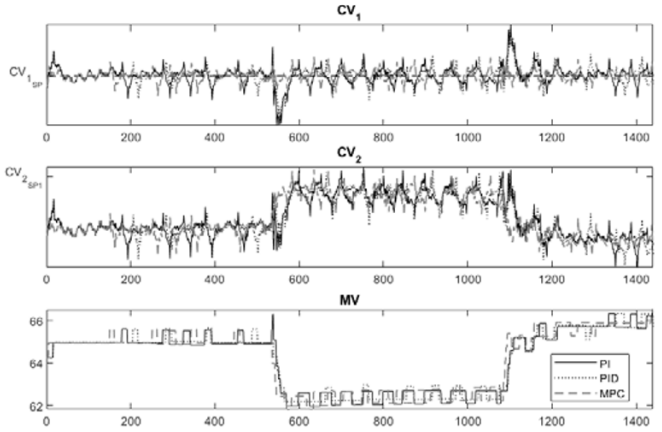


Figure 3: Performance Comparison for Nominal Mode with d=1% for Liquid Valve

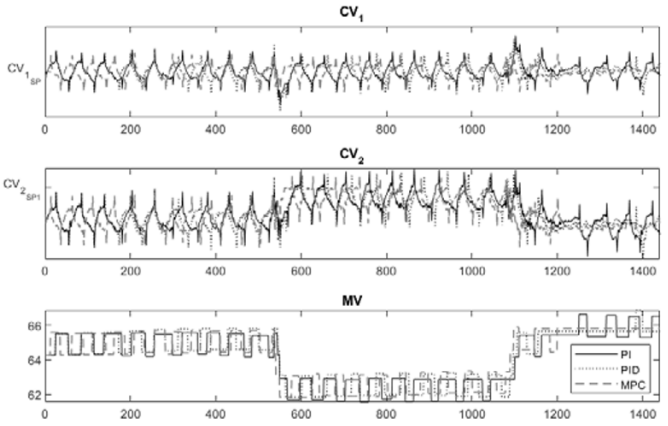


Figure 4: Performance Comparison for Nominal Mode with d=2% for Liquid Valve

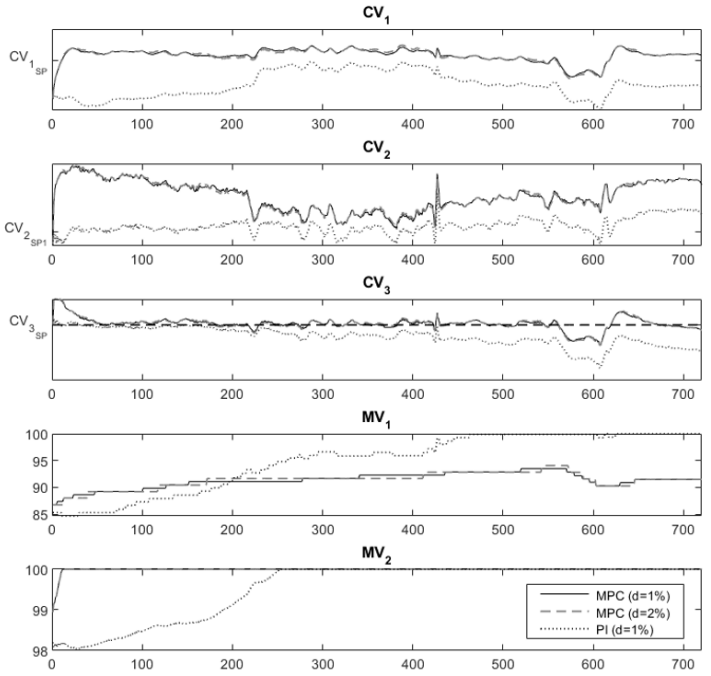


Figure 5: Closed-loop responses for bypass Mode with Sticky Liquid Valve

References

Camacho, E. F., & Bordons, C. (2007). Model predictive control (2nd ed.). London;New York;: Springer.

Choudhury, S. M. A. A., Shah, S. L., & Thornhill, N. F. (2008). Stiction: Definition and Discussions. In S. M. A. A. Choudhury, S. L. Shah & N. F. Thornhill (Eds.), Diagnosis of Process Nonlinearities and Valve Stiction: Data Driven Approaches (pp. 143-151). Berlin, Heidelberg: Springer Berlin Heidelberg.

Krishnamoorthy, D., & Skogestad, S. (2020). Systematic design of active constraint switching using selectors. Computers & chemical engineering, 143.

Kumar, A., Lakshminarayanan, S., Karimi, I. A., & Srinivasan, R. (2021). Critical assessment of control strategies for industrial systems with input-output constraints. Industrial and Engineering Chemistry Research (under review).

Lee, J. H. (2011). Model predictive control: Review of the three decades of development. International Journal of Control, Automation and Systems, 9, 415.

Luyben, W. L. (2011). Heat-Exchanger Bypass Control. Industrial & Engineering Chemistry Research, 50, 965-973.

Patwardhan, R. (2014). Managing the performance of control loops with valve stiction: An Industrial Perspective.

Schaper, C. D., Larimore, W. E., Seborg, D. E., & Mellichamp, D. A. (1994). Identification of chemical processes using canonical variate analysis. Computers & chemical engineering, 18, 55-69.

Srinivasan, R., Rengaswamy, R., Narasimhan, S., & Miller, R. (2005). Control Loop Performance Assessment. 2. Hammerstein Model Approach for Stiction Diagnosis. Industrial & Engineering Chemistry Research, 44, 6719-6728.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50065-8

MILP Formulation for Dynamic Demand Response of Electrolyzers

Florian Joseph Baader[a,b], André Bardow[a,c], Manuel Dahmen[a]*

[a]*Institute of Energy and Climate Research - Energy Systems Engineering (IEK-10), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany*
[b]*RWTH Aachen University, Aachen 52062, Germany*
[c]*Energy & Process Systems Engineering - Department of Mechanical and Process Engineering, ETH Zurich, 8092 Zürich, Switzerland*
**m.dahmen@fz-juelich.de*

Abstract

Electrolyzers can reduce their electricity costs through demand response (DR) by adapting their production rate to time-varying market prices. Although the production rate can often be adapted rapidly, exploiting the full DR potential of an electrolyzer requires to consider slow temperature dynamics, leading to challenging mixed-integer dynamic optimization problems. In this contribution, we propose a dynamic ramping reformulation for real-time scheduling optimization of electrolyzers considering these slow temperature dynamics. Starting from a nonlinear dynamic model, the limits of the temperature gradient are derived to guarantee that the optimization result is feasible on the original model. The limits are then approximated conservatively by piece-wise affine functions leading to a mixed-integer linear program (MILP). Varying the number of piece-wise affine segments allows to explicitly balance model conservativeness against computational burden. We apply our reformulation to a validated alkaline electrolyzer model from literature. Our dynamic temperature ramping approach reduces production costs by 15.9 % compared to nominal operation. A quasi-steady-state optimization, which is restricted to production rates with steady-state temperatures in the allowed range, only leads to 12.8 % improvement. The proposed formulation achieves optimization runtimes below one minute, which is sufficiently fast for real-time scheduling.

Keywords: Electrolysis, Demand response, Mixed-integer linear programming

1. Introduction

Demand response (DR) allows consumers to reduce their electricity costs by adapting production rate to time-varying market-prices and ideally also stabilizes the electricity grid (Zhang and Grossmann, 2016). Particularly suitable for DR are electrochemical production processes such as chlor-alkali or water electrolysis because they can often adapt their production rate rapidly (Burre et al., 2020). More specifically, the time scale of electrochemical reactions is significantly faster than the hourly time scale typical for electricity prices (Simkoff and Baldea, 2020). Therefore, step changes of the production rate can often be assumed in electrolyzer scheduling optimization. Scheduling usually considers a time horizon in the order of one day. However, electrolyzers feature slow temperature dynamics in the order of hours (Gabrielli et al., 2016). Neglecting these slow temperature dynamics during scheduling, reduces the flexibility of the electrolyzer to steady-state-feasible production points that are limited by the minimum and maximum allowed temperature (Simkoff and Baldea, 2020). From a technical point of view,

production rates outside of the steady-state-feasible range can be applied in transient operation for a limited amount of time (Flamm et al., 2021). For example, the energy park Mainz (Germany) has a PEM electrolyzer with 4 MW nominal power that can operate at a peak power of 6 MW for 15 minutes if load is reduced afterwards to allow cooling (Kopp et al., 2017). Simkoff and Baldea (2020) consider temperature effects of a chlor-alkali electrolyzer using dynamic optimization but replace the original nonlinear model with a data-driven surrogate model. On top of dynamic effects, electrolyzers have a minimum allowed current (Ulleberg, 1998). Consequently, the possibility to turn off the electrolyzer can only be considered in scheduling if discrete decisions are included, leading to computationally challenging mixed-integer dynamic optimization (MIDO) problems.

In this contribution, we propose a mixed-integer linear programming (MILP) formulation for electrolyzer scheduling optimization. To this end, we consider temperature dynamics using a dynamic ramping approach (Baader et al., 2021). Accordingly, the limits of the temperature gradient are functions of the electrolyzer state. In contrast to data-driven surrogate models, our reformulation guarantees results that are feasible on the original nonlinear model. The dynamic temperature ramping reformulation is introduced in Section 2. In Section 3, we apply the reformulation to an alkaline electrolyzer model (Ulleberg, 1998) and present results in Section 4. Section 5 concludes the work.

2. MILP Formulation for Electrolyzer Scheduling

2.1. Assumptions

Our reformulation is not restricted to one specific electrolyzer model. The reformulation only relies on a few assumptions that are typically satisfied. The main assumption is that the temperature T is the only differential state. Accordingly, all other states can be considered in (quasi-)steady-state on the scheduling-relevant hourly timescale, which is often valid as electrochemical reactions typically occur on a much faster time scale. Further, we assume that there are two degrees of freedom in electrolyzer operation: the current I and an input u_{cool} acting on the cooling power. For instance, u_{cool} can be the temperature of raw material (Simkoff and Baldea, 2020) or the flow rate of a cooling fluid (Ulleberg, 1998). Thus, the scheduling-relevant variables product flow rate n_{prod}, electric power P_{elec}, and temperature gradient $\frac{dT}{dt}$ can be calculated as nonlinear functions of I, u_{cool}, and T (Figure 1, left). For instance, in a simple model, the voltage U is modeled as an empiric function of I and T (Ulleberg, 1998). The voltage U is needed to calculate both electric power and temperature gradient. In more physically motivated models, the voltage also depends on partial pressures, which are determined by the mass balances (Gabrielli et al., 2016). As these mass balances can be assumed to be in steady state on the hourly time scale, the balances give a nonlinear system of algebraic equations that can be solved for given I, u_{cool}, and T. Thus, our assumption is still valid. Moreover, we assume that the nonlinear functions can be inverted for a given temperature T. More specifically, the input I can be calculated as a function $I = \phi_I(T, n_{prod})$ of temperature T and output n_{prod}, and the input u_{cool} can be calculated as a function $u_{cool} = \phi_{u_{cool}}(T, n_{prod}, \frac{dT}{dt})$, which additionally to T and n_{prod} depends on the output $\frac{dT}{dt}$.

Finally, the variables n_{prod}, u_{cool}, T are bounded by minimum and maximum values. We assume that step changes can be applied on the scheduling time scale to the current I (Flamm et al., 2021). Thus, ramping constraints on the current I are neglected, but can be added in a straightforward way if necessary.

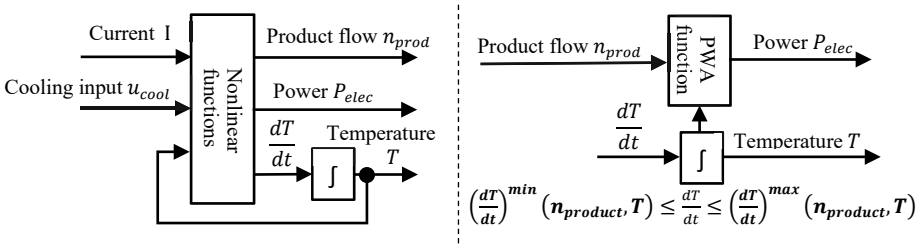


Figure 1: Original nonlinear model (left) and reformulation with piece-wise affine (PWA) function and dynamic ramping constraint in bold font (right).

2.2. Reformulation

In our reformulation, we directly use the scheduling-relevant variables product flow n_{prod}, and temperature gradient $\frac{dT}{dt}$ as degrees of freedom and do not model I and u_{cool} explicitly, which is possible as they can be calculated from n_{prod} and $\frac{dT}{dt}$. The bounds of the temperature gradient depend on both temperature T and product flow rate n_{prod} (Figure 1, right). To choose these bounds, we sample the two-dimensional space given by the bounds of n_{prod} and T. For every pair (n_{prod}, T), we first calculate the current I from the function $\phi_I(T, n_{prod})$ introduced above. Second, we calculate the true nonlinear ramping limits $\left(\frac{dT}{dt}\right)^{min}$, $\left(\frac{dT}{dt}\right)^{max}$ by inserting u_{cool}^{min}, u_{cool}^{max} into the right-hand side function of $\frac{dT}{dt}$. Subsequently, we approximate the nonlinear ramping limits conservatively by piece-wise affine (PWA) functions. Because of this conservativeness, the resulting temperature profile is guaranteed to be feasible on the original nonlinear model. In principle, choosing the conservative limits can be done by bivariate regression (Adeniran and El Ferik, 2017). However, in our case study, we observe that the true nonlinear limits have an almost linear dependence on the electrolyzer temperature T. This observation is likely transferable to other cases because the ramping limits are mainly temperature dependent due to the heat loss to the ambient. This heat loss is essentially proportional to the temperature difference between electrolyzer temperature T and ambient temperature. However, the ramping limits are nonlinear in the production rate n_{prod}. Thus, we set up piece-wise affine functions by dividing the range of n_{prod} into segments. In every segment, the affine functions are parametrized through an optimization that minimizes the distance to the nonlinear bounds. By including more PWA segments, the ramping limits come closer to the true nonlinear limits. However, also the number of binary variables increases.

Finally, we follow the established approach to approximate the electric power P_{elec} as affine function of n_{prod} and T using linear regression (Flamm et al., 2021). Here, small approximation errors are acceptable as they can be compensated by adapting grid electricity consumption. An MILP formulation is achieved by discretizing the temperature evolution using orthogonal collocation on finite elements (Biegler, 2010).

3. Case Study

As case study, we apply our reformulation to a validated alkaline electrolyzer model (Ulleberg, 1998). This model uses empirical functions for the voltage and hydrogen production depending on current I and temperature T. The input u_{cool} is the cooling flow rate, and the temperature gradient $\frac{dT}{dt}$ can be computed as function of I, T, and u_{cool}. Thus, the model satisfies our assumptions (compare to Subsection 2.1 and left part of Figure 1).

The operating range is given by minimum and maximum current density $i^{min} = 40\ \frac{\text{mA}}{\text{cm}^2}$ and $i^{max} = 300\ \frac{\text{mA}}{\text{cm}^2}$ as well as minimum and maximum temperature T^{min} = 50 °C and T^{max} = 80 °C (Ulleberg, 2003).

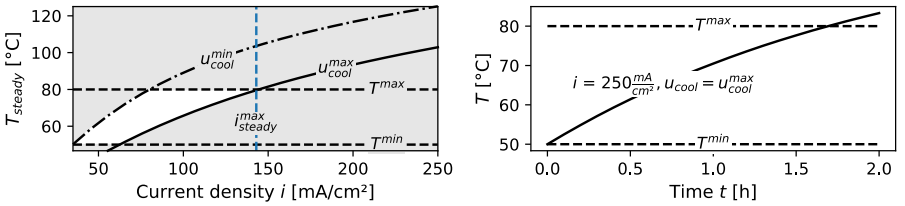


Figure 2: Left: Minimum and maximum temperature, T^{min}, and T^{max}, maximum steady-state-feasible current i_{steady}^{max}, and steady-state temperature T_{steady} for minimum and maximum cooling fluid rate u_{cool}^{min}, and u_{cool}^{max}. Right: Evolution of temperature T for fixed current density i and u_{cool}^{max}.

The steady-state-feasible operating region is given by $i^{min} = 40\ \frac{\text{mA}}{\text{cm}^2}$ and the maximum steady-state-feasible current density $i_{steady}^{max} = 144\ \frac{mA}{cm^2}$ (Figure 2, left). However, significantly higher currents are possible for a scheduling-relevant time. Exemplarily, if the temperature starts from $T^{min} = 50$ °C, a current density of $250\ \frac{\text{mA}}{\text{cm}^2}$ can be applied for more than 1.5 hours until the maximum temperature is reached (Figure 2, right).
In order to describe the temperature dynamics by PWA temperature ramping limits $\left(\frac{dT}{dt}\right)^{min}$, $\left(\frac{dT}{dt}\right)^{max}$, we divide the range of the production rate n_{prod} into equidistant segments, as described in Section 2. We observe that the nonlinearities with respect to n_{prod} are so strong that for less than three affine segments the minimum and maximum ramping limits overlap. We vary the number of affine segments n_{seg} between 3 and 10 and choose 5 segments, which give a reasonable approximation (Figure 3).

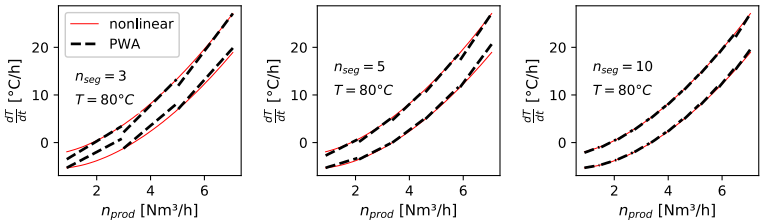


Figure 3: Nonlinear and piece-wise affine (PWA) limits of temperature gradient $\frac{dT}{dt}$ as function of product flow n_{prod} for temperature $T = 80$ °C and different numbers of segments n_{seg}.

In our numerical study, we choose a nominal production rate of $n_{prod}^{nom} = 2.69\ \frac{\text{Nm}^3}{\text{h}}$, which equals 80 % of the production rate achieved with the maximum steady-state feasible current density i_{steady}^{max}. Furthermore, we assume that the nominal production rate must be met on average over the one-day time horizon and use a recent German day-ahead market electricity price profile from April 2nd, 2021. We study the economic performance of our scheduling optimization in a simulation with the original nonlinear process model. This strategy allows us to check the suitability of the chosen time discretization and to verify that the cooling flow rate u_{cool} and the temperature T always stay within the respective

bounds. We benchmark our dynamic temperature ramping approach against (i) a nominal operation with constant production rate n_{prod}^{nom} and (ii) a quasi-steady-state (QSS) optimization that does not consider temperature dynamics and thus can only operate within the steady-state-feasible range. For this QSS optimization, we calculate the efficiency curve assuming that the electrolyzer is at the maximum steady-state-feasible temperature for every current (compare to Figure 2) because the efficiency increases with temperature. To this end, we set the cooling input u_{cool} to zero in the simulation when the temperature is below the maximum allowable temperature and otherwise select u_{cool} such that the temperature stays constant. All optimization problems are solved using gurobi 8.1.0 on an Intel Core i5-8250U processor with an optimality gap of 1 %. Only for the QSS benchmark, zero optimality gap is used such that our dynamic ramping approach is benchmarked against the optimal QSS schedule.

4. Results

Compared to nominal operation at constant production rate n_{prod}^{nom}, QSS optimization based on the steady-state-feasible region reduces electricity costs by 12.8 %. In contrast, our dynamic temperature ramping approach with 5 piece-wise affine segments achieves 15.9 % cost reduction. The optimization runtime is 32 seconds. Using 3 and 10 PWA segments, we achieve 13.6 % in 16 s and 16.5 % in 165 s, respectively. Consequently, dynamic optimization increases savings by up to 29 % compared to QSS.

Figure 4 shows the resulting operation for QSS optimization and for the dynamic temperature ramping with 5 segments. The dynamic temperature ramping approach exploits the fact that the electrolyzer can be cooled down, while it operates at low powers. Afterwards, production rates and electric powers can increase above the steady-state feasible point (see hours 2, 13, 14, and 22 in Figure 4). Interestingly, QSS optimization turns off the electrolyzer for 4 hours of high prices, while our dynamic ramping approach keeps the electrolyzer active for the complete 24 hours. The reason is that for the studied electrolyzer the heat transfer coefficient of the internal heat exchanger increases with the current *I* (Ulleberg, 1998). Consequently, when the electrolyzer is active it can be cooled down faster and thus deeper, which allows to operate at higher powers later in hours 13 and 14. Even if the lowest electricity price occurs at hour 15, the hours 13 and 14 show the highest input powers. The intuitive decision to schedule the highest power in hour 15 is not optimal as the efficiency decreases at lower temperatures. Thus, waiting with the temperature ramp-up until hour 15 would lead to one more hour of operating at low efficiency. These complex temperature dynamics explain why our dynamic approach outperforms the quasi-steady-state benchmark.

5. Conclusion

Electrolyzers are promising demand response (DR) candidates; however, realizing their full DR potential requires challenging mixed-integer dynamic optimization. We propose dynamic temperature ramping which allows to reformulate the nonlinear dynamic model into a mixed-integer linear model. By conservatively approximating the limits of the temperature gradient, the resulting trajectory is guaranteed to be feasible on the original nonlinear model. Our case study considers a validated alkaline electrolyzer model and shows that dynamic temperature ramping reduces costs by 15.9 % compared to nominal operation. A steady-state optimization, which is limited to operate within the steady-state feasible region, only achieves 12.8 % cost reduction. Moreover, our approach allows to explicitly balance computational complexity against solution quality and thereby reaches optimization runtimes below one minute. We expect our approach to be transferable to

many other applications because our main assumption that the temperature evolution is the only dynamic relevant on an hourly timescale is typically true for electrolyzers.

Acknowledgements

The present contribution is supported by the Helmholtz Association under the Joint Initiative 'Energy Systems Integration'.

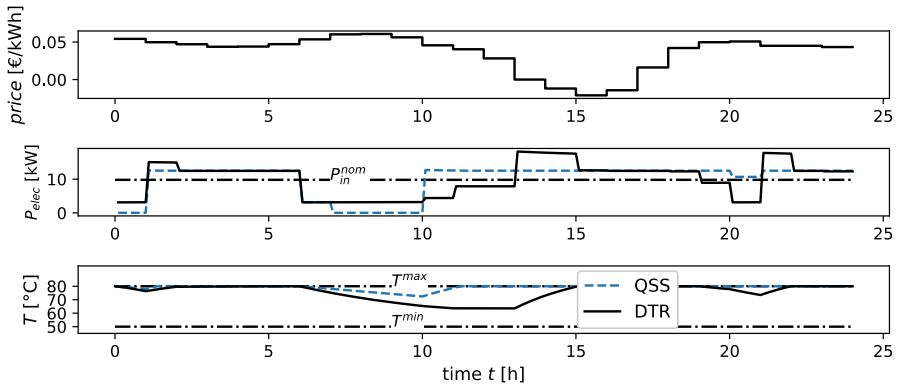


Figure 4: Electricity price (top), simulated electric power P_{elec} (middle), and simulated temperature T (bottom) for quasi-steady-state (QSS) and dynamic temperature ramping (DTR) with $n_{seg} = 5$ segments. The nominal input power P_{in}^{nom} is shown for comparison.

References

A. A. Adeniran, S. El Ferik, 2017, Modeling and identification of nonlinear systems: A review of the multimodel approach—Part 1, IEEE Transactions on Systems, Man, and Cybernetics: Systems, 47 (7), 1149–1159

F. J. Baader, P. Althaus, A. Bardow, M. Dahmen, 2021, Dynamic ramping for demand response of processes and energy systems based on exact linearization, arXiv:2110.08137v1

L. T. Biegler, 2010, Nonlinear programming, SIAM, Philadelphia

J. Burre, D. Bongartz, L. Brée, K. Roh, A. Mitsos, 2020, Power-to-X: Between electricity storage, e-production, and demand side management, Chemie Ingenieur Technik, 92, 74–84

B. Flamm, C. Peter, F. N. Büchi, J. Lygeros, 2021, Electrolyzer modeling and real-time control for optimized production of hydrogen gas, Applied Energy, 281, 116031

P. Gabrielli, B. Flamm, A. Eichler, M. Gazzani, J. Lygeros, M. Mazzotti, 2016, Modeling for optimal operation of PEM fuel cells and electrolyzers, 2016 IEEE 16th International Conference on Environment and Electrical Engineering, 1–7

M. Kopp, D. Coleman, C. Stiller, K. Scheffer, J. Aichinger, B. Scheppat, 2017, Energiepark Mainz: Technical and economic analysis of the worldwide largest power-to-gas plant with PEM electrolysis, International Journal of Hydrogen Energy, 42 (19), 13311–13320

J. M. Simkoff, M. Baldea, 2020, Stochastic scheduling and control using data-driven nonlinear dynamic models: Application to demand response operation of a chlor-alkali plant, Industrial & Engineering Chemistry Research, 21, 10031-10042

Ø. Ulleberg, 1998, Stand-alone power systems for the future: Optimal design, operation and control of solar-hydrogen energy systems, PhD-thesis, Trondheim

Ø. Ulleberg, 2003, Modeling of advanced alkaline electrolyzers: A system simulation approach, International Journal of Hydrogen Energy, 1, 21-33

Q. Zhang, I.E. Grossmann, 2016, Planning and scheduling for industrial demand side management: Advances and challenges, Alternative Energy Sources and Technologies, Ed. M. Mariano, Springer International Publishing, 383-414

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50066-X

Real-Time Optimal Operation of a Chlor-Alkali Electrolysis Process under Demand Response

Erik Esche[a*], Joris Weigert[a], Christian Hoffmann[a], Jens-Uwe Repke[a]

[a]*Process Dynamics and Operations Group, Technische Universität Berlin, D-10623 Berlin, Germany*
erik.esche@tu-berlin.de

Abstract

Real-time implementation of nonlinear model-predictive control (NMPC) for systems under demand response remains a challenge. Deep recurrent neural networks may serve as approximators for online application. Using hyperparameter tuning through Bayesian and Bandit optimization, deep neural networks are trained to high accuracy regarding testing data. An NMPC applied on an industrial chlor-alkali electrolysis example with a reactive distillation section is replaced by deep neural nets. The resulting approximation shows a perfect match to the offline NMPC using a Jordan RNN with 1 or 2 hidden layers, which surpasses the performance of LSTMs.

Keywords: demand response, optimal process operation, chlor-alkali electrolysis, neural network.

1. Introduction & Motivation

Given the surge in renewable energy into electricity markets, heavy fluctuations of electricity prices can be observed leading to both spikes as well as infrequent negative prices. To balance fluctuating production and demand, demand response has become an important tool and opened a market worth billions of euros worldwide, for example, in Germany. Within the European Union, demand response involves load shedding or increases within seconds (FCR), minutes (aFRR), or up to quarter hours (mFRR). Chemical plants with direct electricity input, e.g., air separation units or electrochemical processes, can profit from this market. More recently, investigation of the practical realization of demand response in operations has started. Recurring issues are the sizable and fast load drops or increases with no forewarning: For FCR and aFRR, load changes are directly implemented by the grid operator. Given that these load changes might involve complete shutdown or restart of an entire plant, this is a challenge for standard control solutions. It is unlikely that a nonlinear model predictive controller (NMPC) with a full mechanistic model can be solved in real-time. For speed-up, Vaupel et al. (2020) proposed two different approaches to train artificial neural networks (NN) to (1) serve as initial guess for NMPC or (2) as basis for a control update by quadratic programming. On the other hand, Karg and Lucia (2019) employed deep learning to learn entire robust NMPCs by deep NN.

For learning nonlinear dynamic relationships, a wide range of recurrent neural networks (RNN) exist. RNN feature internal feedback, i.e., outputs are passed as inputs for the next iteration / time step. They exist in various forms, from “fully connected” to leaner types, such as Elman RNN and Jordan RNN (Jordan, 1997). Long short-term memory (LSTM) units (Hochreiter and Schmidhuber, 1997) were developed to learn long-term effects. However, a recent contribution by Gonzalez and Wen (2018) noted that for some basic

nonlinear systems LSTM-based RNNs do not work well and combination with more basic RNN types are necessary for satisfactory results. At this point, RNNs are applied on chemical engineering examples without specific adjustment. It is unclear, which types of RNN are sufficient for which types of systems in chemical engineering.
The present contribution will evaluate deep NNs as approximators for nonlinear model predictive control applied on continuously operated chemical processes under demand response. To this end, Section 2 describes the methodology that we pursue to obtain a representative RNN approximator of the NMPC, before Section 3 discusses the chloralkali electrolysis process and a subsequent reactive distillation as case study.

2. Methodology

For our methodology, we will assume that a suitable dynamic process model exists, which mimics the process behavior throughout the demand response scope. Also, a sampling is available with realistic scenarios given the process dynamics and the energy market. This entails, e.g., scheduling based on fluctuations in the electricity price.

2.1. Optimal Process Control

With such scheduling results and a dynamic process model, optimal control actions may be obtained. We assume a basic discrete-time implementation of an NMPC with a finite prediction horizon N focusing on tracking control, which is solved at sample time *l*:

$$\min_{u_k} \sum_{\nabla k=l\dots l+N} \Phi(x_{k\dots k+N}, u_{k\dots k+N}) \tag{1}$$

$$x_{k+1} := g(x_k, u_k, d_k) \quad \nabla k = l \dots l+N-1, \tag{2}$$

$$0 \le h(x_{k+1}, u_k, d_k) \qquad \nabla k = l \dots l+N-1, \tag{3}$$

$$u^L \le u_k \le u^U \qquad \nabla k = l \dots l+N-1, \tag{4}$$

with states x_k, control inputs u_k, disturbances d_k, dynamic process model g, and inequality constraints h as well as bounds for the control inputs u^L, u^U. Based on the scenarios of step 2, the NMPC formulation may be solved offline, which yields optimal control input u^* for current process state x and disturbances d, which contain the load changes required by the electricity market.

2.2. Approximate Control Law

Using these optimal control inputs u^*, a neural net may be trained as a direct approximator for the nonlinear optimal controller:

$$u^*_{k+} \approx \tilde{u}_{k+} = f(x_{k-}, u_{k-}, d_{k-}), \tag{5}$$

wherein f is a neural network mapping from current state estimate to approximate optimal control input. $k+$ denotes time points in the future, while k- is current and possibly past information. We assume that a good estimate of the current process is always available. A variety of structural choices are at hand for f. Here, we shall limit ourselves to RNNs of type Jordan and LSTM. For these, the number of neurons per layer, the number of hidden layers overall, and regularization parameters need to be selected. Further options concern the number of additional input variables, e.g., the number of past control inputs u to be considered and for Jordan RNNs the number of past states held internally. The former goes well beyond a standard NMPC application and might allow to also include the state estimation step as part of the controller. Choosing these hyperparameters to

obtain a suitable NN is not trivial for general nonlinear systems. This issue can be resolved by hyperparameter optimization: Continuous decisions may be made by Bayesian optimization (Frazier, 2018) and discrete decisions based on Bandit optimization (Dimmery et al., 2019). Both are employed through the python framework Ax (https://ax.dev). During the hyperparameter optimization, all decisions mentioned above are made – apart from the selection of RNN type and number of hidden layers (1 up to 4), which are varied manually. The hyperparameter optimization uses the mean squared error (MSE) regarding the testing data as objective and runs for 50 iterations. Ranges for the hyperparameters are based on prior experience to values set in Table 1. The Jordan-type RNN is constructed in scikit-learn (Pedregosa, 2011), while Keras' LSTM is used as is (https://keras.io). The time series obtained from step 2 is split into training (80 %) and testing (20 %) data set. For scaling of inputs StandardScaler of scikit-learn and MinMaxScaler for outputs is used. The results of the NMPC are rearranged into tuples of input-output pairs with varying size depending on the hyperparameters:

$$\text{Input: } \left(x_{j-l} \dots x_j, d_{j-l} \dots d_j, u_{j-m} \dots u_{j-1}\right), \text{Output: } (\tilde{u}_j \dots \tilde{u}_{j+N}), \tag{6}$$

with m number of past controls, l number of past states, for current time point j.
To train the neural net f, the MSE of the training data between u^* and predicted $\hat{u}$ is minimized. This is amended with a weighted bias penalty term (L2 norm) for regularization. Adam (Kingma and Ba, 2014) is used as solver for training with a fixed batch size of 200 and early stopping with a tolerance of 1.0e-6 and patience of 10.

Table 1. Ranges for hyperparameters chosen during the hyperparameter optimization.

Hyperparameter	Lower Bound	Upper Bound
Number of past controls m	1	50
Number of past states (Jordan) l	10	50
Number of neurons per hidden layer	10	300
Type of activation function	ReLU or tanh	
Regularization parameter (L2)	1.0e-5	10.0

3. Case Study

The operation under demand response of a chlor-alkali electrolysis (CAE) and a subsequent synthesis of 1,2-dichloroethane (DCE) from chlorine and ethene is investigated here (see Figure 1). While storage of chlorine is limited due to safety restrictions, DCE can easily be stored in large quantities. The application of demand response on the CAE causes a fluctuation in both the electrolysis as well as the reactive distillation section producing DCE. Particularly, operation of the DCE under heavy fluctuations of the chlorine stream from the CAE is a challenge. The combined reaction and distillation section of the DCE is modeled with a typical tray-based formulation with a special focus on dynamic load changes (Hoffmann et al., 2020).

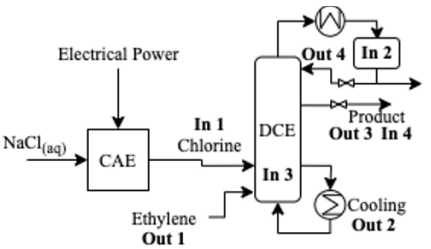


Figure 1. Simplified flowsheet of case study.

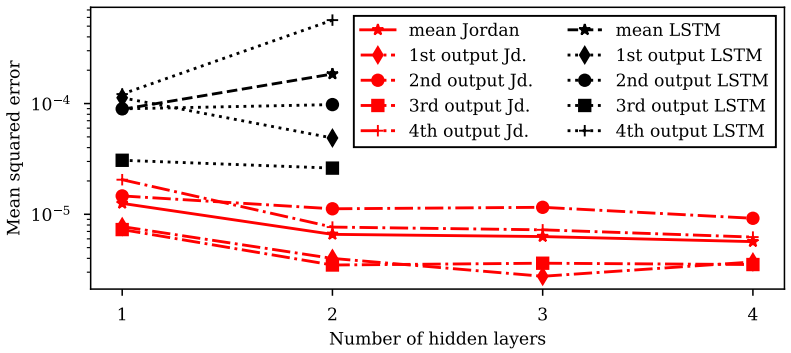


Figure 2. Mean squared error (MSE) of the testing data per output and overall, for Jordan RNN (Jd.) and LSTM with 1 up to 4 hidden layers.

Hoffmann et al. (2021) developed a framework to obtain realistic load profiles for the CAE-DCE process under demand response. A 200-hour profile computed with their framework will serve as a basis for the subsequent case study here. The profile contains load changes of up to 30 % due to either aFRR or mFRR.

3.1. Optimal Process Control

The NMPC formulation described by Eq. (1)-(4) contains an objective ensuring the adherence to the required DCE purity in the outlet of the reactive distillation section. The model consists of the complete dynamic reactive distillation as described in (Hoffmann et al., 2020). As inequality constraints strict ramp restrictions on the changes of the control variables are implemented to ensure technical feasibility. A more detailed discussion may be found in (Hoffmann, 2021). The prediction horizon N is set to 45 minutes as the slow mFRR contains ramps lasting 15 minutes. By solving the NMPC problem with the electricity market profile as disturbance d, optimal control inputs u^* are obtained.

3.2. Approximate Control Law

With the NMPC results, a set of neural nets is trained with 1 to 4 hidden layers of type Jordan and LSTM. Here, we focus on the main inputs of interest for the control of the DCE: The chlorine stream stemming from the CAE represents the market signal and is regarded as an external disturbance (In 1). The state of the DCE is overall captured by the current level in the reflux drum (In 2), the liquid level in the column bottom (In 3), and the concentration of DCE in the product stream (In 4).

Table 2. Training results of the neural nets with two hidden layers of type Jordan and LSTM.

Hyper-parameters	Jordan RNN (2 hidden layers)				LSTM (2 hidden layers)			
	Out 1	Out 2	Out 3	Out 4	Out 1	Out 2	Out 3	Out 4
m	46	47	39	43	46	43	49	37
l (Jordan)	27	7	10	44	-	-	-	-
Neurons	53/115	26/126	128/102	23/257	30/115	86/64	147/88	18/193
Activation	ReLU	tanh	ReLU	ReLU	ReLU	tanh	ReLU	ReLU
Reg. (L2)	5.3e-3	2.5e-3	1.1e-2	2.0e-3	3.7e-5	5.8e-5	3.7e-5	2.1e-4

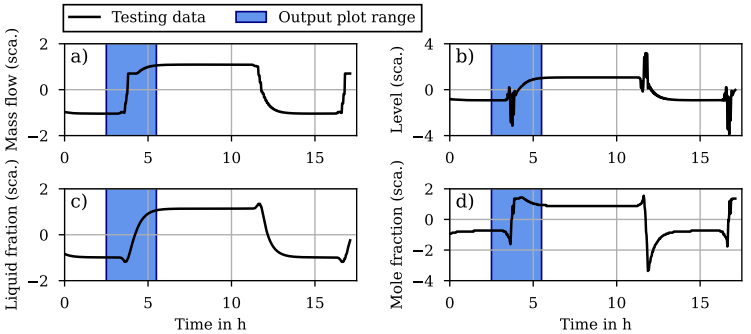


Figure 3. Scaled profiles of the four input variables in the testing data.

The most important control actions on the process are the ethylene dosing (Out 1), the auxiliary column cooling (Out 2) to remove some of the heat of reaction at the bottom, the product flow (Out 3), and the reflux from the reflux drum (Out 4). Results of hyperparameter optimization and training are shown in Figure 2, which states the MSE regarding the testing data per output. For the Jordan RNN with a single layer, the error per output is small and an order of magnitude smaller than for the LSTM. With 2 or more hidden layers, the error decreases further, although the improvements become negligible. Given the increase of the MSE for 2 hidden layers, the further results for LSTM are omitted here. In Table 2, further details are given on Jordan RNN and LSTM with two hidden layers: All hyperparameters are chosen within the specified ranges. Strong variations can be observed with respect to the specific regularization for the various outlets here. Similarly, the choice in activation function appears to lean towards ReLU, although output 2 favors tanh. The number of neurons here is high, but common compared with available publications.

3.3. Quality of the Approximation

To evaluate the quality of the prediction, the testing data will be further examined. Figure 3 shows the testing data for the four input variables named above. Observe that the plots show an interval of 17 hours and the profiles in the minute range are smooth. Figure 4 focusses at a two-hour interval marked in Figure 3 and shows the results for Jordan RNN and LSTM with one and two hidden layers respectively. All four trained models can follow the trajectory computed by the NMPC.

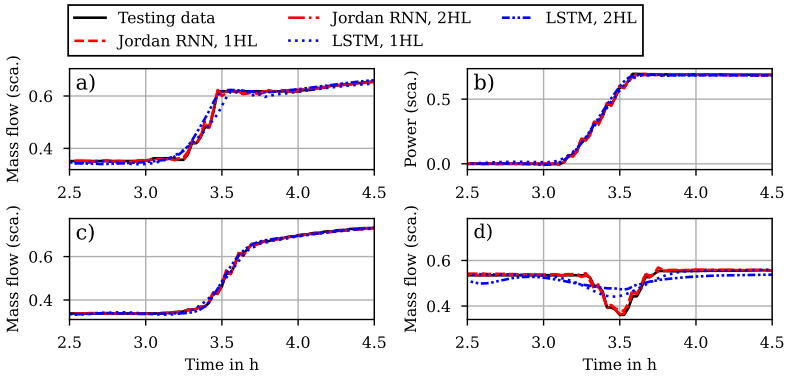


Figure 4. Scaled profiles of the four output variables. Close-up of a two hour interval of the testing data with the results of Jordan RNN and LSTM with one and two hidden layers (HL).

However, the Jordan RNNs manage to follow the slightest control actions of the NMPC, so much so that the testing data disappears behind the plot of the Jordan RNN with two hidden layers. Overall, using the Jordan RNN instead of the original NMPC would here lead to almost no approximation error and the suboptimality is nearly unquantifiable with respect to the solution accuracy of the NMPC itself.

4. Conclusions & Outlook

Direct application of NMPC particularly on fast and highly nonlinear systems remains a challenge. Using offline computation based on realistic demand response scenarios, optimal control trajectories can be computed offline and learned to high accuracy by neural nets. Particularly, Jordan RNNs with two or more hidden layers show a great adaptability in this regard. Using hyperparameter optimization during training greatly eases the effort to obtain suitable NNs. It appears that the learned approximate control law may be applied in real-time to replace slow NMPCs.
Several issues remain to be solved in future work, ensuring the reliability of neural nets beyond the operation window they were originally trained on is always a point of contention. More advanced techniques beyond more extensive sampling are needed, which ensure feasibility of the approximated NMPC results throughout.

5. Acknowledgements

The authors acknowledge the financial support by the Federal Ministry of Economic Affairs and Energy of Germany in the project ChemEFlex (project number 0350013A).

References

D. Dimmery, E. Bakshy, J. Sekhon, 2019, Shrinkage Estimators in Online Experiments, arXiv: 1904.12918v1

P.I. Frazier, 2018, A Tutorial on Bayesian Optimization, arXiv: 1807.02811v1

J. Gonzalez, Y. Wen, 2018, Non-linear System Modeling Using LSTM Neural Networks, IFAC-PaperOnLine, 51, 13, 485-489

S. Hochreiter, J. Schmidhuber, 1997, Long Short-term Memory, Neural Computation, 9, 8, 1735-1780

C. Hoffmann, J. Weigert, E. Esche, J.-U. Repke, 2020, Towards demand-side management of the chlor-alkali electrolysis: dynamic, pressure-driven modeling and model validation of the 1,2-dichloroethane synthesis. Chemical Engineering Science, 214, 115358.

M. Hofmann, R. Müller, A. Christides, P. Fischer, F. Klaucke, S. Vomberg, G. Tsatsaronis, 2021, Flexible and economical operation of chlor-alkali process with subsequent polyvinyl chloride production, AiChE Journal, DOI: https://doi.org/10.1002/aic.17480

M.I. Jordan, 1997, Serial Order: A Parallel Distributed Processing Approach, Neural-Network Models of Cognition, 121, 471-495

B. Karg and S. Lucia, 2019, Learning-based approximation of robust nonlinear predictive control with state estimation applied to a towing kite, 18th European Control Conference (ECC), 16-22, doi: 10.23919/ECC.2019.8796201

D.P. Kingma, J. Ba, 2014, Adam: A Method for Stochastic Optimization, arXiv: 1412.6980v9

F. Pedregoas, G. Varoquaux, A. Gramfor, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, 2011, Scikit-learn: Machine Learning in Python, Journal of Machine Learning, 12, 85, 2825-2830

Y. Vaupel, N.C. Hamacher, A. Caspari, A. Mhamdi, I.G. Kevrekidis, A. Mitsos, 2020, Accelerating nonlinear model predictive control through machine learning, Journal of Process Control, 92, 261-270

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50067-1

Explicit Multi-Objective and Hierarchical Model Predictive Control

Styliani Avraamidou[a*], Iosif Pappas[b,c], Efstratios N. Pistikopoulos[b,c]

[a]*Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Engineering Drive, Madison, WI 53703, United States*
[b]*Texas A&M Energy Institute, Texas A&M University, Address, College Station, TX 77800, United States*
[b]*Artie Mc Ferin Department of Chemical Engineering, Texas A&M University, Address, College Station, TX 77800, United States*
avraaamidou@wisc.edu

Abstract

Model predictive control (MPC) problems can involve multiple, often conflicting objectives, including economic performance, tracking accuracy, disturbance rejection, safety, or environmental criteria. Each of these objectives can be used to design different MPCs that will have different input trajectories and consequently different operational behaviours in closed-loop operation. Various approaches have been proposed in the open literature for the development of multi-objective model predictive controllers in an effort to combine some of the objectives. Multiple objectives can also be ranked in a hierarchy, where every control level in the hierarchy is controlling a subset of the overall control variables, by manipulating a subset of the overall control variables, resulting in hierarchical model predictive controllers. This work utilizes multi-parametric programming to generate both multi-objective and hierarchical explicit model predictive controllers. A case study on a combined stirred tank reactor with two competing objectives, an economic and a tracking objective, is used to illustrate the developed control strategies. The results of this study clearly indicate the effect of the different control strategies on the operation of the reactor.

Keywords: Model Predictive Control; Multi-Parametric Programming; Multi-level Optimization; Multi-Objective Optimization; Explicit Control.

1. Main Text

The optimal operation of a system can involve multiple, often conflicting objectives, due to the complexity of the problems that are studied. These objectives that need to be optimized include economic performance, safety, or environmental criteria. In the context of operational optimization through process control, various approaches have been proposed in the open literature in an attempt to incorporate these objectives simultaneously in the development of optimal control policies. One approach is multi-objective model predictive control (MOMPC), where in two of these strategies the different objectives can be added and combined into a single objective, known as the

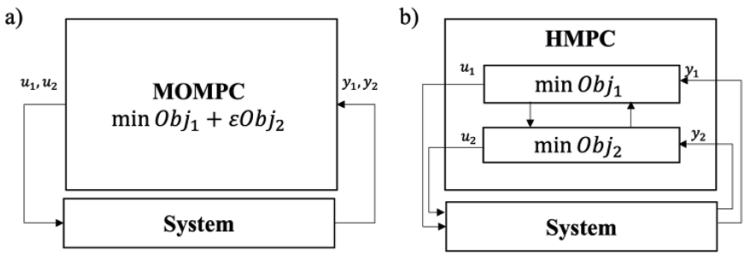


Figure 1 Optimal Controller structures for the incorporation of two objectives for the control of a process system: a) Mutli-objective optimization approach (weighted sum) - MOMPC, b) Hierarchical Optimization Approach – HMPC.

weighted sum method (Fairweather et al. 2012), or objectives can be part of the constraint set, known as the ε-constraint approach (Zavala, 2005; Bemporad et al. 2009).

Another approach that can be used for the consideration of multiple objectives in a control scheme is to rank the objectives in a hierarchy, where every control level in the hierarchy is controlling a subset of the overall control variables, by manipulating a subset of the overall control variables (Avraamidou and Pistikopoulos, 2017; Katebi and Johnson, 1997). Figure 1 illustrates both aforementioned approaches for the consideration of multiple objectives in discrete-time optimal control problems, where u_1 and u_2 are the set of manipulated variables, and y_1 and y_2 are the set of control variables.

MPC strategies require the solution of the optimization problem at every control time step, making the use of most multi-objective or hierarchical solution methods challenging. To this end, this work proposes the use of multi-parametric programming to generate both multi-objective and hierarchical explicit MPCs. More specifically, assuming that there exist two or more control objectives described by linear or convex quadratic functions, we develop multi-parametric based approaches for the derivation of i) the explicit Pareto front of MOMPC, and ii) the explicit solution of HMPC. The MOMPC problem is reformulated into a multiparametric programming problem (Pappas et al. 2021a), which can then be exactly solved using state-of-the-art algorithms (Pappas et al. 2021b), while the HMPC problem is reformulated into a multiparametric multi-level programming problem, which can be exactly solved using the algorithms proposed in (Avraamidou and Pistikopoulos, 2019a).

The following section focuses on the formulation and solution method for the MOMPC, while section 2 focuses on the formulation and solution method of the HMPC. In section 3, a case study on a reactor with two competing objectives, an economic and a tracking objective, is used to illustrate the two developed control strategies. The results of this study indicate the effect of the two strategies and their applicability for the optimal operation of the stirred tank reactor.

2. Multi-Objective Model Predictive Control Structure

2.1. Problem Formulation

Consider problem (1) where q objectives need to be simultaneously minimized by manipulating the inputs, u_i, to calculate the states, x_i, and outputs y_i, at time step i. The process model is a discrete linear state-space model described by the matrices A, B and C and a prediction horizon of N steps along with a terminal set X are assumed.

$$\begin{aligned} \min_{u_i} \quad & \sum_{j=1}^{q} f_j(x,u) \\ \text{s. t.} \quad & x_{i+1} = Ax_i + Bu_i \\ & y_i = Cx_i \\ & x_i^L \le x_i \le x_i^U \\ & u_i^L \le u_i \le u_i^U \\ & x_N \in X \end{aligned} \tag{1}$$

2.2. Solution Strategy

The first step in the proposed solution approach is to reformulate the multi-objective MPC problem in to an ε-constraint problem, following the approach presented by Pappas et al. 2021a, where the reformulated problem is solved using multi-parametric programing while treating the initial states of the system and the ε variables as parameters.

3. Hierarchical Model Predictive Control Structure

3.1. Problem Formulation

$$\begin{aligned} \min_{u_1} \quad & f_1(x,u) \\ \text{s. t.} \quad & \min_{u_2} \ f_2(x,u) \\ & \vdots \\ & \text{s. t.} \quad \min_{u_q} \ f_q(x,u) \\ & \qquad \text{s. t.} \quad x_{i+1} = Ax_i + Bu_i \\ & \qquad\qquad y_i = Cx_i \\ & \qquad\qquad x_i^L \le x_i \le x_i^U \\ & \qquad\qquad u_i^L \le u_i \le u_i^U \\ & \qquad\qquad x_N \in X \end{aligned} \tag{3}$$

3.2. Solution Strategy

To solve this multi-level optimization problem, the algorithm proposed by Avraamidou and Pistikopoulos 2019a and 2019b is utilised. The proposed algorithm transforms the multi-level optimization problem into a series of single-level optimization problems by solving the lower level problems multi-parametrically while treating the states and upper level variables as parameters.

4. Case-study: Continuous Stirred Tank Reactor

4.1. System Definition

Consider a non-isothermal continuously stirred tank reactor (CSTR), adopted from (Kazantzis and Kravaris, 2000), where the following reaction occurs

$$2Na_2S_2O_3 + 4H_2O_2 \rightarrow Na_2SO_3O_6 + Na_2SO_4 + 4H_2O \tag{4}$$

The reactants and the products of the above components are represented by A and B and C, D and E respectively. It is assumed that stoichiometry is preserved in the reactor at all times and hence the reactants are fed to the reactor through a feedstock stream at concentration $C_{A,in}$ and $C_{B,in}$, for A and B respectively, at a ratio $C_{B,in}$:2 $C_{A,in}$, flowrate F,

and temperature T_{in}. The CSTR is assumed to have a constant liquid hold-up. A jacket provides energy to or from the reactor.

It is desired that the outlet concentration and temperature of the controlled variables of the CSTR. The inlet concentrations and temperature are varying between 0.9 and 1.2 mol/L, and 275 to 295 K respectively, and can be considered as measured disturbances. The inlet dilution rate $\left(\frac{F}{V}\right)$ and the coolant temperature T_j can be manipulated by the control system and are therefore considered as manipulated variables. The inlet concentration and temperature are treated as measured disturbances.

4.2. Controller Development – PAROC Framework

To develop the control system for the CSTR defined above, the PAROC framework (Pistikopoulos et al. 2015) was followed.

4.2.1. High-fidelity Model

As a first step a high-fidelity model (5-6) was developed by applying first principles and standard modelling assumptions (constant density and heat capacity, Arrhenius rate, etc.). The details of the model can be found in (Kazantzis and Kravaris, 2000).

$$\frac{dC_A}{dt} = \frac{F}{V}\left(C_{A,in} - C_A\right) - 2k(T)C_A^2 \tag{5}$$

$$\frac{dT}{dt} = \frac{F}{V}(T_{in} - T) + 2\frac{(-\Delta H)_R}{\rho C_p}k(T)C_A^2 - \frac{UA}{V\rho C_p}(T - T_j) \tag{6}$$

where $k = 2k_0 \exp\left(-\frac{E}{RT}\right)$.

4.2.2. Model Approximation

Due to the dynamic nature of the system and its nonlinear components, the original model is linearized around the steady-state of $\begin{bmatrix} C_{A,s} \\ T_s \end{bmatrix} = \begin{bmatrix} 0.076 \\ 376.270 \end{bmatrix}$. Subsequently the linear system ordinary differential equation is discretized using a discretization step of 1 second assuming zero order hold. Consequently, the model is a now a discrete time-invariant state-space model. Here we are also using the variables in deviation form for the inputs and the outputs (e.g. $\hat{C}_A = C_A - C_{A,s}$).

4.2.3. Controller Formulation

Four different control strategies were implemented. The first controller is a classic explicit MPC controller with the tracking objective formulated in (7).

Tracking Objective:

$$\min_{\hat{F},\hat{T}_j} \left[\hat{C}_{A_N} \quad \hat{T}_N\right] P \left[\hat{C}_{A_N} \quad \hat{T}_N\right]^T + \sum_{i=1}^{i=N-1} \left[\hat{C}_{A_i} \quad \hat{T}_i\right] Q \left[\hat{C}_{A_i} \quad \hat{T}_i\right]^T + \left[\frac{\hat{F}}{V_i} \quad \hat{T}_{j,i}\right] R \left[\frac{\hat{F}}{V_i} \quad \hat{T}_{j,i}\right]^T \tag{7}$$

The second controller is an explicit economic MPC with the same constraints as the first controller but with objective (8).

Economic Objective:

$$\min_{\substack{\hat{F} \\ \bar{V}}} \sum_{i=1}^{i=N-1} \left[\frac{\hat{F}}{V_i}\right] V \left[\frac{\hat{F}}{V_i}\right]^T \tag{8}$$

The third controller is a multi-objective economic and tracking controller (MOMPC) with both objectives (7) and (8) in its objective function., while the last controller is a bi-level controller (HMPC) with the economic objective (8) on the upper level optimization problem and the tracking objective (7) on the lower level optimization problem. The constraint set for both the MOMPC and HMPC controllers is identical to the two single objective controllers.

The two single level explicit controllers were solved through POP toolbox, whereas the MOMPC and HMPC were solved as described in sections 2 and 3 respectively. The parameters for all optimization problems consist of the states, the measured disturbances, the previous control action and the output set-point. The prediction horizon was set to 2. The pareto front resulting from the MOMPC controller is presented in Figure 2.

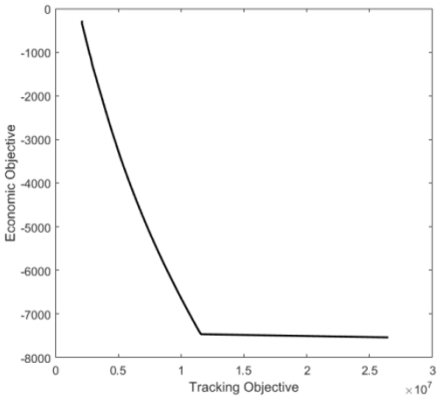


Figure 2 Pareto front of the MOMPC

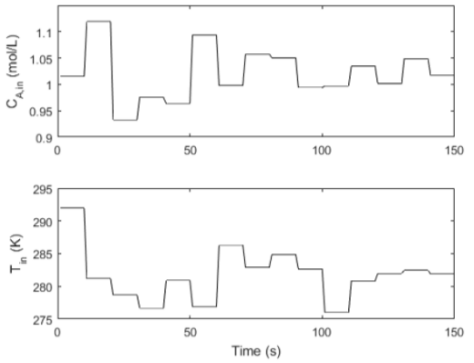


Figure 3 Process Disturbances

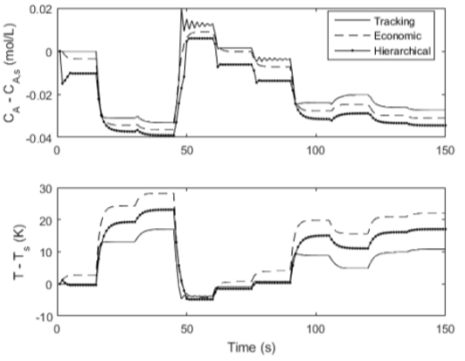


Figure 4 Process Output

4.3. Closed-loop Validation

The last step is the closed-loop validation to evaluate the performance of the designed controllers. The inputs and results of this step are presented in Figures 3 and 4.

5. Conclusion

We presented two multi-parametric based approaches for the incorporation of multiple objectives in model predictive control. A simple CSTR system, with both economic and set-point tracking objectives was used to illustrate the effectiveness of the proposed approaches. The resulting explicit controllers were able to effectively reject disturbances and maintain the system at the given set-points according to their objectives.

Acknowledgements

This work is supported by the DOE-CESMII Energy Efficient Operation of Air Separation Processes Project (DE-EE0007613) and the Rapid Advancement in Process Intensification Deployment (DE-EE0007888-09-04). The authors are also grateful for financial support from UW-Madison, and Texas A&M Energy Institute.

References

M. Fairweather, M. Vallerio, F. Logist, J. F. Impe, 2012, Towards enhanced weight selection for (N)MPC via multi-objective optimization, Proceedings of the 22nd European Symposium on Computer Aided Process Engineering

V. M. Zavala, 2015, A multiobjective optimization perspective on the stability of economic MPC, IFAC-PapersOnLine 48, 8, 974-980

A. Bemporad, D. Munoz de la Pena, 2009, Multiobjective model predictive control, Automatica, 45, 2823-2830

S. Avraamidou, E. N. Pistikopoulos, 2017, A multi-parametric bi-level optimization strategy for hierarchical model predictive control. 27th European Symposium on Computer-Aided Process Engineering (ESCAPE-27); Elsevier, 1591-1596

M. Katebi, M. Johnson, 1997, Predictive control design for large-scale systems, Automatica, 33 (3), 421–425

I. Pappas, D. Kenefake, B. Burnak, S. Avraamidou, H. Ganesh, J. Katz, N. A. Diangelakis, E. N. Pistikopoulos, 2021a, Multiparametric Programming in Process Systems Engineering: Recent Developments and Path Forward, Frontiers in Chemical Engineering, 2, 32

I. Pappas, S. Avraamidou, J. Katz, B. Burnak, B. Beykal, M. Türkay, E. N. Pistikopoulos, 2021b, Multiobjective Optimization of Mixed-Integer Linear Programming Problems: A Multiparametric Optimization Approach. Industrial & Engineering Chemistry Research, 60, 23, 8493-8503

S. Avraamidou, E. N. Pistikopoulos, 2019a, A Multi-Parametric optimization approach for bilevel mixed-integer linear and quadratic programming problems, 125, 98-113

S. Avraamidou, E. N. Pistikopoulos, 2019b, B-POP: Bi-level parametric optimization toolbox. Computers & Chemical Engineering, 122, 193-202

E. N. Pistikopoulos, N. A. Diangelakis, R. Oberdieck, M. M. Papathanasiou, I. Nascu, M. Sun, 2015, PAROC-An integrated framework and software platform for the optimisation and advanced model-based control of process systems, Chemical Engineering Science, 136, 115-138

N. Kazantzis, C. Kravaris, 2000, Synthesis of state feedback regulators for nonlinear processes. Chemical Engineering Science, 55, 17, 3437-3449

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50068-3

A Robust Optimization Strategy for Explicit Model Predictive Control

Iosif Pappas[a,b], Nikolaos A. Diangelakis[a,b], Richard Oberdieck[c], Efstratios N. Pistikopoulos[a,b,*]

[a]*Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX 77843, U.S.A.*
[b]*Texas A&M Energy Institute, Texas A&M University, College Station, TX 77843, U.S.A.*
[c]*Gurobi Optimization, Beaverton, OR 97008, U.S.A.*
stratos@tamu.edu

Abstract

Explicit model predictive control is an established strategy to calculate the model-based optimal control decisions for a process system, while alleviating the computational cost of repetitively solving an optimization problem online. Since models are not ideal representations of the original processes and due to a potentially necessary approximation of the original model for computational efficiency reasons, the optimal solution of the explicit model predictive control problem is based on an imperfect model. Hence, the aforementioned model discrepancy and the presence of unmeasured disturbances facilitate uncertainty that can result to undesirable process behaviour or infeasibility. In this work, a strategy that derives the explicit solution of a robust model predictive control problem with a single multiparametric formulation is exhibited. The proposed approach is founded on the successive robustification of the constraint set of the problem, eliminating the risk of constraint violation, and hence guaranteeing feasibility of closed-loop operation. The benefits of the presented methodology are demonstrated through a linear quadratic regulator problem of an uncertain system.

Keywords: Model Predictive Control; Robust Optimization; Multiparametric Programming

1. Introduction

Model predictive control (MPC) is the established paradigm for the advanced control of multivariable systems, extensively studied and applied by both the academic and industrial communities (Mayne, 2014). In its most encountered form, an MPC problem is a convex quadratic mathematical optimization problem, whose solution is the optimal vector of inputs to regulate the operation of a system. Assuming a finite prediction horizon and a discrete time-invariant linear model, a performance index is minimized to calculate the optimal behavior of the underlying model. Subsequently, only the first control input is applied to the system and the horizon is shifted forward by one step. This procedure is repeated for each sampling time, when new measurements (or estimates of theirs) are made available, facilitating an implicit feedback policy. Explicit MPC refers to the approach of deriving exactly the same optimal vector of inputs, but in an explicit manner. Instead of repetitively solving an optimization problem, the optimal decisions are

expressed analytically (explicitly) by treating the MPC problem as a multiparametric optimization problem. By solving multiparametric optimization problems, the optimal solution of the studied problem is provided as a function of the vector of its uncertain parameters, based on its location at the uncertainty space (critical regions). In the case of MPC, the states are part of the uncertainty vector. The benefits of the explicit solution are, i) the online computational cost of calculating the solution of the problem is substantially reduced by substituting solving an optimization problem with a function evaluation, ii) a complete analysis of the uncertainty (state) space is available a priori, as well as its impact on the solution of the control problem, and iii) the explicit nature of the solution allows for the solution of nested optimization problems. These properties are of particular importance for control applications which are not equipped with the computational power to solve optimization problems online (Pappas et al., 2021).

Robustness is a fundamental element in process control and refers to the ability of the controller to handle uncertainty, and especially unmeasured uncertainty. Since MPC is a model-based control strategy, the quality of the solution is based on the considered model. Nevertheless, process models are not ideal representations of the real system which is to be regulated. In addition, an approximation of the original model is typically required, since the latter is comprised — in many cases — by a large-scale system of differential and algebraic equations that is computationally challenging to be solved online. Finally, the operation of processes includes unmeasured disturbances which affect real-time operations. All of the above sources of uncertainty lead to the undesirable or even infeasible behavior of the plant in closed-loop. For this reason, robust MPC strategies have been proposed to deal with this issue (Kouvaritakis and Cannon, 2016).

Robust explicit MPC aims to derive the explicit solution of the MPC problem by additionally guaranteeing that all sources of uncertainty are taken into account, and at least feasibility is satisfied. In this respect, multiple research efforts have been contributed that tackle the case where the uncertainty source is added to the future prediction (additive uncertainty) and the case where the future prediction is multiplicatively affected by it (multiplicative uncertainty). Sakizlis et al. (2004) included constraints in the design phase of the controller that guarantee that for the worst case of the additive uncertainty, the system is feasible. Bemporad et al. (2003) proposed a min-max approach where the solution is found for problems with a linear objective function and linear constraints for multiplicative uncertainty. Kouramas et al. (2013) tackled the case of explicit MPC problems with a quadratic objective function, linear constraints, and multiplicative uncertainty by employing dynamic programming and robust optimization. More recently, Oberdieck (2016) demonstrated that dynamic programming can be avoided and extended it to hybrid systems by performing projections of the feasible space into the future. This projection operation was achieved my solving a multiparametric linear programming problem. An open question in the robust explicit MPC field is how can the robust solution of an explicit MPC problem with a quadratic cost, linear constraints, and multiplicative uncertainty be developed, by using a single multiparametric optimization formulation.

In this contribution, an algorithm that solves the aforementioned challenge is presented. We formulate a suitable robust control invariant set, successively robustify the constraints, and incorporate linear manipulations to formulate a single multiparametric problem. We solve the optimization program, and as a result, the implementation of the robust policy hedges against the presence of uncertainty and manages to regulate the

system. The remainder of this paper is organized as follows: Section 2 describes the problem formulation, while in Section 3 we present the proposed approach. In Section 4 we demonstrate the benefits of the strategy while in Section 5, we conclude.

2. Problem Formulation

Consider a linear discrete time dynamic model of the following form:

$$x_{k+1} = Ax_k + Bu_k \tag{1}$$

where $x_k \in \mathbb{R}^m$ and $u_k \in \mathbb{R}^n$ are the state and control input vectors respectively at time instant k, and are multiplied by the matrices $A \in \mathbb{R}^{mxm}$ and $B \in \mathbb{R}^{mxn}$. Instead of considering that the system matrices are constant, in this study we assume that the model is uncertain and described by box uncertainty. Specifically:

$$A = A_0 + \Delta A \tag{2}$$

$$B = B_0 + \Delta B \tag{3}$$

$$\Delta A \in \mathbb{A} = \{\Delta A \in \mathbb{R}^{mxm} | -\epsilon_\alpha |A_0| \leq \Delta A \leq \epsilon_\alpha |A_0|\} \tag{4}$$

$$\Delta B \in \mathbb{B} = \left\{\Delta B \in \mathbb{R}^{mxn} \middle| -\epsilon_\beta \middle|B_0\middle| \leq \Delta B \leq \epsilon_\beta |B_0|\right\} \tag{5}$$

A_0 and B_0 are the nominal matrices of the model, while ΔA and ΔB is their uncertain component. This element-wise deviation from the nominal matrix value is prescribed by the matrices ϵ_α and ϵ_β which are of equivalent dimensions to A_0 and B_0. The consideration of box uncertainty allows for its description through halfspace representation, and hence avoids the performance of vertex enumeration which would have been the case if a general polytopic uncertainty set was considered (Oberdieck, 2016). Assuming a prediction horizon N, the following robust explicit linear quadratic regulator problem (LQR) problem can be formulated:

$$\begin{aligned} \min_{u_0,\dots,u_{N-1}} \quad & x_N^T P x_N + \sum_{k=0}^{N-1} x_k^T Q_R x_k + u_k^T R u_k \\ s.t. \quad & x_{k+1} = Ax_k + Bu_k \\ & x_k \in X \\ & u_k \in U \\ & x_N \in T \end{aligned} \tag{6}$$

whose objective is to find the explicit control inputs, $u(x)$, that will drive the system to the origin in the presence of the uncertainty. The weights on the states and inputs are $Q_R \in \mathbb{R}^{mxm}$ and $R \in \mathbb{R}^{nxn}$ respectively, while $P \in \mathbb{R}^{mxm}$ is the terminal cost matrix derived from the solution of the discrete-time algebraic Riccati equation. The states at the end of the prediction are required to belong to the terminal set T (Blanchini, 1999).

3. Methodology

The methodology presented in this section has the goal of ensuring feasibility of the uncertain system. Specifically, the first step of our approach is to derive the robust

counterpart of problem (6), which requires the successive robustification of the state constraints of the formulation. Assume that the state constraints for the first timestep are expressed as:

$$Gx_1 \leq g \tag{7}$$

$$G(Ax_0 + Bu_0) \leq g \tag{8}$$

The robust counterpart of the above constraint is:

$$GA_0x_0 + \epsilon_\alpha |G||A_0||x_0| + GB_0u_0 + \epsilon_\beta |G||B_0||u_0| \leq g \tag{9}$$

The nonlinearity introduced from the absolutes values is addressed by introducing the artificial variables z_0 and v_0, along with their corresponding box constraints:

$$\begin{aligned} &GA_0x_0 + \epsilon_{\alpha|}|G||A_0|z_0 + GB_0u_0 + \epsilon_\beta |G||B_0|v_0 \leq g \\ &-z_0 \leq x_0 \leq z_0 \\ &-v_0 \leq u_0 \leq v_0 \end{aligned} \tag{10}$$

By following the proposed robustification scheme, all constraints are successively robustified for the length of the prediction horizon. That enforces state constraint satisfaction until – and including – the N^{th} step of the horizon. Moreover, the system states are required to enter the invariant set at the N^{th} horizon step, hence the system feasibility thereafter is also ensured. In summary, at each robustification step, an artificial variable for each state and control input is introduced. This challenge is addressed by eliminating the state artificial variables through the Fourier-Motzkin (FM) elimination, which allows for removal of the artificial variables. As an example, assume that $a_{i,j}$ are scalar coefficients and that $a_{1,2} \geq 0$ and that $a_{2,1} \leq 0$:

$$a_{1,1}x_0 + a_{1,2}z_0 + a_{1,3}u_0 + a_{1,4}v_0 \leq g_1 \tag{11}$$
$$a_{2,1}x_0 + a_{2,1}z_0 + a_{2,3}u_0 + a_{2,4}v_0 \leq g_2 \tag{12}$$

That can be rewritten as

$$z_0 \leq \frac{1}{a_{1,2}}\left(g_1 - a_{1,1}x_0 - a_{1,3}u_0 - a_{1,4}v_0\right) \tag{13}$$
$$z_0 \geq \frac{1}{a_{2,1}}\left(g_2 - a_{2,1}x_0 - a_{2,3}u_0 - a_{2,4}v_0\right) \tag{14}$$

Consequently, the variable z_0 can be eliminated by combining the two expressions:

$$\frac{1}{a_{2,1}}\left(g_2 - a_{2,1}x_0 - a_{2,3}u_0 - a_{2,4}v_0\right) \leq \frac{1}{a_{1,2}}\left(g_1 - a_{1,1}x_0 - a_{1,3}u_0 - a_{1,4}v_0\right) \tag{15}$$

However, the drawback of the FM elimination is the introduction of additional inequality constraints in the problem. Hence, after applying the FM algorithm, we eliminate the unnecessary inequality constraints by solving a linear programming problem to check redundancy. As a result, a robustified version of problem (6) is derived and is solved with state-of-the-art multiparametric optimization algorithms. We note that the proposed

approach is applicable to mixed-integer linear models too have recently shown to play an important role in model building for explicit MPC applications (Katz et al., 2020).

4. Results

Consider an uncertain system of form (1), adopted from (Kouramas et al., 2013). The nominal matrices of the system are $A_0 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $B_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. It is assumed that these nominal matrices can deviate from their nominal value by 20% (i.e. $\varepsilon_\alpha = \varepsilon_\beta = 0.2$). The horizon of the problem is $N = 2$, while the cost matrices are $Q_R = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, $R = 0.01$ and $P = \begin{bmatrix} 2.62 & 1.63 \\ 1.63 & 2.64 \end{bmatrix}$. The formulation is solved using the approach presented previously. The problem has eighty two critical regions and is simulated in closed-loop for a constant and random value of the model. The controller can stir the system to the origin.

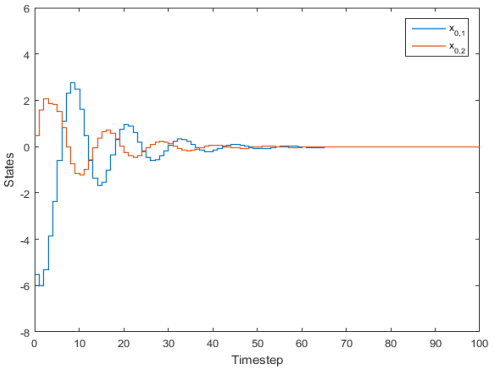


Figure 1: Closed-loop simulation of the system for a random but constant value of the system matrices.

As a next step we generate multiple scenarios of the matrices which are randomly altered at each time step of the closed-loop simulation. The system is regulated for all of them, achieving the control objective of driving the system to the origin.

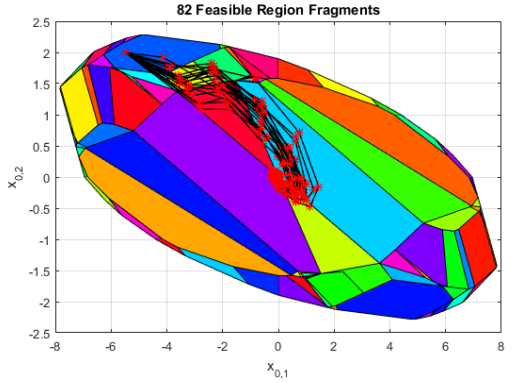


Figure 2: The map of optimal solutions along with closed-loop simulations of the system for multiple scenarios where the values of the matrices are randomly altered.

As expected, the robustification of the constraints forces the system to remain feasible for all different scenarios, while having the benefit of the explicit form of the solution.

5. Conclusions

In this work, we presented an algorithm to solve robust explicit MPC problems. Our approach is based on the successive robustification of the constraints of the problem which along with the terminal set guarantee the feasibility of the system in closed-loop. Additionally, we eliminated the complexity introduced by using the FM elimination algorithm and redundancy checks to remove variables and constraints respectively from the problem formulation. We demonstrated our findings on a numerical MPC case study where we exhibited that the system can be driven to the original for any arbitrary bounded value of the uncertainty. Our next steps include the analysis of the conservativeness of the solution stemming from the robustification.

6. Acknowledgements

The financial support from the Texas A&M Energy Institute, and CESMII (DE-EE0007613, 4550 G WA324) and RAPID (DE-EE0007888-09-04) Institutes are gratefully acknowledged.

7. References

A. Bemporad, F. Borelli, M. Morari, 2003, Min-max control of constrained uncertain discrete time linear systems. IEEE Transactions on automatic control, 48(9), 1600-1606.

B. Kouvaritakis, M. Cannon, 2016, Model Predictive Control, Switzerland: Springer International Publishing, 38.

D. Q. Mayne, 2014, Model predictive control: Recent developments and future promise, Automatica, 50, 12, 2967-2986.

F. Blanchini, 1999, Set invariance in control. Automatica, 35, 11, 1747-1767.

I. Pappas, D. Kenefake, B. Burnak, S. Avraamidou, S. H. Ganesh, J. Katz, N. A. Diangelakis, E. N. Pistikopoulos, 2021, Multiparametric Programming in Process Systems Engineering: Recent Developments and Path Forward. Frontiers in Chemical Engineering, 2.

J. Katz, I. Pappas, S. Avraamidou, E. N. Pistikopoulos, 2020, Integrating Deep Learning Models and Multiparametric Programming. Computers & Chemical Engineering, 138, 106801.

K. Kouramas, C. Panos, N. P. Faísca, E. N. Pistikopoulos, 2013, An algorithm for robust explicit/multi-parametric model predictive control. Automatica, 49, 2, 381-389.

R. H. Oberdieck, 2016, Theoretical and algorithmic advances in multi-parametric optimization and control, Ph.D. Thesis, Imperial College London, London, U.K.

V. Sakizlis, N. M. Kakalis, V. Dua, J. D. Perkins, E. N. Pistikopoulos, 2004, Design of robust model-based controllers via parametric programming. Automatica, 40, 2, 189-201.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50069-5

Data-driven Design of a Feed-forward Controller for Rejecting Measurable Disturbance

Yoichiro Ashida[a*], Masanobu Obika[b]

[a]*Department of Electrical Engineering and Computer Science, National Institute of Technology, Matsue College, 14-4 Nishiikuma-cho, Matsue, 690-8518, JAPAN*
[b]*ADAPTEX Co., Ltd., 3-13-36 Kagamiyama, Higashi-Hiroshima 739-0046, JAPAN*

yashida@matsue-ct.jp

Abstract

Many large-scale multi-input multi-output systems are treated as a combination of single-input single-output systems in reality. At such times, interference from input signals not focused on work as disturbances. For observable disturbances, feed-forward controllers are effective to reject the influence. On the other hand, many data-driven controller tuning schemes are proposed for feed-back controllers. The schemes require not any mathematical models of controlled systems but only operating-data like input and output. This paper proposes a data-driven tuning scheme of feed-forward controllers. Existing data-driven scheme tunes feed-back controller at the same time as the feed-forward controller. In contrast, the proposing scheme only designs the feed-forward controller. By this feature, it is easy to guarantee stability of the control system. Effectiveness of the proposing scheme is verified by a simulation example.

Keywords: process control, disturbance rejection, feed-forward controller

1. Introduction

Most large-scale processes are multi-input multi-output (MIMO) systems. However, the processes are often treated as a group of single-input single-output (SISO) systems because it is difficult to design suitable controllers for a MIMO system. By focusing each SISO system, interference from other SISO systems can be regarded as disturbances. Therefore, disturbance rejection is very important in large-scale process control.

Feed-back controller like PID controller is often employed to realize set-point tracking, and some parameters tuning methods are proposed. Among them, data-driven tuning methods are actively researched. Typical methods are iterative feed-back tuning (IFT) which uses repeated experiment proposed by Hjalmarsson et al. (1998), fictitious reference iterative tuning (FRIT) which uses only off-line optimizations proposed by Soma et al. (2004). The methods can tune controller without any system parameters. Effectiveness of the schemes are verified for experiments. For example, Nakamoto (2003) and Kano et al., (2011) apply IFT and extended-FRIT methods to processes respectively.

Feed-back controllers can improve performance not only set-point tracking but also disturbance rejection. However, it is impossible to reject influence of disturbance perfectly when controlled process has time-delay shown in Alagoz et al. (2015). In addition, the

longer the time-delay are, the larger influence of disturbance are. To solve this problem, feed-forward controllers are often employed like Elso et al. (2013). When disturbance can be observable, the feed-forward controller can reject influence of disturbance completely.

The objective of this paper is to propose a data-driven design method of the disturbance rejection feed-forward controller. In the proposing method, not both feed-back and feed-forward controllers but only feed-forward controller is tuned. Sometimes, control-loop becomes unstable by tuning feed-back controller. Therefore, designing only feed-forward controller is safer than designing both controllers. In addition, FRIT method is employed to determine controller parameters. In the FRIT method, an evaluation function to be minimized is derived directly from tracking error signal. Thus, the evaluation function of FRIT and tracking error have close connection. Effectiveness of the proposing design method is checked by a numerical example.

2. Design Scheme of Disturbance Rejection Controller

2.1. Feed-forward Controller

This research assumes a control system as shown in Figure 1. $r(t), u(t), y(t)$, and $\nu(t)$ denote reference, input, output, and unknown noise signals respectively. Additionally, $d(t)$ denotes trigger signal of disturbance.

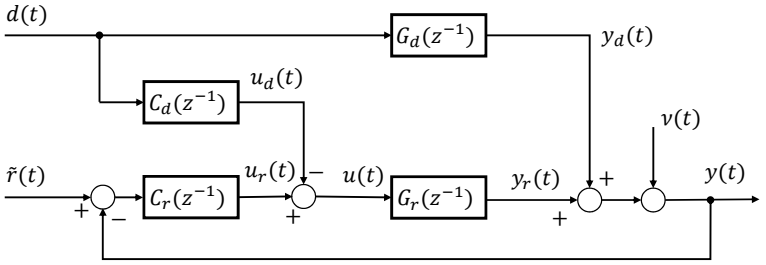


Figure 1: Block diagram of the proposing control system.

Aim of this research is to design disturbance rejection controller $C_d(z^{-1})$. Feed-back controller $C_r(z^{-1})$ is assumed to exist, and proposing design method does not touch the controller. This is because it is easy to ensure stability of control system and to employ the method to industries. Controlled system $G_r(z^{-1})$ and disturbance system $G_d(z^{-1})$ are

$$G_r(z^{-1}) = \frac{B_r(z^{-1})}{A_r(z^{-1})} z^{-k_r}, \tag{1}$$

$$G_d(z^{-1}) = \frac{B_d(z^{-1})}{A_d(z^{-1})} z^{-k_d}, \tag{2}$$

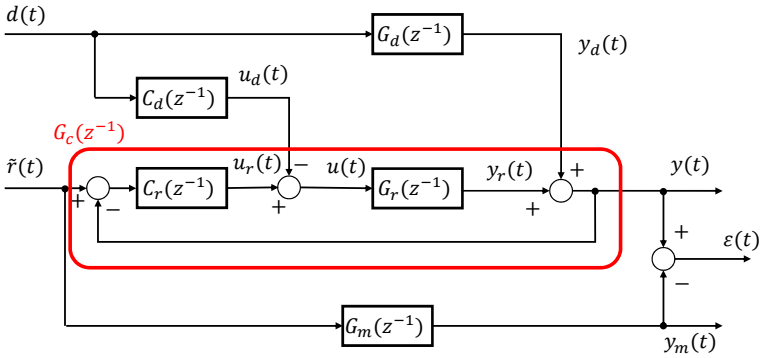


Figure 2: Conceptual diagram of the proposing FRIT method.

where $A_r(z^{-1}), B_r(z^{-1}), A_d(z^{-1})$, and $B_d(z^{-1})$ are

$$A_r(z^{-1}) = 1 + a_{r1}z^{-1} + \cdots + a_{rna}z^{-rna}, \tag{3}$$

$$B_r(z^{-1}) = 1 + b_{r1}z^{-1} + \cdots + b_{rnb}z^{-rnb}, \tag{4}$$

$$A_d(z^{-1}) = 1 + a_{d1}z^{-1} + \cdots + a_{dna}z^{-dna}, \tag{5}$$

$$B_d(z^{-1}) = 1 + b_{d1}z^{-1} + \cdots + b_{dnb}z^{-dnb}. \tag{6}$$

When $C_d(z^{-1})$ is designed as

$$C_d(z^{-1}) = G_d(z^{-1})G_r(z^{-1})^{-1}, \tag{7}$$

influence of $d(t)$ is rejected from $y(t)$. To realize the controller, $k_d \geq k_r$ must be hold.

2.2. Data-driven controller tuning

Soma et al.(2004) proposes data-driven controller tuning named fictitious reference iterative tuning (FRIT) to tune feed-back controllers for set-point tracking. This paper extends FRIT method for designing disturbance rejection feed-forward controller $C_d(z^{-1})$. Conceptual diagram of proposing FRIT is shown in Figure 2.

$\varepsilon(t)$ denotes error as $\varepsilon(t) = y(t) - G_m(z^{-1})\tilde{r}(t)$, thus minimizing $\varepsilon(t)$ means minimizing control error between reference trajectory and control output directly.

Assuming that one-set of operating-data $u(t), y(t)$ and $d(t)$ has been obtained, and closed-loop transfer functions with $d(t) = 0$ is set as $G_c(z^{-1})$. When Eq.(7) holds, influence of $d(t)$ is neglected from $y(t)$, and $y(t)$ can be expressed as $y(t) = G_c(z^{-1})\tilde{r}(t)$. Therefore, $\varepsilon(t)$ becomes zero when $G_m(z^{-1})$ and $G_c(z^{-1})$ are identical. By using these relations, proposing FRIT determines $C_d(z^{-1})$.

In Figure 2, $u(t)$ is calculated as

$$u(t) = C_r(z^{-1})r(t) - C_r(z^{-1})y(t) - C_d(z^{-1})d(t). \tag{8}$$

Based on Eq.(8), fictitious reference signal $\tilde{r}(t)$ is defined as

$$\tilde{r}(t) := y(t) + C_r(z^{-1})^{-1}\left\{u(t) + C_d(z^{-1})d(t)\right\}. \tag{9}$$

Therefore, $\varepsilon(t)$ is expressed as

$$\varepsilon(t) = y(t) - G_m(z^{-1}) \left[y(t) + C_r(z^{-1})^{-1} \left\{ u(t) + C_d(z^{-1}) d(t) \right\} \right]. \tag{10}$$

From the previous discussion, suitable $C_d(z^{-1})$ can be obtained by minimizing the following cost function J:

$$J := \sum_{i=1}^{N} \varepsilon(i)^2, \tag{11}$$

where N denotes size of operating-data.

The following two sets of optimization variables can be considered in Eq.(10).

1. $C_d(z^{-1})$ and $G_m(z^{-1})$,
2. $G_r(z^{-1})$ and $G_d(z^{-1})$.

The first approach directly determines controller $C_d(z^{-1})$ and closed-loop model $G_m(z^{-1})$. Although ideal $C_d(z^{-1})$ of Eq.(7) and closed-loop $G_c(z^{-1})$ both include $G_r(z^{-1})$, first approach determines both of them independently. In contrast, the second approach determines controlled system and disturbance transfer function. From controlled system $G_r(z^{-1})$ and known $C_r(z^{-1})$, $G_m(z^{-1})$ is easily calculated, and $C_d(z^{-1})$ is also calculated using $G_r(z^{-1})$ and $G_d(z^{-1})$ by Eq.(7). Although it looks system identification, this is still data-driven tuning because minimized error is not modeling error but control error $\varepsilon(t)$. This paper employs the second approach.

3. Numerical examples

Simulations of this section was executed as Figure 1. $G_r(z^{-1})$, $G_d(z^{-1})$, and $C_r(z^{-1})$ were set as

$$G_r(z^{-1}) = \frac{0.0004821 z^{-1} + 0.0004648 z^{-2}}{1 - 1.895 z^{-1} + 0.8958 z^{-2}} z^{-30}, \tag{12}$$

$$G_d(z^{-1}) = \frac{0.002415 z^{-1} + 0.002332 z^{-2}}{1 - 1.9000 z^{-1} + 0.9003 z^{-2}} z^{-50}, \tag{13}$$

$$C_r(z^{-1}) = \frac{0.02 - 0.01 z^{-1}}{\Delta}. \tag{14}$$

In addition, unknown noise was introduced as

$$v(t) = \frac{0.004988 z^{-1}}{1 - 0.995 z^{-1}} \xi(t), \tag{15}$$

where $\xi(t)$ is a Gaussian white noise with zero mean and 3.0^2 variance.

Initial operating-data which was obtained with $C_d(z^{-1}) = 0$ is shown as Figure 3. It is clear that controlled output is affected by disturbance.

Next, the following $G_r(z^{-1})$ and $G_d(z^{-1})$ were obtained by the proposing method. For minimization of J, fminunc function of MATLAB R2021a software was employed. The

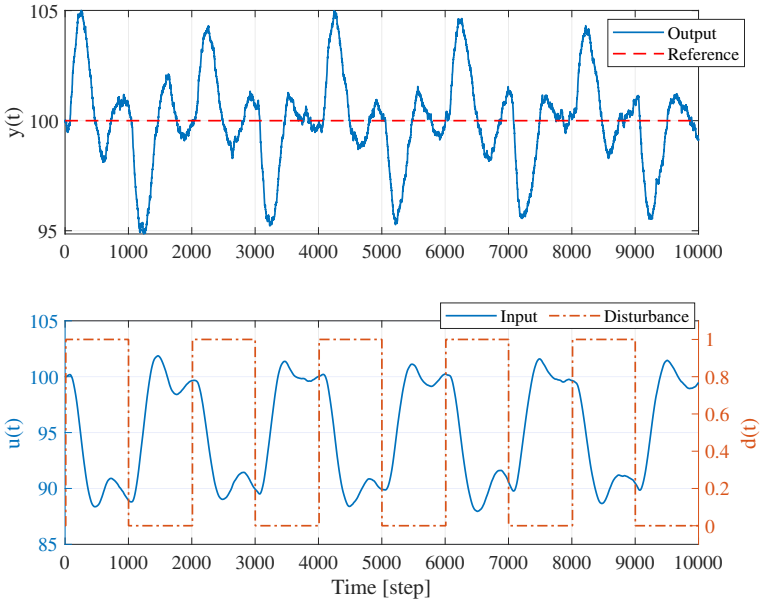


Figure 3: Initial operating-data with $C_d(z^{-1}) = 0$.

function uses quasi-Newton method with bfgs algorithm for optimization. Time-delays were given as known parameters.

$$G_r(z^{-1}) = \frac{0.0086z^{-1}}{1 - 0.9923z^{-1}} z^{-30}, \tag{16}$$

$$G_d(z^{-1}) = \frac{0.0448z^{-1}}{1 - 0.9962z^{-1}} z^{-50}. \tag{17}$$

Even though $G_r(z^{-1})$ and $G_d(z^{-1})$ are both second order systems, they are modeled as first order systems in this simulation.

Figure 4 shows the result using $C_d(z^{-1})$ calculated by the proposing method. Upper figure shows that proposing method mostly rejected influence of disturbance. Shown as lower figure, input signal quickly changed after varying disturbance signal. As a result, influence of disturbance was canceled before appearing to the output. $G_r(z^{-1})$ and $G_d(z^{-1})$ are second order systems, and unknown noise is added. Therefore, the proposing scheme is considered effective for some uncertain elements.

4. Conclusions

This paper has proposed a design method of feed-forward disturbance rejection controller. FRIT method has been employed to tune the controller. A feature is to tune not feed-forward and feed-back controllers but only feed-forward controller. Effectiveness of proposing scheme has been verified by a numerical example. It is considered that proposing scheme is effective for system with some uncertain elements.

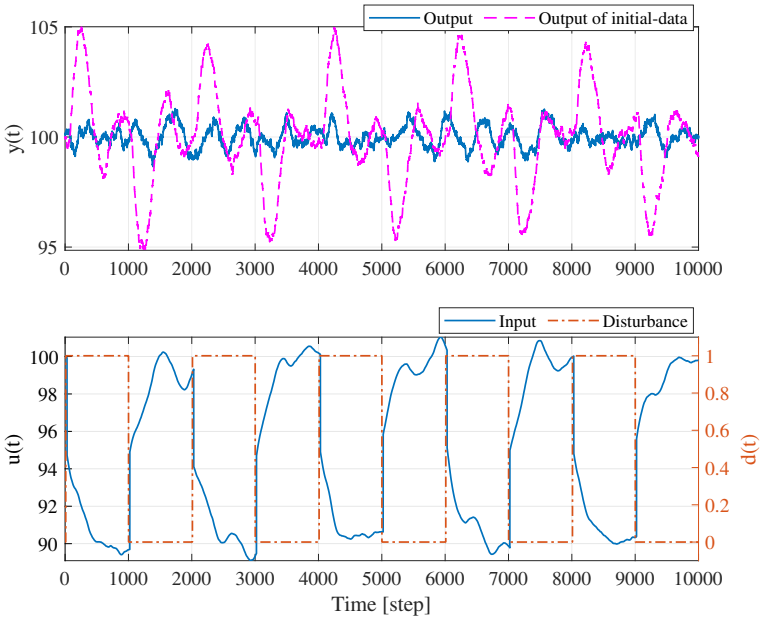


Figure 4: Control result with $C_d(z^{-1})$ calculated by the proposing method.

References

Hjalmarsson, H., Gevers, M., Gunnarsson, S., and Lequin, O., 1998, Iterative Feedback Tuning: Theory and Applications, IEEE Control Systems Magazine, 18, 8, 26-41

Soma, S., Kaneko, O., and Fujii, T., 2004, A New Method of Controller Parameter Tuning based on Input-Output Data – Fictitious Reference Iterative Tuning (FRIT) –, IFAC Proceedings Volumes, 37, 12, 789–794

Nakamoto, M., 2003, An Application of Iterative Feedback Tuning for a Process Control, Trans. of the Society of Instrument and Control Engineers, 39, 10, 924-932

Kano, M., Tasaka, K., Ogawa, M., Takinami, A., Takahashi, S., and Yoshii, S., 2011, Extended Fictitious Reference Iterative Tuning and Its Application to Chemical Processes, Proc. of 2011 International Symposium on Advanced Control of Industrial Processes

Elso, J., Gil-Martínez, M., García-Sanz, M., 2013, Quantitative feedback–feedforward control for model matching and disturbance rejection, IET Control Theory & Applications, 7, 6, 894-900

Alagoz, B.B., Deniz, F.N., Keles, C., Tan, N., 2015, Disturbance rejection performance analyses of closed loop control systems by reference to disturbance ratio, ISA Transactions, 55, 63-71

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50070-1

Optimal Operation of Heat Exchanger Networks with Changing Active Constraint Regions

Lucas Ferreira Bernardino[a*], Dinesh Krishnamoorthy[a], Sigurd Skogestad[a]

[a]*Department of Chemical Engineering, Norwegian University of Science and Technology, Trondheim, Norway*

lucas.f.bernardino@ntnu.no

Abstract

In this paper, we study the optimal operation of heat exchanger networks with stream splits. In particular, we extend previous approaches on the unconstrained optimization of the system to the constrained case, with temperature constraints on each flow branch, and with changing disturbances so that the set of optimally active constraints changes during operation. The simplest way to achieve optimal operation when some of the constraints are active, is to control the constraints to their limiting value, known as active constraint control. For the remaining unconstrained degrees of freedom, we propose to control linear combinations of the gradient as self-optimizing controlled variables. To automatically switch between the different active constraint regions, we use classical advanced control elements such as selectors, thereby achieving optimal operation using only the temperature measurements as feedback in different active constraint regions. The performance of the proposed feedback optimizing control structure for the heat exchange network is compared with the traditional model-based real-time optimization using simulations. In the presence of structural plant-model mistmach, we show that our proposed approach performs optimally for all disturbances, while traditional real-time optimization fails to converge for some cases, as the optimization problem becomes infeasible depending on the estimated disturbances.

Keywords: process control, optimal operation, self-optimizing control, applications

1. Introduction

In the context of optimal operation of process systems, the choice of controlled variables plays a vital role, as it will dictate how efficiently a process can operate without interference of higher layers (Skogestad, 2000). The ideal design of a supervisory control layer would result in a structure that is able to operate optimally under constant setpoints. This concept is known as self-optimizing control, and recent developments aim for systematic choice of control objectives (Krishnamoorthy and Skogestad, 2019). A known challenge in supervisory layer design is the change in optimally active constraints during operation, which can be caused by changes in disturbances that affect process objectives. When that happens, reconfiguration of the controlled structure is usually desired to minimize the operational losses. If that does not happen, interactions with the higher optimization layer become stronger, as the sensitivity of the optimal setpoint values with relation to the changing disturbances is high when there are no changes in the control structure. Krishnamoorthy and Skogestad (2019) discusses the handling of changes in active constraints through

feedback control, without the solution of online optimization problems, by selector-based control structures. This approach is to be evaluated in this work, compared to the solution of real-time optimization (RTO) problems, which can be problematic in the presence of model-plant mismatch.

2. Case study modeling

The case study considered in this work consists of three heat exchangers in parallel, see Figure 1. Each exchanger has its own source of hot fluid, such that the cold fluid is split and sent to the exchangers, and the operational goal is to maximize the outlet temperature of the cold fluid, subject to constraints related to the maximum temperature in the individual exchangers.

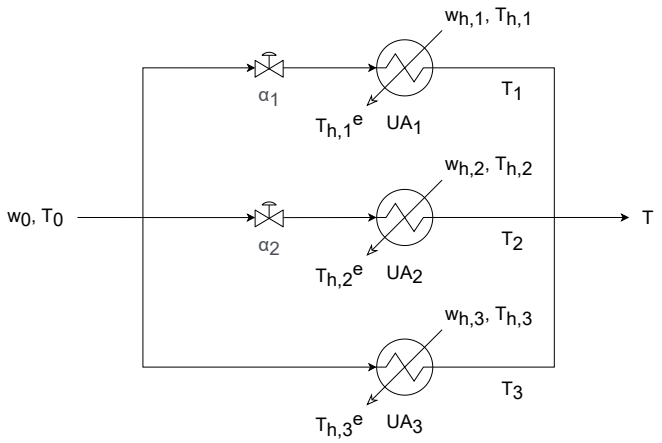


Figure 1: Heat exchanger network scheme

In addition to the mass and energy balances, an additional relation is necessary for calculating the total exchanged heat in each equipment, Q_i. The analytic solution, assuming constant heat capacities and countercurrent flow, is given by Eq.(1).

$$Q_i = UA_i \, \Delta T_{LM,i} \tag{1}$$

In this equation, $\Delta T_{LM,i}$ represents the logarithmic mean of temperature differences inside the heat exchanger. Although exact, this model presents some numerical challenges, especially when the heat capacities are too close, or when the temperature differences assume opposite signs during iteration. A simplified linear version of this model makes use of the arithmetic mean of temperature differences, $\Delta T_{AM,i}$, and for this model, simple analytic expressions for the gradient can be derived (Jäschke and Skogestad, 2014).

The steady-state optimization problem considered for the optimal operation of this system can therefore be written as:

$$\begin{aligned} \min_{\alpha} \quad & J = -T \\ \text{s.t.} \quad & g_i = T_i - T_{max} \leq 0, \quad i = 1, 2, 3 \end{aligned} \tag{2}$$

3. Proposed control structure

The optimal operation of heat exchanger networks has been extensively studied by Jäschke and Skogestad (2014) for the unconstrained case. In this case, the gradient J_u to be driven to zero can be approximately written in terms of the Jäschke temperatures. For the constrained case, however, the set of controlled variables need to change so that optimal operation is achieved. Given that the active constraints g_A are effectively controlled, there are still unconstrained degrees of freedom that need to be used for optimal operation. As proven by Krishnamoorthy and Skogestad (2019), we can find the additional controlled variables as a linear combination of the gradient such that the necessary conditions of optimality are satisfied. These correspond to $c = N^T J_u$, where N is the nullspace of the gradient of the active constraints with relation to the inputs, $\nabla_u g_A$, at the optimal point. This procedure results in a set of controlled variables per region, defined by the respective set of active constraints.

For this case study, there are 7 feasible operating regions, one of which is fully unconstrained, 3 being partially constrained (one active constraint per region), and the remaining being fully constrained (two active constraints per region). The case with all 3 constraints being active is infeasible with the available degrees of freedom, and will therefore not be considered. The fully unconstrained region can be optimally operated by controlling the plant gradient to zero, and the fully constrained regions are optimally operated through active constraint control. For the optimal operation in the partially constrained regions, the combinations of the gradient to be controlled in addition to the active constraints are given in Table 1.

Active constraint	N^T
g_1	$\begin{bmatrix} 0 & 1 \end{bmatrix}$
g_2	$\begin{bmatrix} 1 & 0 \end{bmatrix}$
g_3	$\begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$

Table 1: Linear combinations of gradient per active constraint

The next step for the design of a simple control structure is defining the pairing between manipulated and controlled variables, and the switching between active controllers. In the current case study, there are 2 manipulated variables and 3 constraints, which means that the constraints cannot be assigned to one specific input if optimal operation over all regions is desired. Therefore, at least one of the constraints needs to be controlled by multiple inputs.

Based on this reasoning, this work proposes an adaptive control structure to deal with all possible active constraint regions. The full control structure, showing the logic blocks and controllers, is presented in Figure 2. and the pairing between manipulated and controlled

variables is summarized in Table 2. All presented controllers have integral action, so that steady-state offset is eliminated.

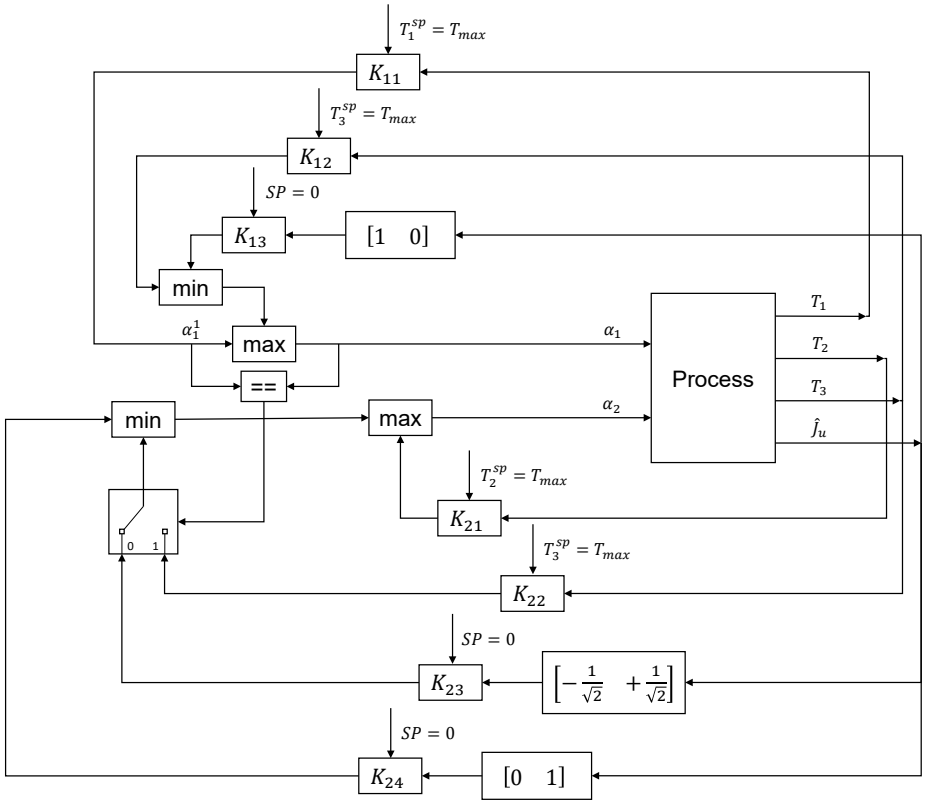


Figure 2: Proposed adaptive control structure

α_1	α_2 (T_1 inactive)	α_2 (T_1 active)
T_1	T_2	T_2
$\begin{bmatrix}1 & 0\end{bmatrix} J_u$	$\begin{bmatrix}0 & 1\end{bmatrix} J_u$	$\begin{bmatrix}0 & 1\end{bmatrix} J_u$
T_3	$\begin{bmatrix}-\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix} J_u$	T_3

Table 2: Proposed adaptive pairing for all operating regions

4. Simulation results and discussion

The control structure previously presented is now evaluated in closed-loop simulation face to changing disturbances. Figure 3 shows the simulation results, where all 7 possible regions are explored. As the process itself is considered to be at steady state at all times, the dynamics of the system is fully attributed to the tuning of the controllers. Operation

in the fully constrained regions is optimal at steady state, whereas there is some deviation from the optimal conditions in the partially constrained and unconstrained regions. This is due to the estimation of gradients by Jäschke temperatures, which does not fully represent the plant model, but gives a reasonable estimate for control, so that low operational loss is achieved.

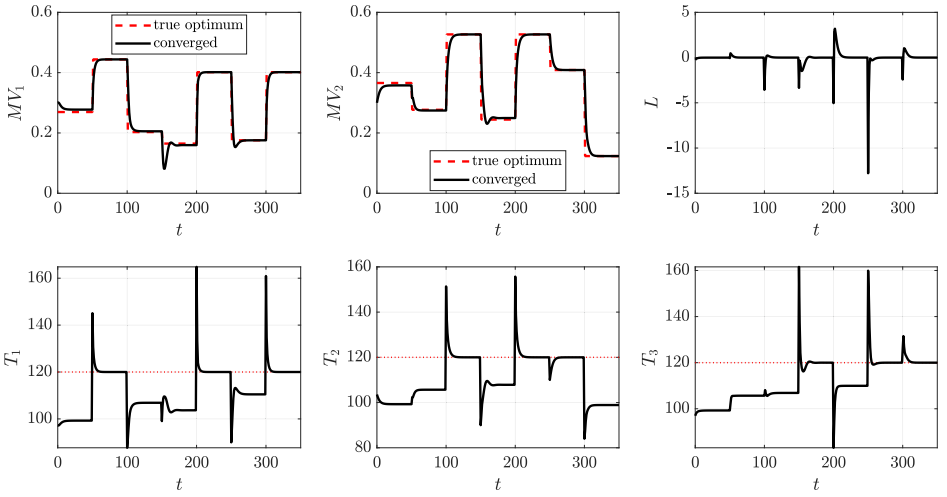


Figure 3: Simulation of region-based control structure using Jäschke temperatures

These results are compared with a traditional RTO implementation, see Figure 4. This implementation consists of a two-step approach, with disturbance estimation followed by model-based constrained optimization. The system converges in few iterations, with similar steady-state behavior to the region-based control structure. The unconstrained and partially constrained regions suffer from deviations from the true optima, due to model-plant mismatch, and the converged state is quite similar to that of the region-based control structure. This is to be expected, as Jäschke temperatures represent the gradient information extracted from the model used in the RTO framework.

In the RTO simulation, a curious undesired behavior is observed. From $t = 40$, in the fifth simulated region, the system converges to an infeasible point. This happens because the disturbance estimation step returns parameter values that make the optimization problem infeasible, meaning that there are no inputs that satisfy all constraints on the model with the given parameters, even if the estimation step returns parameters that agree with the plant measurements. Some workarounds are therefore deemed necessary for the effective implementation of the RTO strategy, such as the adaptation of the optimization problem itself, based on the estimation of gradients from the true plant (Marchetti et al., 2009).

5. Conclusion

In this work, we extended previous work on the optimal operation of heat exchanger networks to the constrained case, where the ideal self-optimizing variables known as Jäschke

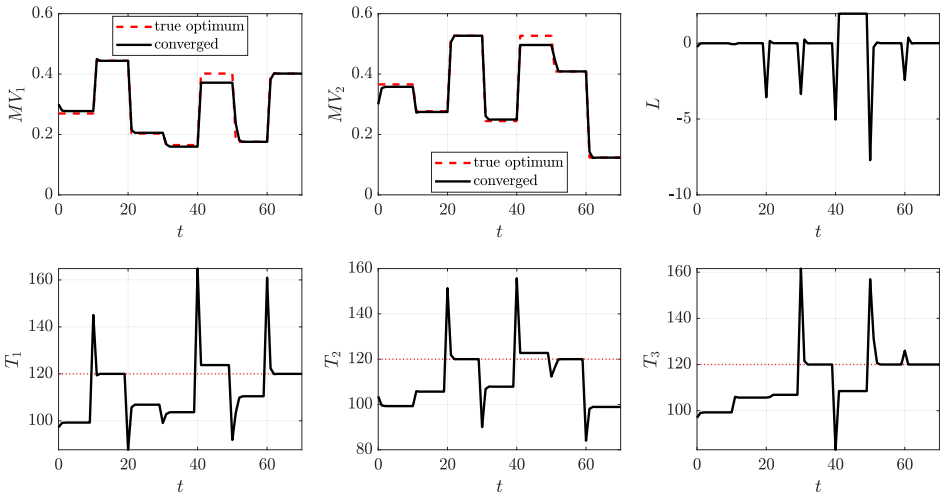


Figure 4: Simulation of steady-state RTO with model-plant mismatch

temperatures cannot be applied to every operating condition. Instead, control of the active constraints becomes necessary for optimal operation, and the challenge lies in deciding automatically what are the best controlled variables during operation. This has been achieved with the use of selectors, with steady-state performance comparable to a traditional model-based RTO implementation. With the proposed control implementation, one avoids the solution of online optimization problems, which can be problematic, as highlighted by the presented results. However, the simultaneous use of the presented tools is encouraged, so that near-optimal operation is achieved in the faster timescales, and optimization tools can correct for mismatches under more careful evaluation of the results.

References

Jäschke, J., Skogestad, S., 2014. Optimal operation of heat exchanger networks with stream split: Only temperature measurements are required. Computers & chemical engineering 70, 35–49.

Krishnamoorthy, D., Skogestad, S., 2019. Online process optimization with active constraint set changes using simple control structures. Industrial & Engineering Chemistry Research 58, 13555–13567.

Marchetti, A., Chachuat, B., Bonvin, D., 2009. Modifier-adaptation methodology for real-time optimization. Industrial & engineering chemistry research 48, 6022–6033.

Skogestad, S., 2000. Plantwide control: The search for the self-optimizing control structure. Journal of process control 10, 487–507.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50071-3

Iterative Feedback Tuning for Regulatory Control Systems Using Estimate of Sensitivity Function

Shiro Masuda[a*]

[a]*Dept. of Systems Design, Tokyo Metropolitan University, 6-6 Asahigaoka, Hino, 191-0065, Japan*

smasuda@tmu.ac.jp

Abstract

The paper provides a method for Iterative Feedback Tuning (IFT) for regulatory control systems. The IFT estimates the gradient of a prescribed cost criterion using collected data. However, the original approaches require special experiments for gradient estimate. The proposed method makes it possible to estimate the sensitivity function that leads to the gradient estimate from regulatory control data. The proposed approach uses two sets of regulatory control data. The first one is used for rough estimate of the sensitivity function. Then, the second data are used for the correction of the estimation error of the sensitivity function. A numerical example shows that the proposed method can optimize the cost criterion even in the case where the identifiability condition does not hold.

Keywords: Iterative feedback tuning, Regulatory control systems, Sensitivity function

1. Introduction

The Iterative Feedback Tuning (IFT), which was originally initiated by Hjalmarsson (1998), is a data-driven controller parameter tuning method that achieves optimal control parameters by way of parameter updating laws using gradient estimates for a prescribed cost criterion. While the original IFT requires a special experiment for gradient estimate, the present work concerns a gradient estimation method using regulatory control data. However, the methods discussed in Kammer (2000) required a certain plant test for estimating sensitivity functions. The method in Masuda (2019) sometimes brings numerical problems in the derivation of gradient estimates. Therefore, the paper provides an estimating method for sensitivity functions leading to the gradient estimate. The proposed approach uses two sets of regulatory control data. The first one is used for the estimation of plant and disturbance model. Then, after calculating a sensitivity function using the estimated plant model, the estimation error is compensated by using the second regulatory control data. The effectiveness of the proposed approach is shown through a numerical example.

2. IFT for regulatory control

2.1. Process description and regulatory control Systems

Consider the following process model described by

$$y(t) = G(q)u(t) + H(q)w(t), \tag{1}$$

where q is a shift operator, i.e. $q^k y(t) = y(t+k)$. $u(t)$ and $y(t)$ are process input and output signals at discrete time instant t. $w(t)$ is zero mean white noise with the variance σ_w^2. The process model $G(q)$ and the disturbance model $H(q)$ are expressed as

$$G(q) = \frac{B(q^{-1})}{F(q^{-1})}, \quad H(q) = \frac{C(q^{-1})}{D(q^{-1})}, \tag{2}$$

$$B(q^{-1}) = b_0 + b_1 q^{-1} + \cdots + b_{n_b} q^{-n_b}, \tag{3}$$

$$F(q^{-1}) = 1 + f_1 q^{-1} + \cdots + f_{n_f} q^{-n_f}, \tag{4}$$

$$C(q^{-1}) = 1 + c_1 q^{-1} + \cdots + c_{n_c} q^{-n_c}, \tag{5}$$

$$D(q^{-1}) = 1 + d_1 q^{-1} + \cdots + d_{n_d} q^{-n_d}, \tag{6}$$

where n_b, n_f, n_c, and n_d are the orders of $B(q^{-1})$, $F(q^{-1})$, $C(q^{-1})$, and $D(q^{-1})$, respectively. It can be supposed that the disturbance model $H(q^{-1})$ is a stable, minimum phase, and bi-proper rational function without loss of generality, so both the zeros of $C(q^{-1})$ and $D(q^{-1})$ lie in a unit circle in the complex plane.

The present work deals with the regulatory control system with the the following feedback controller $K(q)$

$$K(q) = \frac{X(q^{-1})}{Y(q^{-1})} \tag{7}$$

$$X(q^{-1}) = x_0 + x_1 q^{-1} + \cdots + x_{n_x} q^{-n_x}, \tag{8}$$

$$Y(q^{-1}) = y_0 + y_1 q^{-1} + \cdots + y_{n_y} q^{-n_y}, \tag{9}$$

where n_x and n_y are the orders of $X(q^{-1})$ and $Y(q^{-1})$, respectively. The process input signal is calculated as $u(t) = -K(q)y(t)$. As for the controller $K(q)$, the paper considers the case where the parameters of the numerator are tuning parameters for control performance improvement. Hence, the paper represents the controller $K(q)$ in the following parametrization.

$$K(q, \boldsymbol{\rho}) = \boldsymbol{\varphi}(q)^{\mathrm{T}} \boldsymbol{\rho}, \tag{10}$$

$$\boldsymbol{\rho} = \begin{bmatrix} x_0 & x_1 & \cdots & x_{n_x} \end{bmatrix}^{\mathrm{T}}, \ \boldsymbol{\varphi}(q) = \begin{bmatrix} \frac{1}{Y(q^{-1})} & \frac{q^{-1}}{Y(q^{-1})} & \cdots & \frac{q^{-n_x}}{Y(q^{-1})} \end{bmatrix}^{\mathrm{T}}, \tag{11}$$

where $\boldsymbol{\varphi}(q)$ is a basis vector that specifies controller structures and $\boldsymbol{\rho}$ is a controller parameter vector, which will be tuned for control performance improvement.

The assumptions of the present work are as follows.

(A1) The parameters of the process $G(q)$ and the disturbance model $H(q)$ are unknown, but the order of the numerators and the denominators, n_b, n_f, n_c, and n_d are known.

(A2) The controller structure is predetermined. Namely, $\boldsymbol{\varphi}(q)$ and the dimension of parameter vector n_x is assumed to be given beforehand.

(A3) Two different controller parameters that stabilize the closed-loop system for the prescribed controller structure are attainable, and the closed-loop output implemented by the stabilized controller parameters can be collected for a certain interval.

2.2. Control objective

The control objective of the present work is the disturbance attenuation for the regulatory control systems. To this end, the present work will reduce the problem into the optimization problem minimizing the cost criterion.

$$J = J_y + \lambda J_u,\ \lambda > 0, \quad J_y = \mathrm{E}\left[\frac{1}{T}\sum_{t=1}^{T} y(t)^2\right],\ J_u = \mathrm{E}\left[\frac{1}{T}\sum_{t=1}^{T} u(t)^2\right], \tag{12}$$

where $\mathrm{E}[\cdot]$ represents the expectation operator. However, the present work employs the IFT approach for controller parameter tuning, which tunes the controller parameters every T steps by using collected process output signals at the corresponding interval. Hence, the paper updates the controller parameters in order to decrease the following cost criterion at every interval instead of the minimization of Eq.(12).

$$J^{(i)} = J_y^{(i)} + \lambda J_u^{(i)},\ \lambda > 0, \quad J_y^{(i)} = \frac{1}{T}\sum_{t=iT+1}^{(i+1)T}\left(y^{(i)}(t)\right)^2,\ J_u^{(i)} = \frac{1}{T}\sum_{t=iT+1}^{(i+1)T}\left(u^{(i)}(t)\right)^2, \tag{13}$$

where $i = 0, 1, 2, \cdots$ is trial numbers, and $y^{(i)}(t)$ and $u^{(i)}(t)$ are the process output and input signals when the controller parameters $\boldsymbol{\rho}^{(i)}$ is implemented. The objective of the paper is to provide a way how the controller parameters $\boldsymbol{\rho}^{(i)}$ are tuned so that the cost criterion Eq.(13) is decreased at every interval $iT + 1 \le t \le (i+1)T,\ i = 0, 1, 2\cdots$ under the assumption (A1), (A2), and (A3).

3. Controller parameter tuning via IFT

3.1. Gradient estimate

In order to achieve the control objective, the present work estimates the gradient of the cost criterion Eq.(13) from the collected data. The following is the gradient of the cost criterion at the i-th trial.

$$\frac{\partial J^{(i)}}{\partial \boldsymbol{\rho}^{(i)}} = \frac{\partial J_y^{(i)}}{\partial \boldsymbol{\rho}^{(i)}} + \lambda \frac{\partial J_u^{(i)}}{\partial \boldsymbol{\rho}^{(i)}}, \tag{14}$$

$$\frac{\partial J_y^{(i)}}{\partial \boldsymbol{\rho}^{(i)}} = \frac{2}{T}\sum_{t=1}^{T} y^{(i)}(t)\frac{\partial y^{(i)}(t)}{\partial \boldsymbol{\rho}^{(i)}}, \quad \frac{\partial J_u^{(i)}}{\partial \boldsymbol{\rho}^{(i)}} = \frac{2}{T}\sum_{t=1}^{T} u^{(i)}(t)\frac{\partial u^{(i)}(t)}{\partial \boldsymbol{\rho}^{(i)}} \tag{15}$$

In Eq.(15), $\frac{\partial y^{(i)}}{\partial \boldsymbol{\rho}^{(i)}}$ and $\frac{\partial u^{(i)}}{\partial \boldsymbol{\rho}^{(i)}}$ are calculated as

$$\frac{\partial y^{(i)}}{\partial \boldsymbol{\rho}^{(i)}} = -\frac{\boldsymbol{\varphi}(q)^{\mathrm{T}}}{K(q, \boldsymbol{\rho}^{(i)})} T_f(\boldsymbol{\rho}^{(i)}) y^{(i)}(t), \quad \frac{\partial u^{(i)}}{\partial \boldsymbol{\rho}^{(i)}} = -\boldsymbol{\varphi}(q)^{\mathrm{T}} S_f(\boldsymbol{\rho}^{(i)}) y^{(i)}(t), \tag{16}$$

where $S_f(\boldsymbol{\rho}^{(i)})$ is the sensitivity function, and $T_f(\boldsymbol{\rho}^{(i)})$ is the complementary sensitivity function represented as

$$S_f(\boldsymbol{\rho}^{(i)}) = \frac{1}{1 + G(q)K(q, \boldsymbol{\rho}^{(i)})}, \quad T_f(\boldsymbol{\rho}^{(i)}) = 1 - S_f(\boldsymbol{\rho}^{(i)}) \tag{17}$$

From Eq.(14), Eq.(15), and Eq.(16), it follows that if the sensitivity function $S_f(\boldsymbol{\rho}^{(i)})$ is estimated from the collected data $y^{(i)}(t), iT+1 \leq t \leq (i+1)T$, the gradient $\frac{\partial J^{(i)}}{\partial \boldsymbol{\rho}^{(i)}}$ could be estimated. Therefore, the next section shows how the sensitivity function $S_f(\boldsymbol{\rho}^{(i)})$ could be estimated.

4. Estimate of sensitivity function

As shown in the assumption (A3), it is assumed that two different controller parameters stabilize the closed-loop, and each closed-output are collected. The subsection considers the case where one set of controller parameters are the i-th trial controller parameters $\boldsymbol{\rho}^{(i)}$, and another set of controller parameters are the controller parameters $\boldsymbol{\rho}^c$ for the correction of the estimate of the sensitivity function. The proposed estimation method firstly estimates the sensitivity function using the collected data $y^{(i)}(t), iT+1 \leq t \leq (i+1)T$ implemented by the controller parameters $\boldsymbol{\rho}^{(i)}$. Then, the estimate of the sensitivity function is corrected by using the collected data $y^c(t), t_c+1 \leq t \leq t_c+T$ implemented by the controller parameters $\boldsymbol{\rho}^c$. t_c is a certain starting time instant for the data collection of $y^c(t)$. The detail procedure is as follows.

4.1. Rough estimate of sensitivity function

This stage roughly estimates the sensitivity function and the disturbance model in the case of the controller parameters $\boldsymbol{\rho}^{(i)}$. Let $\hat{\boldsymbol{\theta}}$ denote the estimated plant and disturbance model parameter vector, and let $\tilde{H}(q, \hat{\boldsymbol{\theta}})$ and $\tilde{S}_f(\hat{\boldsymbol{\theta}}, \boldsymbol{\rho}^{(i)})$ denote the estimated disturbance model and the estimated sensitivity function, respectively. The proposed method does not require the preciseness of each estimate of the sensitivity function and the disturbance model. Meanwhile, suppose that the proposed method successfully estimates the product of sensitivity function and disturbance model. The request will hold because time series analysis of the collected data $y^{(i)}(t)$ would lead to the estimates of $\tilde{H}(q, \hat{\boldsymbol{\theta}})\tilde{S}_f(\hat{\boldsymbol{\theta}}, \boldsymbol{\rho}^{(i)})$.

Additionally, Gevers (2009) proves that the prediction error method makes the request hold even in the case where the identifiability condition does not hold. Hence the estimated disturbance model and the sensitivity function can be parametrized by a bi-proper, minimum phase, stable rational function $\Delta^{(i)}(q)$.

$$H(q, \boldsymbol{\theta}^*) = \tilde{H}(q, \hat{\boldsymbol{\theta}})\Delta^{(i)}(q), \quad S_f(\boldsymbol{\theta}^*, \boldsymbol{\rho}^{(i)}) = \tilde{S}_f(\hat{\boldsymbol{\theta}}, \boldsymbol{\rho}^{(i)})\Delta^{(i)}(q)^{-1}, \tag{18}$$

where the $\boldsymbol{\theta}^*$ is the true parameters of the disturbance model. The next step determines the $\Delta^{(i)}(q)$ so that the roughly estimated sensitivity function becomes close to the true sensitivity function $S_f(\boldsymbol{\theta}^*, \boldsymbol{\rho}^{(i)})$.

4.2. Correction of the roughly estimated sensitivity function

The step uses the the collected data $y^c(t)$ implemented by the controller parameters $\boldsymbol{\rho}^c$. Let the controller and the sensitivity function using $\boldsymbol{\rho}^c$ be defined as $K(\boldsymbol{\rho}^c)$ and $S_f(\boldsymbol{\theta}^*, \boldsymbol{\rho}^c)$, respectively. By cancellation of the process model $G(q)$ between the $S_f(\boldsymbol{\theta}^*, \boldsymbol{\rho}^c)$ and $S_f(\boldsymbol{\theta}^*, \boldsymbol{\rho}^{(i)})$, the following equation can be derived.

$$S_f(\boldsymbol{\theta}^*, \boldsymbol{\rho}^c)^{-1} = 1 + \left(S_f(\boldsymbol{\theta}^*, \boldsymbol{\rho}^{(i)})^{-1} - 1\right)\frac{K(\boldsymbol{\rho}^c)}{K(\boldsymbol{\rho}^{(i)})} \tag{19}$$

Now, note that the prediction error for the collected data $y^c(t)$ can be represented as

$$\varepsilon^c(t,\boldsymbol{\theta}) = H(q,\boldsymbol{\theta}^*)^{-1} S_f(\boldsymbol{\theta}^*,\boldsymbol{\rho}^c)^{-1} y^c(t). \tag{20}$$

Hence, using Eq.(18), the prediction error Eq.(20) can be expressed as

$$\varepsilon^c(t,\hat{\boldsymbol{\theta}}) = \tilde{H}(q,\hat{\boldsymbol{\theta}})\Delta^{(i)}(q)\left(1+\left(\tilde{S}_f(\hat{\boldsymbol{\theta}})^{-1}\Delta^{(i)}(q)-1\right)\frac{K(\boldsymbol{\rho}^c)}{K(\boldsymbol{\rho}^{(i)})}\right) y^c(t). \tag{21}$$

The ideal $\Delta^{(i)}(q)$ that leads to the true sensitivity function minimizes the prediction error Eq.(21). Thus, the optimal $\hat{\Delta}^{(i)}(q)$ is obtained so that the mean square error of the prediction error Eq.(21) is minimized. Namely, $\hat{\Delta}^{(i)}(q)$ can be described as

$$\hat{\Delta}^{(i)}(q) = \arg\min_{\Delta^{(i)}(q)} \frac{1}{T}\sum_{t=1}^{T} \varepsilon^c(t,\boldsymbol{\theta})^2 \tag{22}$$

Finally, the corrected sensitivity function can be obtained as

$$\hat{S}_f(\boldsymbol{\theta}^*,\boldsymbol{\rho}^{(i)}) = \tilde{S}_f(\hat{\boldsymbol{\theta}},\boldsymbol{\rho}^{(i)})\hat{\Delta}^{(i)}(q)^{-1} \tag{23}$$

5. Numerical Example

Consider the following process model and disturbance model.

$$G(q) = \frac{0.3q^{-1}}{1-0.45q^{-1}},\ H(q) = \frac{1-0.75q^{-1}}{1-1.2q^{-1}+0.36q^{-2}} \tag{24}$$

The proposed method was applied to the numerical example. We set the time interval $T = 5000$, the variance of white noise is $\sigma_w^2 = 1$. The controller structure is $n_x = 1$ and $\varphi(q) = \frac{1}{1-0.8q^{-1}}$. Hence, the controller is parametrized as $K(q,\rho) = \varphi(q)\rho$. Note that the orders of the process model, the disturbance model, and the controller are $n_b = 1, n_f = 1, n_c = 1, n_d = 2, n_x = 0, n_y = 1$, and the identifiability condition does not hold. i.e. $\max(n_x - n_f, n_y - n_b) = 0 < 2 = n_d$. Hence, the prediction error method does not estimate the true model. In the numerical example, the first controller parameter $\rho^{(0)} = 1.8$ was used. On the other hand, as the controller parameter for the correction of the sensitivity estimates $\rho^c = 1.5$ was used. The rough estimate of the sensitivity function was calculated using the estimated process and disturbance model parameters. Since the identifiability condition does not hold, the estimated parameter remains bias error. Hence, the rough estimate also remains bias error. Fig.1 shows the comparison results. From the figure, we can see that the proposed method successfully corrects the estimate of the sensitivity function. Fig.2 and Fig.3 show how the cost criterion and controller parameter varies by applying the gradient of corrected estimate and rough estimate. From the figures, it follows that the proposed method works effectively.

6. Conclusions

The paper proposed IFT for regulatory control systems by way of estimating sensitivity functions. The numerical example showed that the proposed method can optimize the cost criterion even in the case where the identifiability condition does not hold. The further theoretical analysis remains open problems.

This work was partly supported by JSPS KAKENHI Grant Number 19K04456.

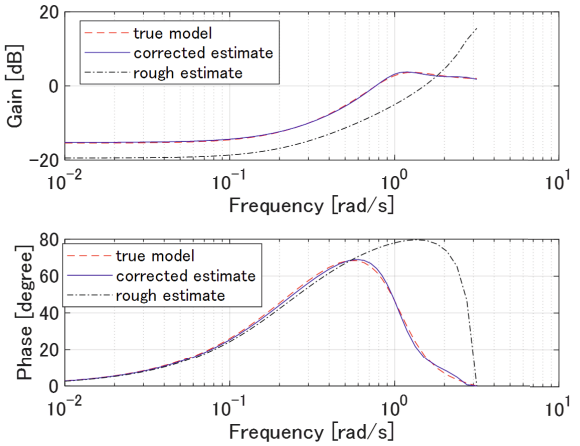


Figure 1: Sensitivity function oftrue model, corrected estimate, and rough estimate

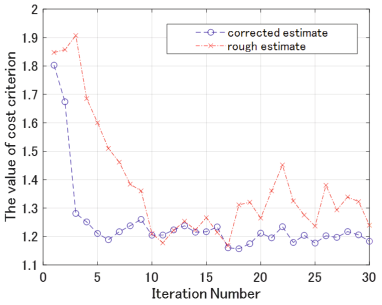


Figure 2: Cost criterion in case of corrected estimate and rough estimate

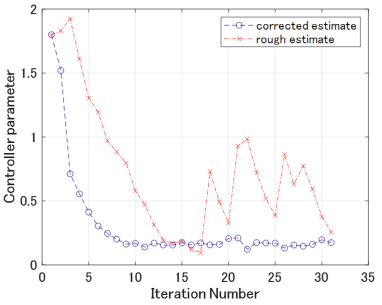


Figure 3: Controller parameter in case of corrected estimate and rough estimate

References

M. Gevers, A. S. Bazanella, X. Bombois, and L. Mišković, 2009. Identification and the Information Matrix: How to Get Just Sufficiently Rich. IEEE Transaction Automatic Control, Vol. 54, No. 12, pp.2828-2840.

H. Hjalmarsson, M. Gevers, S. Gunnarsson, and L. Lequin, 1998. Iterative feedback tuning: Theory and applications. IEEE Control System Magazine, Vol. 18, No. 4, pp.26-41.

L. C. Kammer, R. R. Bitmead, P. L. Bartlett, 2000. Direct iterative tuning via spectral analysis. Automatica, Vol. 36, pp.1301-1307.

S. Masuda, 2019. Iterative Controller Parameters Tuning Using Gradient Estimate of Variance Evaluation. In Proc. of 2019 International Conference on Advanced Mechatronic Systems (ICAMechS).

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50072-5

D-RTO as Enabler for Green Chemical Processes – Systematic Application and Challenges in Reactive Liquid Multiphase Systems

Markus Illner[a*], Volodymyr Kozachynskyi[a], Erik Esche[a], Jens-Uwe Repke[a]

[a] *Process Dynamics and Operations Group, Technische Universität Berlin, Straße des 17. Juni 135, Berlin, D-10623, GERMANY*
markus.illner@tu-berlin.de

Abstract

To date the realization of processes following the principles of Green Chemistry is still challenging due to their novelty, unknown properties of applied feedstocks and solvents, or unidentified system phenomena. Process operation and control is impeded by high system dynamics and unknown behavior. To enable early-stage realization of such "green" processes, optimal process control and especially dynamic real-time optimization (D-RTO) is advised. However, for implementation on real processes key requirements on model adequacy, measurement sufficiency, and robustness must be fulfilled. This is investigated for a reactive liquid multiphase system. Based on identified critical challenges, a tailored D-RTO framework is developed and tested using mini-plant operations. Results indicate greatly improved process operation and reaction performance.

Keywords: Microemulsions, Real-Time Optimization, Multi-Rate State Estimation, Catalysis, Mini-Plant Operation.

1. Introduction and Motivation

With an increasing need for sustainability, Green Chemistry processes considering, e.g., new synthesis paths for renewable feedstocks and the application of reactants, solvents, or additives with low environmental impact are strived for. However, large-scale industrial application of (continuous) production processes are still impeded by the complexity of novel component systems, unknown thermodynamics, and challenging process control (Ivanković, 2017). Developed methods in process systems engineering (PSE) are considered as enabler and assist process development (Mitsos et al., 2018) and operation (Müller et al., 2017; Rafiei and Ricardez-Sandoval, 2020). Their application depends on the specific challenges arising from the process and requirements for a robust implementation of suitable PSE methods, which need to be tailored based on a systematic approach. Such a procedure is outlined within this contribution for the realization of a novel "green" process concept for the hydroformylation of long-chained oily substrates in microemulsion systems and demonstrated for continuous mini-plant operations. Surfactant-based microemulsions offer beneficial properties in providing large interfacial areas for contacting aqueous catalyst solutions with oily substrates in a reactor and a thermomorphic phase separation behavior allowing for product separation and recycling of catalyst and surfactant using simple gravity settlers. However, process operation of such systems using standard automation usually fails due to the complex phase separation behavior and immeasurable states (Illner et al., 2016). Hence, a systematic analysis is conducted to identify critical operation challenges and collect sufficient information on

the system behavior. This enables the tailored development of a D-RTO framework based on multi-rate state estimation and dynamic optimization, while systematically considering key requirements such as suitable process models, measurability of plant states, and communication structures. With regard to process industry, a real-life application of D-RTO was tested using long-term mini-plant runs of up to 200 h, aiming for stable and continuous operation of the crucial phase separation and optimal reaction performance.

2. System Information and Technical Application

As example system the long-chained 1-dodecene, the surfactant Marlipal® 24/70, and an aqueous catalyst solution from the rhodium precursor (CAS: 14874-82-9) and Sulfo-XantPhos are used to form a microemulsion. The hydroformylation reaction network with the product tridecanal and possible by-products is found in (Pogrzeba et al., 2019).

2.1. Separation Behavior of the Microemulsion System

Microemulsions are mixtures of oil, water, and an amphiphile, which are characterized by complex rheology and phase separation behavior. Several phase states (1, $\underline{2}$, $\overline{2}$, 3) are possible (Figure 1), developing surfactant-rich emulsion phases and highly pure excess phases (oil or water). The separation dynamics show a distinct minimum of the separation time for the three-phase region, making it the sole feasible operation region.

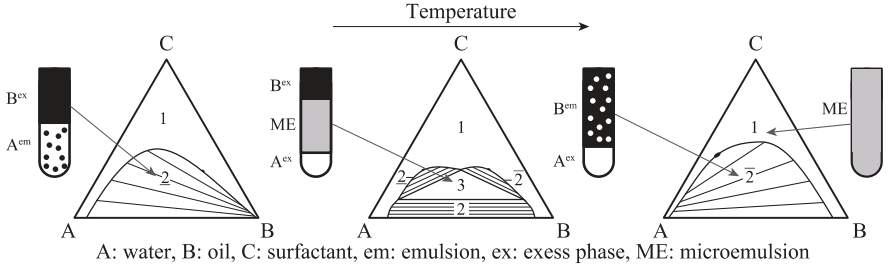


Figure 1: Schematic isothermal Gibbs triangles for microemulsion system. Phases are labeled according to the continuous liquid. Figure adapted from (Sottmann and Stubenrauch, 2009).

2.2. Mini-Plant Configuration

For testing microemulsions as green and superior reaction media, a mini-plant is operated by our group at Technische Universität Berlin. Following Figure 2, a CSTR is used for emulsification of catalyst solution, surfactant, and substrate. With addition of syngas, the reaction is conducted at 15bar and 90°C. Phase separation into up to three liquid phases is then carried out in a gravity settler holding a flow sight glass and three phase drains.

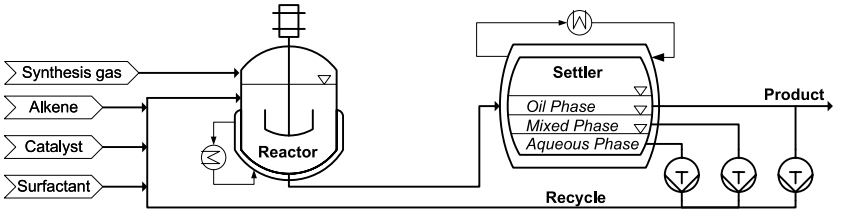


Figure 2: Simplified sketch of the mini-plant at TU Berlin consisting of reactor, settler, and recycles.

The reaction product is syphoned off from the top-most liquid phase, while the rest is recycled into the reactor. Plant automation is realized with Siemens PCS7 using an OPC

server-client structure for data communication (OPC Foundation, 2021), while reaction tracking is done via offline gas chromatography (GC). Additional information on safety measures, automation layers, and analytics can be found in (Illner, 2020).

2.3. Operational Challenges and Requirements for D-RTO Application

Successful plant operation depends on efficient control of the phase separation, which is complicated by the complex phase behavior of microemulsions. A systematic analysis of relevant influences regarding sensitivity on phase separation operation, measurability, and controllability is deployed (Illner, 2020). This reveals small and dynamically shifting operation regions (due to concentration shifts by recycling and reaction) and an unmeasurability of relevant concentrations (surfactant). To support plant start-up and to enable continuous operation with an optimal reaction performance, D-RTO is considered based on a review of existing methods and initial case studies. Here, state estimation is deployed to adapt the state of a process model to current plant measurements before calculating control trajectories by dynamic optimization. However, real-life application of D-RTO faces several theoretical and practical requirements, which have to be fulfilled for the given process (Biegler and Zavala, 2009; Bonvin and Srinivasan, 2013):
R1: A suitable model describing influences of relevant disturbances, feasible operation regions (plant optimum), and (active) process constraints is required. It needs to be twice continuously differentiable and of fast and reliable convergence behavior.
R2: State and optimality of the plant need to be quantifiable, which requires availability of specific measurements and sensitivity of the objective function regarding model states.
R3: By consequence, it needs to be verified that the problem formulation of D-RTO for the plant is robust with respect to model or measurement uncertainty.
R4: Weights in objective functions of estimators and optimizers demand tuning.
R5: A suitable communication structure between state estimation, optimization, distributed control systems (DCS), and additional analytics is required for data handling.
R6: State estimation needs to treat different sampling rates, as valuable but rare concentration measurements (gas chromatography) are mandatory to be incorporated.
R7: The D-RTO framework should allow for re-initialization after operator interactions, as unexpected events can occur, which might be out of the model's scope.

3. Model Development and D-RTO Framework

Cornerstone for the application of D-RTO on the mini-plant for the hydroformylation of 1-dodecene in microemulsions systems is a suitable dynamic process model, describing all relevant phenomena in the system (R1-R3). Key element is a model-based description of the three-phasic separation of the microemulsion, possible phase changes (constrain feasible operation region) and the derivation of a soft-sensor for otherwise inaccessible concentration information. Based thereon, a D-RTO framework is built to track and optimize plant operation using DCS, offline GC, and soft-sensor information.

3.1. Dynamic Mini-Plant Model

A dynamic model of the full mini-plant system, including relevant tanks, actuators, and available measurements is set up based on first principles and aiming for representing plant states from *fully empty* to *continuous operation*. Of special interest are the reactor and the settler model. The former considers a mechanistic reaction network for the hydroformylation reaction, which is systematically adapted to incorporate influences of the microemulsion on the reaction (Pogrzeba et al., 2019). A power-law formulation for

reaction enhancement by the surfactant concentration and a twice continuously differentiable selectivity switch (sigmoidal function) have been implemented (Illner, 2020). Regarding the settler unit, one suffers from the lack of profound thermodynamic descriptions of microemulsions (VLLLE model). For plant operation, it is however mandatory to track and describe the desired three-phase state, constrain it from shifting into undesired states, and obtain concentration information on all present phases (feedback on reaction via recycle). To achieve this, a polynomial surrogate model is derived from lab experiments, connecting experimental inputs (integral concentration x_i and temperature T) with composition x_i^{Phase} and volume fraction ϕ^{Phase} of each existing phase. This enables the formulation of a soft-sensor working on ϕ^{Phase} and T to predict the otherwise inaccessible concentrations of surfactant and water (Illner, 2020):

$$x_i^{Phase}, x_i = g(T, \phi^{Phase}) \tag{1}$$

The soft-sensor is a vital element for enabling D-RTO since it enables observation of the plant state using state-estimation, while also optimality conditions become tractable (R2). The separation model is then implemented into a settler unit model holding multiple liquid hold-ups and three fixed phase drains. Twice continuously differentiable sigmoidal functions are deployed to enable switching of the outlet flow composition according to the present phase at the respective outlet. The whole process model is available as DAE system or fully discretized via MOSAICmodeling (Esche et al., 2017).

3.2. D-RTO Framework

The structure of the developed D-RTO scheme is given in Figure 3. As a first step, moving horizon (state-)estimation (MHE) is chosen due to its superior features in handling nonlinearities and constraints (Weigert et al., 2018). As a special feature, multiple sampling rates are considered (R6). The MHE continuously provides estimates based on fast measurements (temperature, flow, level), which are updated on a second layer, whenever slow but valuable concentration measurements are available (2 to 4 h). Deployed objective functions for both layers are structured as in Eq. (2) and contain matrices P, Q, and R to weight arrival cost of estimates z, measurements y, and process noise ξ.

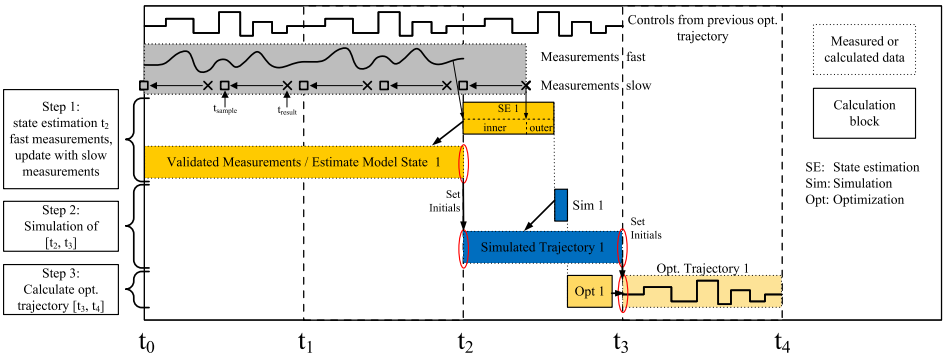


Figure 3: Graphical representation of interaction between state estimator, optimizer, and plant.

Convergence of the MHE is found to be sensitive to the choice of respective weights and online tuning is advised, if largely differing operation modes are considered (R4). Given a state estimate at t_2, a future plant state t_3 is simulated based on the available control trajectory applied on the plant. This way, feasible initials for optimization are obtained.

The dynamic optimization considers feed and recycle flows, reactor and settler temperature, and product flow as manipulables and yields trajectories of controller setpoints for a horizon of 4 h based on an economic objective function considering the product stream and penalties for catalyst loss. In parallel, MHE continuously captures the plant's state and re-optimization is triggered on larger estimated deviations. The efficient interaction of MHE, optimizer, analytics, and the plant is realized by an OPC UA-based server client structure, allowing each tool to directly communicate with the DCS.

$$\min_{z,\xi} \sum \Delta z^T P^{-1} \Delta z + \Delta y^T R^{-1} \Delta y + \Delta \xi^T Q^{-1} \Delta \xi \tag{2}$$

4. Case Study

D-RTO application was tested for long-term mini-plant runs of more than 200 h each. Pre-calculated (gPROMS) optimal trajectories ensured a stable operation throughout the critical start-up phase and provided suitable initials for the D-RTO framework. The latter was used to successfully stabilize continuous operation, as shown in Figure 4. For the shown horizons, conversion, product yield, and selectivity are stabilized on high levels, while the optimizer takes action on adaption of reaction conditions in the reactor. This resulted in further increase of the product selectivity (track optimality, R2) and successful phase separation operation in the desired three-phase state (purity of 99.5 % of oily compounds in oil phase obtained). Hence, an overall (optimal) reaction performance with a yield of 38 % and a product selectivity of 92 % was obtained. Both values are in perfect agreement with reference lab-scale experiments. Furthermore, online applicability of the D-RTO scheme is proven. For the given horizon of 4 h, feasible solutions were obtained within 72 min (e.g., horizon 1: state estimation - 735 CPU seconds; optimization - 3535 CPU seconds) given a maximum of 120 min for calculations and result implementation. However, convergence behavior significantly depends on the choice of the solver and its parameters. Here, CONOPT is preferred due to superior handling of high nonlinearities.

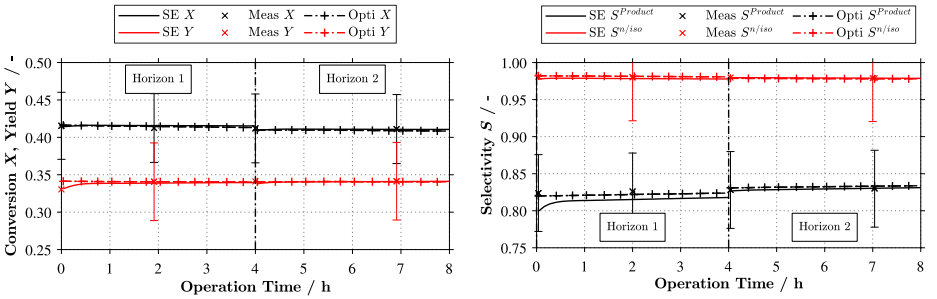


Figure 4: D-RTO application on the mini-plant. Left: Reaction conversion X and product yield Y. Right: Product selectivity $S^{Product}$ and regio-selectivity $S^{n/iso}$. Two application horizons with trajectories from state estimation (SE) and optimization (Opti) compared to measurement data (Meas).

5. Conclusion and Outlook

D-RTO is ideally suited to assist the realization of complex liquid multiphase systems as reaction media. However, key challenges for application are identified in providing suitable model structures, measurement availability and accuracy. This is demonstrated for a complex reactive multiphase system and handled with a first-time implementation of a phase separation model for microemulsions, as well as a soft-sensor for unmeasurable

model states. Based on this, a tailored D-RTO framework is successfully tested using mini-plant operation runs over 200 h each. However, future work is required on the task: how to systematically derive adequate models including phenomena relevant for the desired process, online tuning procedures for weights of state-estimator and optimizer formulations, as well as the incorporation of uncertainty.

Acknowledgements

Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) - TRR 63 “Integrierte chemische Prozesse in flüssigen Mehrphasensystemen” (Teilprojekt D2, D4) - 56091768.

References

L. T. Biegler, V. M. Zavala, 2009, Large-scale nonlinear programming using IPOPT: An integrating framework for enterprise-wide dynamic optimization, Computers & Chemical Engineering, 33, 575–582.

D. Bonvin, B. Srinivasan, 2013, On the role of the necessary conditions of optimality in structuring dynamic real-time optimization schemes, Computers & Chemical Engineering, 51, 172–180.

E. Esche, C. Hoffmann, M. Illner, D. Müller, S. Fillinger, G. Tolksdorf, H. Bonart, G. Wozny, J.-U. Repke, 2017, MOSAIC - Enabling Large-Scale Equation-Based Flow Sheet Optimization, Chemie Ingenieur Technik, 89, 620–635.

M. Illner, 2020, Rigorous analysis of reactive microemulsion systems for process design and operation, PhD Thesis, Technische Universität Berlin.

M. Illner, T. Pogrzeba, M. Schmidt, D. Müller, E. Esche, R. Schomäcker, J.-U. Repke, G. Wozny, 2016, Hydroformylation of 1-dodecene in Microemulsions: Operation and Validation of Lab Results in a Miniplant, In: Technical Transactions, 1-M (1).

A. Ivanković, 2017, Review of 12 Principles of Green Chemistry in Practice, International Journal of Sustainable and Green Energy, 6, 39.

A. Mitsos, N. Asprion, C. A. Floudas, M. Bortz, M. Baldea, D. Bonvin, A. Caspari, P. Schäfer, 2018, Challenges in process optimization for new feedstocks and energy sources, Computers & Chemical Engineering, 113, 209–221.

D. Müller, M. Illner, E. Esche, T. Pogrzeba, M. Schmidt, R. Schomäcker, L. T. Biegler, G. Wozny, J.-U. Repke, 2017, Dynamic real-time optimization under uncertainty of a hydroformylation mini-plant, Computers & Chemical Engineering, 106, 836–848.

OPC Foundation, 2021. https://opcfoundation.org (accessed October 18, 2021).

T. Pogrzeba, M. Illner, M. Schmidt, N. Milojevic, E. Esche, J.-U. Repke, R. Schomäcker, 2019, Kinetics of Hydroformylation of 1-Dodecene in Microemulsion Systems Using a Rhodium Sulfoxantphos Catalyst, Industrial & Engineering Chemistry Research, 58, 4443–4453.

M. Rafiei, L. A. Ricardez-Sandoval, 2020, New frontiers, challenges, and opportunities in integration of design and control for enterprise-wide sustainability, Computers & Chemical Engineering, 132, 106610.

T. Sottmann, C. Stubenrauch, 2009, Phase Behaviour, Interfacial Tension and Microstructure of Microemulsions, In: Microemulsions, 1–47.

J. Weigert, M. Illner, E. Esche, J.-U. Repke, Development of a State Estimation Environment for the Optimal Control of a Mini-plant for the Hydroformylation in Microemulsions. Chemical Engineering Transactions, 2018, 973–978.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50073-7

Design of PID controllers using semi-infinite programming

Evren Mert Turan[a], Rohit Kannan[b], Johannes Jäschke[a*]

[a]*Department of Chemical Engineering, Norwegian University of Science and Technology (NTNU), Trondheim, 7491, Norway*
[b]*Center for Nonlinear Studies and Theoretical Division (T-5), Los Alamos National Laboratory, Los Alamos, NM 87545 USA*
johannes.jaschke@ntnu.no

Abstract

The PID controller is widely used, and several methods have been proposed for choosing the controller parameters to achieve good performance. The controller tuning problem is set up as a semi-infinite program (SIP), with the integrated squared error (ISE) or the H_∞ norm of the frequency domain error function ($|E(s)|_\infty$) as the objective function, and H_∞ constraints for robustness and noise attenuation. Previous authors considered discrete points to enforce the H_∞ constraints, however this is an outer approximation that does not guarantee a feasible point. When a feasible point can be found, it may require multiple iterations with a finer and finer discretisation. Here, the SIP is solved using a global optimisation algorithm. Several numerical experiments show that the proposed formulation converges quickly (<10 seconds) and gives sensible controller tuning values without the need to apply expert knowledge to the tuning problem. These results suggest that this is an attractive method for automated controller tuning.

Keywords: Controller tuning; Global Optimisation; Process Dynamics and Control; Semi-infinite Programming

1. Introduction

The PID controller has found widespread use in industry and there are many methods in the literature to tune PID parameters. Typically, tuning involves a trade-off between rejecting disturbances and robustness to uncertainty (Åström and Hägglund, 2006). Finding parameters by trial and error is time-intensive, which has led to the formulation of tuning rules, e.g. the Ziegler-Nichols tuning rule and SIMC, see Åström and Hägglund (2006) for an overview. An alternative to tuning rules, is to find controller parameters by solving an optimisation problem. Optimisation-based tuning is a powerful tool, especially when system complexity, non-standard parameterisations, or requirements on performance and robustness mean that tuning rules are ill-suited (Grimholt and Skogestad, 2018; Åström and Hägglund, 2006).

Balchen (1958) presented the first "modern" formulation of the PID optimisation problem, that explicitly included a performance and robustness trade off. Since then, various authors have proposed different formulations, see e.g. Soltesz et al. (2017). Here, we place constraints on the H_∞ norm of transfer functions, i.e. the constraints should be satisfied for all considered frequencies ($w \in \Omega \subset \mathbb{R}_+$), which means there are an infinite number of constraints (Grimholt, and Skogestad, 2018; Soltesz et al. 2017).

Previous authors (Grimholt, and Skogestad, 2018; Soltesz et al. 2017) discretised the frequencies to form a finite problem, e.g., Grimholt, and Skogestad (2018) used 10 000 points. This is an outer approximation that does not guarantee a feasible point. It also raises the problem of how to select the discretisation frequencies. If we consider the PID tuning problem as one in which the constraints must be satisfied, then this means that multiple iterations with a finer discretisation or the use of expert knowledge to choose a good prior discretisation may be necessary.

In this work we use the global optimisation algorithm proposed by Djelassi and Mitsos (2017) to solve the semi-infinite PID tuning problem. This algorithm iteratively solves discretised subproblems, where at each iteration a new discretisation point is added at the frequency that results in the largest constraint violation at the incumbent solution. To facilitate the global optimisation algorithm, we use an objective function in the frequency domain. Initial results show that the proposed formulation converges in reasonably quick computation times (<10 seconds) and gives sensible controller tuning values without the need to apply expert knowledge to the tuning problem.

1.1. System

We consider the closed loop linear system in Figure 1, with disturbances at the plant input and output (d_u and d_y), and noise (n) entering the system at the measurement output. The system is represented by the following transfer functions (Åström and Hägglund, 2006):

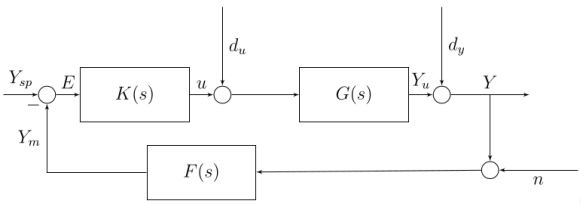


Figure 1. Block diagram of closed loop system. K(s) is the controller, G(s) is the process and F(s) is the filter.

$$S(s) = \frac{1}{1 + G(s)K(s)}, \qquad T(s) = 1 - S(s), \qquad TF(s) = T(s)F(s),$$

$$GS(s) = G(s)S(s), \qquad KS(s) = K(s)S(s), \qquad KFS(s) = K(s)F(s)S(s),$$

where s is the complex frequency ($s = iw$), and $S(s)$ and $T(s)$ are the sensitivity and complementary sensitivity functions, respectively. Here, we consider the case of pure error feedback ($F = 1$). The controller error, E, is the difference between the measured output (y) and setpoint (y_s):

$$-E(s) = y - y_s = S(s)d_y + GS(s)d_u - T(s)n. \tag{1}$$

In this work we consider PID controllers that are parameterised in the linear form:

$$K(s) = k_p + \frac{k_i}{s} + k_d s, \tag{2}$$

where k_p, k_i, and k_d are the tuning parameters. In this form the optimiser can selected a PID subtype, e.g. setting k_d to zero yields a PI controller.

1.2. Objective

We wish to pick control parameters that minimise the error after some disturbance. Various performance indices have been proposed, with the most widely used measure being the integral absolute error (IAE):

$$IAE = \int_0^\infty |e(t)|\, dt. \tag{3}$$

This formulation requires the error function in the time domain ($e(t)$). Finding the time domain error function generally involves explicit simulation or taking the inverse Laplace transform. Balchen (1958) proposed the use of a performance index in the frequency domain that approximates the IAE. The rationale behind the approximation is that $|e(t)| = e(t)\frac{|e(t)|}{e(t)}$, where if $e(t)$ is oscillatory then the fraction defines a square wave. The IAE can then be approximated by introducing a sine wave with free parameters w and a, that are chosen to maximise the integral, i.e. reduce the approximation error. This allows one to write the objective in the frequency domain:

$$IAE = \int_0^\infty |e(t)|\, dt \approx \max_{a,w} \int_0^\infty e(t)\sin(wt + a)\, dt \tag{4}$$

$$= \max_w |E(iw)| = |E(s)|_\infty = HIE,$$

where $|\cdot|_\infty$ is the H_∞ norm. For convenience, we shall refer to this as the H-infinity error (HIE). The HIE is bounded by the integral error (IE) and IAE: $IE \le HIE \le IAE$. If the system is well-dampened, then $IE \approx HIE \approx IAE$. Using Parseval's theorem, the integral squared error can be (exactly) represented in the frequency domain:

$$ISE = \int_0^\infty e(t)^2\, dt = \frac{1}{\pi}\int_0^\infty |E(iw)|^2 dw. \tag{5}$$

1.3. Robustness

We enforce robustness by constraining the maximums in the sensitivity and complementary sensitivity functions M_S and M_T, where

$$M_S = |S(iw)|_\infty, \qquad M_T = |T(iw)|_\infty.$$

The magnitude of M_S and M_T, describe the sensitivity of the system to process uncertainty or change, e.g., M_S gives the worst-case amplification of a disturbance and, on a Nyquist plot, is the distance from the loop transfer function to the point (-1,0).

Constraining the magnitude of M_S and M_T defines circles on the Nyquist plot that the loop transfer function must lie out of. A combined sensitivity constraint can be defined that covers both excluded regions. For $M = M_S = M_T$, this constraint is a circle on the Nyquist plot with centre (C, 0) and radius R given by (Åström and Hägglund, 2006):

$$C = -\frac{2M^2 - 2M + 1}{2M^2 - 2M}, \qquad R = -\frac{2M - 1}{2M^2 - 2M}.$$

1.4. Noise attenuation

It is also desirable to limit control usage due to noise. This can be performed by bounding the noise amplification ratio, $\frac{\sigma_u^2}{\sigma_n^2}$, where σ_u^2 and σ_n^2 are the variances of the control and noise respectively. Let $\phi_n(w)$ be the unknown spectral density of the (unclassified) noise, and Q be the transfer function from noise to the control signal ($Q = -KFS$, see Figure 1). The following inequality holds (Soltesz, et al. 2017):

$$\sigma_u^2 \leq |Q|_\infty^2 \sigma_n^2. \tag{6}$$

Thus, the constraint $|Q|_\infty \leq M_Q$ conservatively constrains the noise amplification ratio. This inequality can be written in the form:

$$|KF(iw)| - M_Q|1 + L(iw)| \leq 0, \qquad \forall w \in \Omega \subset \mathbb{R}_+, \tag{7}$$

where Ω defines the range of frequencies considered.

1.5. Optimisation problem

Semi-infinite programs are optimisation programs with a finite number of variables, and an infinite number of constraints. In the PID problem we have an infinite number of constraints as the constraint must hold for all considered frequencies ($w \in \Omega \subset \mathbb{R}_+$). The optimisation problem for some performance index (P_I) in the frequency domain is:

$$\min_{k_p,k_i,k_d} \eta \tag{8.a}$$

$$P_I(iw) - \eta \leq 0, \qquad \forall w \in \Omega \subset \mathbb{R}_+, \tag{8.b}$$

$$R^2 - |C - L(iw)| \leq 0, \qquad \forall w \in \Omega \subset \mathbb{R}_+, \tag{8.c}$$

$$|KF(iw)| - M_Q|1 + L(iw)| \leq 0, \qquad \forall w \in \Omega \subset \mathbb{R}_+, \tag{8.d}$$

where the constraints are explicitly parameterised by the frequency.

2. Numerical examples

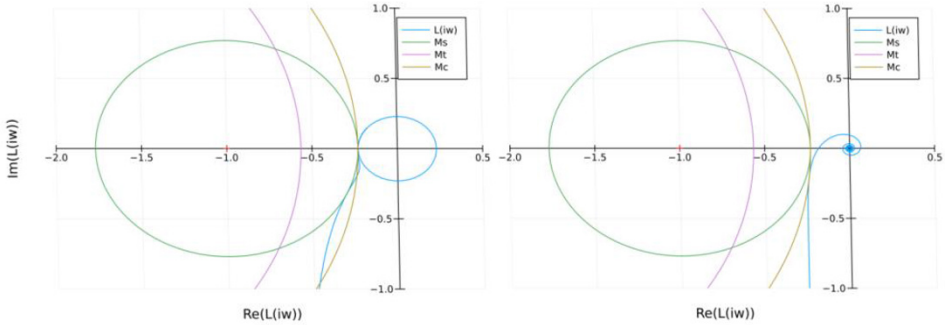
This work is coded in Julia and with the use of the global optimisation package EAGO.jl (Wilhelm and Stuber, 2020), GLPK (Makhorin, 2008), IPOPT (Wächter and Biegler, 2006), and the JuMP modelling language (Dunning, et al. 2017).

2.1. First order process with time delay

Consider the system from Grimholt and Skogestad (2018) with transfer functions:

$$G(s) = \frac{\exp(-s)}{s+1} \qquad\qquad F(s) = \frac{1}{0.001s+1}$$

To compare with the published results, we use the same weighted cost of the error from a step disturbance in u and y: $\eta = \frac{1}{1.56} HIE_{dy} + \frac{1}{1.42} HIE_{du}$. We enforce constraints on the sensitivity and complementary sensitivity with $M_S = M_T = 1.3$ and only consider frequencies w in the interval $[0.01\ 100]$. No constraint is used for the input usage.



(a) (b)

Figure 2: Nyquist plots of first order process with time delay. Left plot has constraints on maximum sensitivity and complementary sensitivity. Right plot has constraints on combined sensitivity and noise attenuation.

The optimiser finds the parameters $[0.51, 0.54, 0.23]$ in 2.6 seconds, with the Nyquist plot shown in Figure 2a. This closely matches the reported solution of $[0.52, 0.53, 0.22]$, despite the use of HIE instead of the IAE (Grimholt, and Skogestad, 2018).

For comparison, introducing a constraint on input usage ($M_Q = 1.0$) and using the combined circle constraint gives the control parameters [0.32 0.28 and 0.01], with the Nyquist plot shown in Figure 2b.

2.2. Third order process with inverse response

Consider the system process transfer functions:

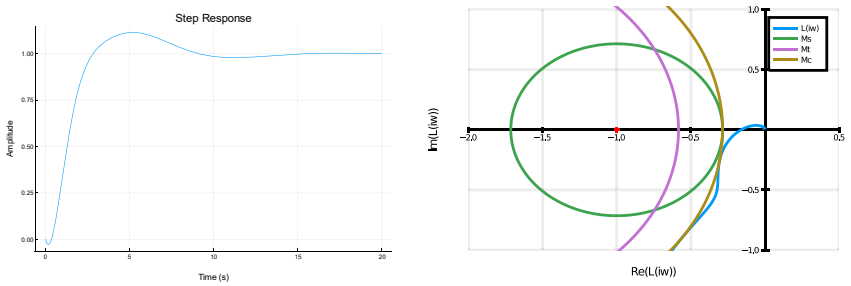
$$G(s) = \frac{1 - 0.2s}{(s+1)^3}, \qquad F(s) = 1.$$

We consider a constraint on the maximum combined sensitivity (≤ 1.3) and error function $E(s) = GS(s)d_u$. We consider frequencies in the interval $[0.01\ 100]$, and bounds on controller parameters of 0.0 and 2.0.

The optimisation is performed with HIE and ISE as the objective, giving parameters of [1.58, 1.00, 1.73] and [1.54, 1.05, 1.87] respectively, in less than 5 seconds each. The system response using the HIE parameters is shown in Figure 3.

2.3. Discussion

Despite the potential for HIE to go to zero, this did not occur in the above examples. Numerical experiments have shown that this generally occurs with oscillatory systems or large upper bounds on the control parameters and no constraint on input usage. Providing good bounds on the control parameters (e.g. by using a tuning rule) can improve the speed of optimisation. If the bounds could ensure that the control system is well-dampened, then $HIE \approx IAE$. The proposed SIP formulation can be readily extended to other linear fixed-order controllers.



(a) (b)

Figure 3: Step response and Nyquist plot for third order process with inverse response. HIE is used as the objective with no constraint on input usage.

3. Conclusions

We demonstrate that the robust PID tuning problem can be formulated and solved as a semi-infinite program, entirely in the frequency domain, using the HIE or ISE as objective functions. Robustness is enforced via H_∞ constraints on the sensitivity and complementary sensitivity functions, or an H_∞ constraint on the combined sensitivity. Control usage is restricted via an H_∞ constraint on the noise amplification ratio. On a range of systems, sensible controller parameters were found, typically in less than 10 seconds. Potential further work could include an extension to multiple output systems, or other controllers.

4. Acknowledgments

J.J. and E.M.T. acknowledge the support of the Norwegian Research Council through the AutoPRO project. R.K. acknowledges the support of the U.S. Department of Energy through the LANL/LDRD Program and the Center for Nonlinear Studies.

References

K.J. Åström, and T. Hägglund, 2006, Advanced PID control, Vol. 461, Research Triangle Park, NC: ISA-The Instrumentation, Systems, and Automation Society.

J. G. Balchen, 1958, A performance index for feedback control systems based on the Fourier transform of the control deviation, Norges tekniske vitenskapsakademi.

H. Djelassi, and A. Mitsos, 2017, A hybrid discretization algorithm with guaranteed feasibility for the global solution of semi-infinite programs, Journal of Global Optimization, 68(2), 227-253.

I. Dunning, J. Huchette and M. Lubin, 2017, JuMP: A modeling langauge for mathematical optimization, SIAM Reivew, 59, 2, 295-320.

C. Grimholt, and S. Skogestad, 2018, Optimization of fixed-order controllers using exact gradients. Journal of Process Control, 71, 130-138.

A. Makhorin, 2008, GNU linear programming kit, http://www.gnu.org/s/glpk/glpk.html

K. Soltesz, C. Grimholt, & S. Skogestad, 2017, Simultaneous design of proportional–integral–derivative controller and measurement filter by optimisation, IET Control Theory & Applications, 11(3), 341-348.

A. Wächter and L.T. Biegler, 2006, On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming, Mathematical Programming, 106.1, 25-57.

M.E. Wilhelm and M.D. Stuber, 2020, EAGO.jl: easy advanced global optimization in Julia, Optimisation Methods and Software, 1-26.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50074-9

Safe Chance Constrained Reinforcement Learning for Batch Process Optimization and Control

Max Mowbray[a], Panogiotis Petsagkourakis[b], Antonio Del Rio Chanona[b], and Dongda Zhang[a,*]

[a] *Centre for Process Integration, School of Chemical Engineering and Analytical Science, The University of Manchester, Manchester, M13 9PL, United Kingdom*
[b] *Centre for Process Systems Engineering (CPSE), Department of Chemical Engineering, Imperial College London, London, SW7 2AZ, United Kingdom*
dongda.zhang@manchester.ac.uk

Abstract

Reinforcement Learning (RL) has generated excitement within the process industries within the context of decision making under uncertainty. The primary benefit of RL is that it provides a flexible and general approach to handling systems subject to both exogenous and endogenous uncertainties. Despite this there has been little reported uptake of RL in the process industries. This is partly due to the inability to provide optimality guarantees under the model used for learning, but more importantly due to safety concerns. This has led to the development of RL algorithms in the context of 'Safe RL'. In this work, we present an algorithm that leverages the variance prediction of Gaussian process state space models to a) handle operational constraints and b) account for mismatch between the offline process model and the real online process. The algorithm is then benchmarked on an uncertain Lutein photo-production process against nonlinear model predictive control (NMPC) and several state-of-the-art Safe RL algorithms. Through the definition of key performance indicators, we quantitatively demonstrate the efficacy of the method with respect to objective performance and probabilistic constraint satisfaction.

Keywords: Safe Reinforcement Learning; Optimal Control; Dynamic Optimization; Bioprocess Operation; Machine Learning

1. Introduction

There are two main drivers for the development of (nonlinear) model based controllers within the context of the process industries. The first of those is how best to account for the expression of model uncertainties and account for the various scenarios in decision making. The second of the drivers is inspired by the conceptualisation behind the 4th Industrial Revolution and pertains to the best use of data from the ongoing process to inform control decisions. The primary approaches to these drivers within the academic community exist in the form of stochastic (as well as scenario and robust) variants of (nonlinear) model predictive control (sNMPC), learning model predictive control (L-MPC) and Reinforcement Learning (RL) (and various hybrids). In this work, we further investigate the application of RL for online optimization of (bio)chemical processes.

RL has gained traction within the community for its ability to identify an approximately optimal control policy for a Markov decision process (MDP) independently of explicit assumption regarding the underlying process dynamics or the way in which process

uncertainty is propagated. This is because RL instead *learns* a control policy (or a functionalisation of it) via trial and error of various control strategies. This process of interaction, and general policy iteration, ultimately enables the learning of an approximately optimal policy. The major benefit of RL over the other avenues discussed is that the other avenues are dependent upon identification of a closed form, finite dimensional description of the underlying system and the associated uncertainties. This is not the case in RL, and enables the flexible expression of a wide range of process uncertainties even when the model structure is nonlinear or nonsmooth. RL also has the ability to address the desire to account for process data from the real process. Conceivably, policies could either be updated online or from batch to batch using conceptually similar algorithms to those used in offline policy learning.

Despite the apparent potential of RL, there has been few reported incidents of deployment to the process industries. This is primarily because of the inability of RL to naturally handle constraints within the MDP framework and the potential for errors introduced from process-model mismatch to propagate through the control function when it is deployed to provide control on the real process. Use of a model is however absolutely required for initial policy learning due to the expense and potential safety issues arising from conducting the RL process online *from scratch.* The framework for identifying RL policies is elucidated by Figure 1.

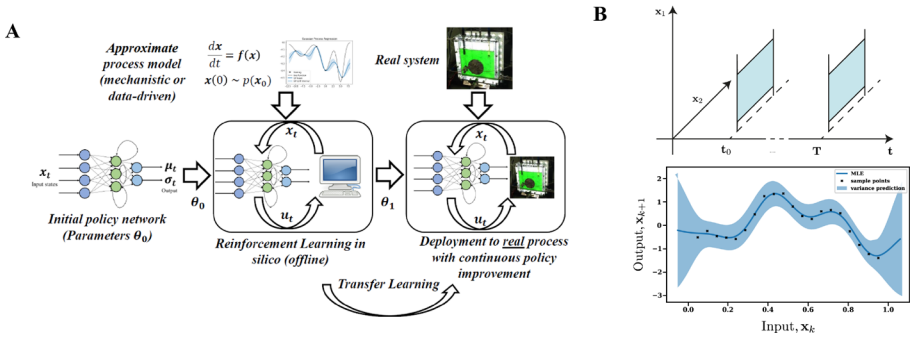


Figure 1: Demonstration of the A) Framework for RL policy training and deployment, B) Considerations for satisfaction of constraint sets and accounting for process model mismatch.

Recent works have addressed the handling of constraints via translation of the concept of constraint tightening as founded in the domain of SMPC (Pan *et al.,* 2021). Other works, based in the domain of offline or batch RL, have considered the development of algorithmic mechanisms to account for learning from a finite and stationary dataset or an approximate process model. A dominant idea in this domain is that of pessimism, which in practice is a mechanism to bias the policy away from regions of the model or data manifold characterised by high epistemic uncertainty (i.e. areas of the domain where there is little information) (Yu *et al.,* 2020). However, few works have considered the development of algorithms that consider both factors (Brunke *et al.,* 2021). In this work, we present an algorithm that considers precisely this joint problem of constraint satisfaction and mismatch (known as Safe RL), through the integration of a Gaussian process state space models as the offline process model, as well as the concepts of constraint tightening and pessimism to handle joint chance constraints and mismatch, respectively. The method is demonstrated on a fed-batch Lutein photo-production process, which is both highly nonlinear and uncertain. The benchmarks consist of state-

of-the-art safe RL algorithms and NMPC (which represents the best case deterministic method).

2. Methodology

In this work we assume that the system concerned is Markovian and expresses uncertain process dynamics, such that discrete time process evolution may be described as follows:

$$\boldsymbol{x}_{t+1} = f(\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{s}_t) \tag{1}$$

where $\boldsymbol{x} \in \mathbb{X} \subseteq \mathbb{R}^{n_x}$ are states; $\boldsymbol{u} \in \mathbb{U} \subseteq \mathbb{R}^{n_u}$ are control inputs from a given control set; $\boldsymbol{s} \in \mathbb{S} \subseteq \mathbb{R}^{n_s}$ are realisations of process uncertainty termed generally to describe various sources of uncertainty; and, $t \in \{0, \dots, T\}$ is a discrete time index within a discrete time, finite horizon. We would like to solve the following chance constrained problem:

$$P(\pi_c) = \begin{cases} \max\limits_{\pi_c} \mathbb{E}_{\pi_c}\left[\sum\limits_{t=0}^{T-1} R_{t+1}\right] \\ \boldsymbol{x}_0 \sim p(\boldsymbol{x}_0) \\ \boldsymbol{x}_{t+1} = f(\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{s}_t) \\ \boldsymbol{u}_t \sim \pi_c(\boldsymbol{u}_t|\boldsymbol{x}_t) \\ \boldsymbol{u}_t \in \mathbb{U}_t \\ P\left(\bigcap\limits_{t=0}^{T} \boldsymbol{x}_t \in \overline{\mathbb{X}}_t\right) \geq 1 - \alpha \\ \forall\, t \in \{0, \dots, T\} \end{cases} \tag{2}$$

where $\pi_c(\boldsymbol{u}_t|\boldsymbol{x}_t)$ defines a conditional probability density function, that provides a distribution over controls given observation of state; $\overline{\mathbb{X}}_t = \left\{\boldsymbol{x}_t \in \mathbb{R}^{n_x} : A_j^T \boldsymbol{x}_t - b_j \leq 0, \forall\, j \in \{1, \dots, n_g\}\right\}$ is the set of states that satisfy a given affine ($A_j \in \mathbb{R}^{n_x}$ and $b_j \in \mathbb{R}$) constraint set at a given time index; $\alpha = (0,1]$ is the probability allowed for violation of the constraint set for all time indices; $p(\boldsymbol{x}_0)$ defines the initial state distribution; and, $R_{t+1} \in \mathbb{R}$ defines a scalar cost, provided by a function, $R: \mathbb{X} \times \mathbb{U} \times \mathbb{X} \to \mathbb{R}$, that ranks process evolution with respect to control objectives.

Conventional RL algorithms identify a policy, $\pi^* = argmax_\pi \mathbb{E}_\pi[\sum_{t=0}^{T-1} R_{t+1}]$ and do not naturally handle constraints. In this work, we are concerned with identifying a policy, π_c, that also satisfies a joint constraint set, with a given probability $1 - \alpha$, that may be defined by the implementation.

To handle this, we propose to translate the concept of constraint tightening (Valdez-Navarro and Ricardez-Sandoval, 2019) to tighten the constraint set, such that we can reformulate our probabilistic expressions into deterministic surrogates. The tightened constraint set is expressed as $\bar{\mathbb{X}}_t = \left\{\boldsymbol{x}_t \in \mathbb{R}^{n_x} : A_j \bar{\boldsymbol{x}}_t + \varepsilon_{j,t} - b_j \leq 0,\ \forall\, j = \{1, \dots, n_g\}\right\}$, where $\varepsilon_{j,t} \in \mathbb{R}$ is a constraint tightening mechanism (known as a *backoff*) specific to both time index and constraint, and $\bar{\boldsymbol{x}} \in \mathbb{R}^{n_x}$ is the nominal process state. The idea of the backoff is to essentially back the nominal state away from the constraint boundary to allow for process variation. Explicit, closed form expressions for the backoffs may be derived from the Cantelli-Chebyshev inequality. Together, with Boole's inequality, this enables us to decompose the probability of constraint satisfaction across the various constraints, and identify the backoffs required by the following:

$$\varepsilon_{j,t} = \sqrt{\frac{1-\iota_j}{\iota_j}}\sqrt{A_j^T \Sigma[\boldsymbol{x}_t] A_j} \tag{3}$$

where $\Sigma[\boldsymbol{x}_t]$ is the finite variance of the system state at time index t and $\iota_j = \alpha/n_g$ is the probability with which one may allow for violation of constraint j. This enables formalisation of the tightened constraint set. Solving for this enables the satisfaction of the joint chance constraints with desired probability. In this work, we parameterise both $\bar{\boldsymbol{x}}$ and $\Sigma[\boldsymbol{x}]$ of the state by a Gaussian process state space model. Expressions for both may then be obtained in closed form, providing deterministic expressions of the constraint set that can be incorporated into a penalty function. Additionally, we incorporate $\Sigma[\boldsymbol{x}]$ to penalise exploitation of the regions of the state space characterised by high epistemic uncertainty (this is the mechanism for pessimism). In short, what we propose here is to identify an optimal constrained policy that maximises a penalty function in expectation:

$$\pi_c{}^* = argmax_\pi \mathbb{E}_\pi \left[\sum_{t=0}^{T-1} R_{t+1} - tr(\zeta \Sigma[\boldsymbol{x}_{t+1}]) - \kappa \| [A^T \bar{\boldsymbol{x}}_{t+1} + \boldsymbol{\epsilon}_{t+1} - \boldsymbol{b}]^- \|_2 \right] \tag{4}$$

where $\zeta \in \mathbb{R}^{n_x \times n_x}$ is a diagonal matrix; $\kappa \in \mathbb{R}$ is a large real value; $A = \left[A_1, \dots, A_{n_g}\right]$; $\boldsymbol{b} = \left[b_1, \dots, b_{n_g}\right]^T$ and $\boldsymbol{\epsilon}_t = \left[\varepsilon_{1,t}, \dots, \varepsilon_{n_g,t}\right]^T$. In practice, the Cantelli-Chebyshev backoffs identified are typically conservative, so we propose to tune them by a set of multipliers $\boldsymbol{\xi} \in \mathbb{R}^{n_g}$. This problem is an expensive black box optimization and so we deploy use of Bayesian optimization (BO) to identify the multipliers. At a given iterate i of optimization the backoffs are defined, $\Xi^i = \{\boldsymbol{\xi}^i \boldsymbol{\epsilon}_t\}_{t=1:T}$. The objective for BO, J_{BO}, is formalised:

$$\boldsymbol{\xi}^* = argmax_{\boldsymbol{\xi}} J_{BO} = \left(G_{\pi_c^*} - \beta \sigma_{\pi_c^*}\right) \exp(-cU) \tag{5}$$

where $c \in \mathbb{R}$, $\beta \in \mathbb{R}$ are constants; $U = \left(F_{LB} - (1-\alpha)\right)^2$, where F_{LB} is a robust statistic evaluating the probability of joint chance constraint satisfaction; $\sigma_{\pi_c^*} = \Sigma_{\pi_c^*}[\sum_{t=0}^{T-1} R_{t+1}]$ is the variance of the returns with respect to the process objective; $G_{\pi_c^*} = \mathbb{E}_{\pi_c^*}[\sum_{t=0}^{T-1} R_{t+1}]$ is the expected returns with respect to the process objective. The statistic, F_{LB}, is a statistically robust sample approximation of probabilistic constraint satisfaction.

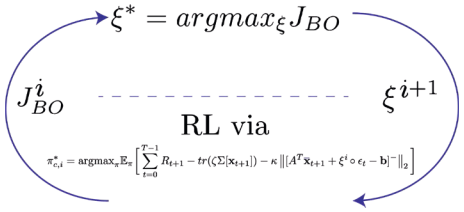


Figure 2: Figurative description of the algorithm proposed. In practice, we propose the use of Gaussian process based Bayesian Optimization.

3. Case Study

3.1. A Fed-batch Lutein Photo-production process

To demonstrate the methodology, we propose the use of a computational study involving a fed-batch Lutein photo-production process, which was first detailed in Del Rio-Chanona

et al. (2017) and described by a set of nonlinear ordinary differential equations (ODEs). We direct the interested reader to that paper for more information on the underlying system and parameter values.

3.2. Case Study Design and Benchmarks

Additional to the parameter values detailed by the original work, we assume the presence of 5% parametric uncertainty on the point estimates provided. We then leverage the availability of the uncertain, mechanistic model (that is equivalent to Eq. 1) and conceptualise that it represents the *real* system. Description of the case study follows: a) generate a dataset (that in practice could be available from e.g. design of experiments) by sampling the uncertain mechanistic model with space filling control trajectories, b) deploy the methodology described by building a GP state space model with the dataset and identify a policy through the framework detailed by Figure 1, c) deploy the policy identified to optimize the *real* uncertain process (model) and d) benchmark the performance against NMPC, first order constrained optimisation in policy space (FOCOPS) (Zhang *et al.,*2021), the model-based offline policy optimization (MOPO) algorithm (Yu *et al.,* 2020), and the conservative offline model based policy optimization (COMBO) algorithm (Yu *et al.,* 2021). The first two benchmarks represent the best case deterministic method, and a state-of-the-art constrained RL method, respectively. The latter two represent state-of-the-art offline (batch) RL methods that are designed to account for mismatch. In all RL benchmarks, constraints were handled by incorporating deterministic expressions for the original constraint sets into a penalty function (i.e. they were treated as hard constraints) within the GP state space model.

3.3. Process Constraints, Objective and Key Performance Indicators

The objective of process operation is productivity maximisation. The control inputs, $\boldsymbol{u} = [I_0, F_{in}]$, are the incident light intensity to the reactor, and the nitrate inflow rate and they are bounded. The path constraint set is defined by Eq. 6. The set defined represents constraints on the maximum concentration of biomass c_x (g/L), the minimum nitrate concentration and the ratio between biomass and Lutein concentration. The desired probability of constraint violation, $\alpha = 0.01$. The process objective is provided by Eq. 7.

$$A = \begin{bmatrix} 1 & 0 & -1.67 \\ 0 & -1 \times 10^{-3} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 2.6 \\ 0.15 \\ 0 \end{bmatrix} \tag{6}$$

$$R_{t+1} = \begin{cases} \boldsymbol{d}^T \boldsymbol{x}_{t+1} - \Delta \boldsymbol{u}_t^T \boldsymbol{C} \Delta \boldsymbol{u}_t & if\ t = T-1 \\ -\Delta \boldsymbol{u}_t^T \boldsymbol{C} \Delta \boldsymbol{u}_t & else \end{cases} \tag{7}$$

where $T = 6$, $\boldsymbol{d} = [0, -0.001, 4]^T$, $\Delta \boldsymbol{u}_t = \boldsymbol{u}_t - \boldsymbol{u}_{t-1}$, $\boldsymbol{x} = [c_x, c_N, c_L]^T$ and $C = diag([0.16, 8.1 \times 10^{-5}])$. Key performance indicators include F_{LB}, F_{SA} and $G_{\pi_c^*}$.

4. Results and Discussion

The results from validation of the control policies identified by each respective method over 500 Monte Carlo simulations on the *real* uncertain process model are outlined by Table 1.

Table 1: Results from benchmarks and method proposed (SCCPO) in online optimization of the real uncertain process

Method	Sample approx. of probability of constraint satisfaction, F_{SA}	Robust approx. of probability of constraint satisfaction, F_{LB}	Fulfilment of process operational objective, $G_{\pi_c^*}$.
NMPC	0.12	0.148	11.58 +/- 4.07
MOPO	1.0	1.0	10.98 +/- 0.072
COMBO	1.0	1.0	10.69 +/- 0.070
FOCOPS	1.0	1.0	13.11 +/- 0.090
SCCPO	1.0	1.0	**14.17 +/- 0.095**

The results demonstrate the ability of the method proposed to handle both constraints and mismatch. This is especially reinforced by the relative performance to NMPC, where the model is exactly the same as the uncertain real process and the only difference that exists is the presence of parametric uncertainty in the real process. All RL benchmark methods (MOPO, COMBO, FOCOPS) handle constraints with the desired probability. It is thought that this arises due to the implementation of a backoff as introduced through a) the pessimism term, $tr(\zeta\Sigma[x_{t+1}])$, present in MOPO and FOCOPS, and b) the nature of the conservative mechanisms as present in COMBO. However, it should be noted that the action of these mechanisms is not specific to any given constraint, which may go some way to explaining the conservative control performance.

5. Conclusions

In this work, we have presented an algorithm that handles both operational constraints and process model-mismatch for the deployment of RL policies for the online optimization of uncertain, nonlinear fed-batch process systems. The algorithm has been benchmarked against best case deterministic methods in the form of NMPC and state-of-the-art safe and offline RL methods. The performance was demonstrated to be competitive if not advantageous relative to the benchmarks proposed.

References

L. Brunke, M. Greeff, A.W. Hall, Z. Yuan, S. Zhou, J. Panerati, A.P. Schoellig, Safe Learning in Robotics: From Learning-Based Control to Safe Reinforcement Learning, arXiv:2108.06266 (2021)

E. Pan, P. Petsagkourakis, M. Mowbray, D. Zhang, E. A. del Rio-Chanona, Constrained Model-Free Reinforcement Learning for Process Optimization, arXiv preprint arXiv: 2011.07925 (2021)

Y. I. Valdez-Navarro, L. A. Ricardez-Sandoval, A novel back-off algorithm for integration of scheduling and control of batch processes under uncertainty, Industrial & Engineering Chemistry Research 58 (48) (2019)

T. Yu, G. Thomas, L. Yu, S. Ermon, J. Zou, S. Levine, C. Finn, T. Ma, Mopo: Model-based offline policy optimization, . arXiv:2005.13239 (2020)

E.A. Del Rio-Chanona, N. Ahmed, D. Zhang, Y. Lu, K. Jing, Kinetic modeling and process analysis for Desmodesmus sp. lutein photo-production,

Y. Zhang, Q. Vuong, K. W. Ross, First Order Constrained Optimization in Policy Space, arXiv preprint arXiv: 2002.06506 (2020)

T. Yu, A. Kumar, R. Rafailov, A. Rajeswaran, S. Levine, C. Finn, COMBO: Conservative Offline Model-Based Policy Optimization, arXiv:2102.08363 (2021)

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50075-0

Combining Machine Learning with Mixed Integer Linear Programming in Solving Complex Scheduling Problems

Iiro Harjunkoski[a,b*], and Teemu Ikonen[a]

[a]*Department of Chemical and Metallurgical Engineering, Aalto University, 00076 Finland*
[b]*Hitachi Energy Research, 68309 Mannheim, Germany*
iiro.harjunkoski@aalto.fi / iiro.harjunkoski@hitachienergy.com

Abstract

With the increasing digitalization of industrial production processes and the quest for maximizing the synergies through more integrated operations, there is an increasing need also to automatize the decision making. In terms of scheduling, problems are becoming larger and need to consider more aspects making both the modeling and the solution of the resulting problems cumbersome. Suitable methods to deal with these problems include, e.g., simplifying the problem as necessary to speed up the optimization (i.e., balancing the optimality and solution speed where possible), using heuristics to support faster solution, deploying simulation tools to predict the values of most complex variables, using decomposition methods to divide the problem into smaller subproblems, and a rich mixture of all of the above. This paper discusses various approaches to support optimization by using machine learning and related challenges in implementing them.

Keywords: scheduling, machine learning, hybrid models, efficiency

1. Introduction

The topic of combining machine learning (ML) with optimization, foremost the combinatorially complex mixed integer linear programming (MILP) problems, has many facets and has recently received increasing attention in the literature. In its simplest form, the focus can lie either on modeling or on the solution procedure. In modeling, the main target is to reduce the workload of a modeler – or the large, often experimental, efforts in parameter tuning that is coupled with complex analysis. In solving complex problem related to scheduling, the main target is either to reduce the size of the search space or provide more guidance (similar to strong branching) while traveling through the search tree. An inspiring and insightful view into artificial intelligence (AI) and chemical engineering is provided by Venkatasubramanian (2019), who also highlights the challenges of reasoning and explainability of ML-based decisions.

Production scheduling problems are often modeled and solved as MILP problems, at least partially, as this provides a framework to systematically embed problem-specific constraints and facilitates solving the models by state-of-the-art solvers. Theoretically, scheduling problems are NP-complete (Garey et al., 1979) for generic approaches in finding the shortest-length (makespan) schedule, as well as minimizing the mean-flow time. When focusing on MILP approaches (Méndez et al., 2006), the main complexities are related to the combinatorial space arising from the large number of binary decision variables that stem both from the assignment (yes/no) and sequencing (before/after)

decisions involved in scheduling. Often, the sequencing decisions are more complex especially in continuous-time approaches due to the typically deployed big-M constraints that in general show poor relaxation characteristics.

As scheduling problems typically lie in between the control (local) and business (enterprise) layers, they need to adopt some characteristics from both worlds. Lower-level decisions need to ensure the feasibility of operations by taking into account sometimes complex equipment-related choices – or even such a simple task to avoid simultaneous overlaps of multiple jobs in a producing equipment. Business decisions may also be complex to model in practice due to the multitude of conditions that may impact the scheduling. One example can be found in scheduling of power generation units, where both the dispatching of electricity and providing electricity reserves to account for sudden losses of generation must be procured with different pricing schemes. In other scheduling problems, complex modeling tasks may apply when estimating highly non-linear processes, equipment degradation, taxation, or inventory policies. Thus, there are potential challenges both in the modeling and solution of scheduling problems that may be improved by using ML-based methods.

Figure 1 shows a rough overview of the procedure where the original problem data is first transformed into a scheduling model using both domain knowledge (here we assume also including the standard well-known model structures), as well as process data and ML to create data-driven model components. Second, the resulting model is passed onto the solution procedure, which typically deploys mathematical programming (e.g., Branch & Bound or Branch & Cut) or evolutionary methods (e.g., genetic algorithms). The solution procedure can be supported either by heuristics to speed up the solution process, leading to a local optimum or a good starting point, or ML-based decision support. In this paper, we focus on the latter. In the following, we discuss both the modeling (Section 2) and solving aspects (Section 3) of using ML to enhance scheduling.

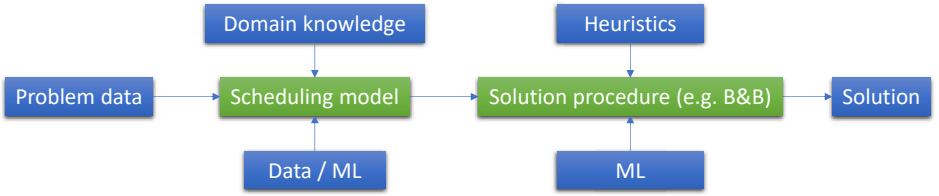


Figure 1. Main options for using ML in formulating and solving large MILP problems

2. Modeling

For the efficient modeling of complex constraints based on e.g., experimental or simulated data, machine learning based tools have been proposed, such as ALAMO (Wilson and Sahinidis, 2017) to derive automatically generated constraints, where the accuracy between the observed points and the resulting model equations is optimally balanced with the resulting numerical complexity. As opposed to typical parameter estimation schemes, the ALAMO approach also selects the structural constraint components. This approach can be very efficient when the order of the resulting equations is low. However, good linear representations, necessary for MILP models, may be difficult to identify. Other aspects to consider are whether there is enough data available for a reliable result or whether the existing equations can directly be mathematically re-formulated or approximated through piece-wise linear functions or over/under-

estimators. Also, if the domain is well-defined and scope relatively narrow, alternative self-written codes may well support the modeling by benefiting from the existing ML toolkits. The main idea of ML-supported modeling is to reduce the sometimes significant and error-prone engineering efforts that can occupy highly-trained personnel on routine tasks, which could be better handled if automated. The challenge, on the other hand, is to identify the use cases where domain expertise can be reliably generated by a system that mainly builds on already existing data – leaving very little room for creative approaches, should a new and unexpected operational situation take place. One option could be whether the ML tool would be able to identify its own capability limitations and notify an operator/planner about a situation that is out of its designed scope.

Already before the increased attention on ML, the concept of surrogate models has been utilized also for scheduling (Bhosekar and Ierapetritou, 2018) and recently also methodologies have been used to effectively utilize the data available for more accurate modeling (Shi and You, 2018). The number of related activities is increasing, indicating that there are still significant opportunities ahead on this research avenue.

3. Solution efficiency

Apart from the modeling step, probably even more efforts have been made in expediting the solution procedures of large and complex problems. Such approaches comprise methods to reduce the combinatorial complexity; in scheduling applications this typically leads to reducing the number of binary variables through pre-fixing the values of some of these or adding constraints that relate the values of different variables to each other (falling into the category of tightening constraints). When using rigorous MILP models for scheduling, this can also be achieved through analyzing the LP-relaxation, based on which one can derive e.g. variable lower bounds for the remaining MILP problem (Castro et al., 2020). This method can be performed per instance and can provide a significant reduction in solution time, as efficient cuts are added to the original MILP model. Other decomposition approaches have been suggested e.g. in Terrazas-Moreno and Grossmann (2011). If, on the other hand, there are sufficiently large data sets available, one can also use AI (Venkatasubramanian, 2019), or more specifically, ML to efficiently single out some decision options that either should be always selected (i.e., fixing related binary variables to one) or systematically excluded (i.e., fixing binary variables to zero).

Here we cannot dive into individual case studies but some related results are reported in Harjunkoski et al. (2020), where several options to speed up the solution of large-scale MILP problems are discussed. If sufficient data is available, e.g. in the paper industry case described by Mostafaei et al. (2020), one can use machine learning to dynamically generate more accurate (up-to-date) scheduling parameters such as change-over times for a grade-change in papermaking using machine learning over a large set of production data. Having access to multiple years of operational data also poses the challenge of how to balance between the recent and past data instances (often referred to as the *forgetting factor*), in order to ensure that the estimates remain accurate and are responsive for possible changes in the process. Apart from the scheduling parameters, one can also use the above data to eliminate product sequences that do not appear in the process history – these might be either unpractical or costly and have therefore not been selected by skilled operators. The approach results in a fraction of possible sequences, making intractable problems solvable. In short, instead of selecting from all possible sequences, the optimization focuses only on those that have been applied in the past.

In the presence of equipment degradation, it may be complex to decide whether a unit needs to be maintained or replaced – decisions that are often represented by binary variables. Having good statistical data at hand, it may be possible to either enforce, as well as exclude possible maintenance actions for components that with a very high certainty can be expected to fail or not to fail before the next planned maintenance stop (Ikonen et al., 2020a). Thus, instead of using a single instance as a basis of knowledge, as is done in the case of analyzing the problem through its LP-relaxation or specific decomposition schemes, the use of advanced AI/ML-techniques for creating more generic decision-making patterns can be a very efficient way of reducing the combinatorial complexity of the problem.

The above approach results in models where the variable bounds are tightened or fixed before the solution of the problem. There are many other avenues reported in the literature. One of the very interesting approaches focuses on trying to deploy several approaches to collect more understanding of the problem and the structure of its search tree in order to reduce the complexity and in a way learn to solve a problem in a better way (Xavier et al., 2019). This is especially promising for problems that are solved over-and-over-again with little variations and also involves implementing machine learning into the Branch & Bound (or Branch & Cut) tree search. There exists already a platform ECOLE (Provoust et al., 2020) for supporting the research in doing this. Commercial solver providers are already exploring and equipped with similar capabilities, e.g. the optimization provider Gurobi has done quite a bit of investigations on this and implemented some ideas (see reference on https://www.gurobi.com). Also, IBM CPLEX uses ML to automatically decide whether to use some algorithmic choices in solving quadratic optimization problems (reference https://developer.ibm.com). It is certain that this research challenge is going to be addressed by many researchers in the next years.

Recently, some research groups have investigated how to enhance process scheduling by reinforcement learning (Sutton and Barto, 2018). Hubbs et al. (2020a) examines the use of deep reinforcement learning in process scheduling. Another option is to deploy reinforcement learning at a higher level. Ikonen et al. (2020b) propose a framework where a reinforcement learning (RL) agent is trained to decide the timing of rescheduling procedures, select the scheduling algorithms to be used (e.g., MIP or heuristic) and estimate the time budget needed for the optimization. The Python-based library OR-GYM (Hubbs et al., 2020b) provides test environments for developing reinforcement learning algorithms to address operations research problems.

4. Main Challenges

As the problem space is very broad, the challenges are manifold so here we highlight only a few:

1. **Data.** Using machine learning in supporting optimization requires the presence of sufficient and high-quality data. In any of the above approaches a successful application must get hold on sufficient – often business critical – data, which often limits the work to company-internal exercises, hindering experts from the outside to get involved. Because of this, there are unfortunately almost no open data-sets available from the process industries. Possible tampering with the data may be difficult to detect and without dedicated domain experts, the interpretation of data may be close to impossible.

2. **Automated modeling.** Even if there exist well-documented model alternatives the modeling of scheduling problems is often almost an "art", while combining physical/business/operational requirements in a way that meets the company policies. It should be well considered which part of this can be automated without compromising any of the company objectives.
3. **Algorithmic complexity.** As mathematical modeling alone is often complex, adding another layer of ML makes it even more difficult to manage without good support. Today, there are limited number of experts that can master both disciplines and targeted training efforts are needed to ensure sufficient in-house knowledge.
4. **Deployability.** So far, purely ML-based approaches are able to handle mainly "routine tasks", where a human operator would need to react fast and perform a limited number of actions. How to widen the scope of ML is still open but a balanced mixture of e.g. ML and MILP could be one way to increase this capability.
5. **Balancing multiple objectives.** The strength of mathematical optimization is to be able to mix various objective components and balance them through given weights also in highly varying conditions. As ML alone may not be able to do this, the combination with optimization technologies can be very useful.

A relatively recent approach of Physics-informed neural networks (PINNs) reported e.g. in Raissi et al. (2019) provides an interesting approach of combining the use of data for learning with existing models of physical and biological systems and allows to benefit from the vast amount of existing prior knowledge that are not utilized in most common ML-methods. Encoding such structured information into a learning algorithm results in amplifying the information content of the data that the algorithm sees, enabling it to quickly steer itself towards the right solution and generalize well even when only a few training examples are available.

5. Conclusions

The topic of combining machine learning to support or complement mixed integer linear programming in solving scheduling problems is extremely interesting and has many possible avenues for successful applications. It is, however, important to have realistic expectations and start by working on tangible problems, where the benefits can be observed and quantified. Simultaneously, it would be important to create sufficiently large, shared datasets for development and comparison of different approaches. Through continued research, new discoveries can without doubt be made and some form of collaborative approaches will likely very soon become the future standard for the efficient and successful solutions of complex scheduling problems, once the major challenges have been overcome.

References

Bhosekar, A., and Ierapetritou, M. (2018). Advances in surrogate based modeling, feasibility analysis, and optimization: A review. Computers and Chemical Engineering, 108, 250-267

Castro, P. M., Dalle Ave, G., Engell, S., Grossmann, I. E., and Harjunkoski, I. (2020). Industrial demand side management of a steel plant considering alternative power modes and electrode replacement. Industrial and Engineering Chemistry Research, 59(30), 13642-13656

Harjunkoski, I., Ikonen, T., Mostafaei, H., Deneke, T., and Heljanko, K. (2020). Synergistic and intelligent process optimization: First results and open challenges. Industrial and Engineering Chemistry Research, 59(38), 16684-16694

https://developer.ibm.com/docloud/blog/2019/11/28/using-machine-learning-in-cplex-12-10/

https://www.gurobi.com/resource/integrating-machine-learning-with-mathematical-optimization-resource-matching/

Hubbs, C. D., Li, C., Sahinidis, N. V., Grossmann, I. E., & Wassick, J. M. (2020a). A deep reinforcement learning approach for chemical production scheduling. Computers & Chemical Engineering, 141, 106982

Hubbs, C.D., Perez, H.D., Sarwar, O., Sahinidis, N.V., Grossmann, I.E., & Wassick, J.M. (2020b). OR-Gym: A Reinforcement Learning Library for Operations Research Problem. ArXiv, abs/2008.06319.

Ikonen, T. J., Mostafaei, H., Ye, Y., Bernal, D. E., Grossmann, I. E., & Harjunkoski, I. (2020a). Large-scale selective maintenance optimization using bathtub-shaped failure rates. Computers and Chemical Engineering, 139, 106876

Ikonen, T. J., Heljanko, K., and Harjunkoski, I. (2020b). Reinforcement learning of adaptive online rescheduling timing and computing time allocation. Computers and Chemical Engineering, 141, 106994

Garey, M. R., Johnson, D. S., & Sethi, R. (1976). Complexity of Flowshop and Jobshop Scheduling. Mathematics of Operations Research, 1(2), 117-129

Méndez, C. A., Cerdá, J., Grossmann, I. E., Harjunkoski, I., & Fahl, M. (2006). State-of-the-art review of optimization methods for short-term scheduling of batch processes. Computers and Chemical Engineering, 30(6-7), 913-946

Mostafaei, H., Ikonen, T., Kramb, J., Deneke, T., Heljanko, K., and Harjunkoski, I. (2020). Data-driven approach to grade change scheduling optimization in a paper machine. Industrial and Engineering Chemistry Research, 59(17), 8281-8294

Prouvost, Antoine et al. “Ecole: A Gym-like Library for Machine Learning in Combinatorial Optimization Solvers.” ArXiv abs/2011.06069 (2020)

Raissi, M., Perdikaris, P., and Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 378, 686-707

Shi, H., and You, F. (2015). A novel adaptive surrogate modeling-based algorithm for simultaneous optimization of sequential batch process scheduling and dynamic operations. AIChE Journal, 61(12), 4191-4209

Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. MIT press.

Terrazas-Moreno, S., and Grossmann, I. E. (2011). A multiscale decomposition method for the optimal planning and scheduling of multi-site continuous multiproduct plants. Chemical Engineering Science, 66(19), 4307-4318

Venkatasubramanian, V. (2019). The promise of artificial intelligence in chemical engineering: Is it here, finally? AIChE Journal, 65(2), 466-478

Wilson, Z. T., and Sahinidis, N. V. (2017). The ALAMO approach to machine learning. Computers and Chemical Engineering, 106, 785-795

Xavier, A.S., Qiu, F., and Ahmed, S. (2019). Learning to Solve Large-Scale Security-Constrained Unit Commitment Problems, arXiv:1902.01697 [math.OC]

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50076-2

Knowledge-guided Hybrid Approach for Scheduling Multipurpose Batch Plants

Dan Li[a], Dongda Zhang[a], Nan Zhang[a], Liping Zhang[b], and Jie Li[a]*

[a] *Centre for Process Integration, Department of Chemical Engineering, The University of Manchester, Manchester M13 9PL, UK*
[b] *Department of Industrial Engineering, Wuhan University of Science and Technology, Wuhan, Hubei, 430081 P. R. China*
jie.li-2@machester.ac.uk

Abstract

In this work, a novel hybrid algorithm integrating knowledge-guided GA and sequence-based mixed-integer linear programming (MILP) model is proposed for scheduling of industrial multipurpose batch plants. The computational results demonstrate that the proposed hybrid algorithm can generate the optimal solutions within 5 minutes for all tested industrial-scale examples. It can generate the same or better solutions using less computational effort than the existing methods.

Keywords: Scheduling; Multipurpose batch plants; Hybrid algorithm; Genetic algorithm

1. Introduction

Multipurpose batch plants have shared facilities and product-specific processing steps, whose flexibility provides industrial application prospect but poses challenges on scheduling. A plethora of mixed-integer linear programming (MILP) models have been proposed, including time-grid-based and sequence-based formulations (Harjunkoski et al., 2014). Their capabilities for small-size problems are well established. However, they may fail to solve industrial-scale problems because appropriate number of time points is unknown a priori and feasible solutions are hard to yield in short time frames. Although various decomposition approaches (Nishi et al., 2010) have been attempted, large computational efforts are still inevitable to obtain near optimal or optimal solutions.

Genetic algorithm (GA) is widely embraced to address industrial-scale problems (Woolway and Majozi, 2018) because it can generate good-quality solutions quickly with strong global search capability and inherent parallelism. However, GA is at disadvantage in solution optimality. Han and Gu (2021) showed that worse solutions were obtained for some large examples than MILP models, although computational effort could be significantly reduced. The hybrid algorithm combining advantages of GA on fast convergency and mathematical programming on solution optimality may eliminate limitations of a single algorithm and solve challenging large-scale problems.

In this work, we propose a hybrid algorithm integrating GA and the sequence-based MILP model to generate near-optimal or optimal solutions for industrial-scale scheduling problems. The computational results show that the proposed algorithm can solve large-size problems to optimality within 5 minutes and yield the same or better solutions within shorter computational time compared to the existing MILP methods. Also, the hybrid framework has better performance on both computational effort and optimality than GA and the sequence-based MILP model.

2. Sequence-based MILP formulation

We define three binary variables including $X_{imi'm'}$, $XS_{imi'm's}$, and z_{im}. Specifically, $X_{imi'm'}$ equals 1 if batch m of task i is performed before batch m' of task i' on one unit. $XS_{imi'm's}$ equals 1 when batch m of task i is transferred earlier or simultaneously than batch m' of task i' for storage vessel of state s. z_{im} is 1 if batch m of task i is performed. A task having multiple processing units is split into different tasks. Batches of a task should be processed in sequence, as Eq.(1).

$$z_{im} \le z_{i(m-1)} \qquad \forall i, m > 1 \tag{1}$$

$$X_{imi'(m'+1)} \ge X_{imi'm'} \qquad \forall j, i \in \mathbf{I}_j, i' \in \mathbf{I}_j, i \ne i', m, m' < M \tag{2}$$

$$X_{i(m-1)i'm'} \ge X_{imi'm'} \qquad \forall j, i \in \mathbf{I}_j, i' \in \mathbf{I}_j, i \ne i', m > 1, m' \tag{3}$$

$$X_{imi'm'} + X_{i'm'im} \ge z_{im} + z_{i'm} - 1 \qquad \forall j, i \in \mathbf{I}_j, i' \in \mathbf{I}_j, i < i', m, m' \tag{4}$$

Eq.(2) ensures that if batch m of task i is processed before the batch m' of a task i' on a unit j, then this batch m must also be processed before the batch $(m'+1)$ of task i'. If batch m' of task i' is processed after the batch m of task i on a unit j, then this batch m' must be processed after the batch $(m-1)$ of task i, as Eq.(3). Eq.(4) tells the sequential relation between any two batches of two tasks in the same unit. Batch size b_{im} is bounded by the maximum B_i^{max} and minimum B_i^{min} capacity, as indicated in Eq.(5).

$$B_i^{min} \cdot z_{im} \le b_{im} \le B_i^{max} \cdot z_{im} \qquad \forall i, m \tag{5}$$

The duration of a batch is ensured by Eq.(6), where a_i and β_i are fixed and variable terms in the processing time, respectively. If a state s is subject to zero-wait ($s \in \mathbf{S}^{ZW}$), the duration is exactly equal to the processing time in Eq.(7). Sequencing constraints for the same task and different tasks in a unit are given in Eq.(8) and Eq.(9).

$$T_{im}^{\mathrm{f}} \ge T_{im}^{\mathrm{b}} + a_i \cdot z_{im} + \beta_i \cdot b_{im} \qquad \forall s \in \mathbf{S} \backslash \mathbf{S}^{ZW}, i \in \mathbf{I}_s^P, m \tag{6}$$

$$T_{im}^{\mathrm{f}} = T_{im}^{\mathrm{b}} + a_i \cdot z_{im} + \beta_i \cdot b_{im} \qquad \forall s \in \mathbf{S}^{ZW}, i \in \mathbf{I}_s^P, m \tag{7}$$

$$T_{i(m+1)}^{\mathrm{b}} \ge T_{im}^{\mathrm{f}} \qquad \forall i, m < M \tag{8}$$

$$T_{i'm'}^{\mathrm{b}} \ge T_{im}^{\mathrm{f}} - H \cdot (1 - X_{imi'm'}) \qquad \forall j, i \in \mathbf{I}_j, i' \in \mathbf{I}_j, i \ne i', m, m' \tag{9}$$

T_{ims}^{s} is defined as the transfer time of batch m of task $i \in \mathbf{I}_s$ into or out from the storage vessel of state s. Eq.(10) enforces the transfer time of batch m of task $i \in \mathbf{I}_s^P$ producing state s into storage equals its finish time T_{im}^{f}. The start time (T_{im}^{b}) of batch m of task i consuming state s is equal to the transfer time out from the storage. Sequence constraints on storage for batches of the same tasks and different tasks are given by Eqs.(12-13). Sets $\mathbf{S}^P, \mathbf{S}^R$ and $\mathbf{S}^{in}$ indicate product, raw material and intermediate state, respectively.

$$T_{ims}^{\mathrm{s}} = T_{im}^{\mathrm{f}} \qquad \forall s \in \mathbf{S}^{in}, i \in \mathbf{I}_s^P, m \tag{10}$$

$$T_{ims}^{\mathrm{s}} = T_{im}^{\mathrm{b}} \qquad \forall s \in \mathbf{S}^{in}, i \in \mathbf{I}_s^C, m \tag{11}$$

$$T_{i(m+1)s}^{\mathrm{s}} \ge T_{ims}^{s} \qquad \forall s \in \mathbf{S}^{in}, i \in \mathbf{I}_s, m < M \tag{12}$$

$$T_{i'm's}^{\mathrm{s}} \ge T_{ims}^{s} - H \cdot (1 - XS_{imi'm's}) \qquad \forall s \in \mathbf{S}^{in}, i \in \mathbf{I}_s, i' \in \mathbf{I}_s, m, m', i \ne i' \tag{13}$$

Eqs.(14-16) are formulated to enforce precedence of batches in different tasks on storage, which are similar to Eqs.(2-4). When two tasks can be processed in the same unit and

related to the same state s, sequence relations for their batches keep consistent on unit and storage, as formulated in Eq.(17).

$$XS_{imi'(m'+1)s} \geq XS_{imi'm's} \qquad \forall s \in \boldsymbol{S}^{in}, i \in \mathbf{I}_s, i' \in \mathbf{I}_s, i \neq i', m, m' < M \quad (14)$$

$$XS_{i(m-1)i'm's} \geq XS_{imi'm's} \qquad \forall s \in \boldsymbol{S}^{in}, i \in \mathbf{I}_s, i' \in \mathbf{I}_s, i \neq i', m > 1, m' \quad (15)$$

$$XS_{imi'm's} + XS_{i'm'ims} \geq z_{im} + z_{i'm'} - 1 \quad \forall s \in \mathbf{S}^{in}, i \in \mathbf{I}_s, i' \in \mathbf{I}_s, m, m', i < i' \quad (16)$$

$$XS_{imi'm's} = X_{imi'm'} \qquad \forall j, s \in \boldsymbol{S}^{in}, i, i' \in \mathbf{I}_s \cap \mathbf{I}_j, m, m', i \neq i' \quad (17)$$

$CB_{imi'm's}$ is introduced to monitor batches m' transferred before batch m of different tasks in Eqs.(18-20) or the same tasks by Eq.(21). It equals to $b_{i'm'}$ if batch m' of task i' is transferred before or at the same time as batch m of task i. Otherwise, it equals to 0. Parameter $\rho_{i,s}$ is the fraction for task *i* to produce ($\rho_{i,s}$>0) or consume ($\rho_{i,s}$<0) state *s*. The inventory level in storage is calculated by Eq.(22), being positive and smaller than the maximum storage capacity ST_s^{max} after any transfer. Eq.(23) enforces total amount transferred for batches must satisfy storage limitation, where $\mathrm{In0}_s$ is the initial inventory.

$$CB_{imi'm's} \leq B_{i'}^{max} \cdot XS_{i'm'ims} \qquad \forall s \in \boldsymbol{S}^{in}, i, i' \in \mathbf{I}_s, m, m', i \neq i' \quad (18)$$

$$CB_{imi'm's} \geq b_{i'm'} - B_{i'}^{max} \cdot (1 - XS_{i'm'ims}) \quad \forall s \in \boldsymbol{S^{in}}, i, i' \in \mathbf{I}_s, m, m', i \neq i \quad (19)$$

$$CB_{imi'm's} \leq b_{i'm'} \qquad \forall s \in \boldsymbol{S}^{in}, i, i' \in \mathbf{I}_s, m, m', i \neq i' \quad (20)$$

$$CB_{imi'm's} = b_{i'm'} \qquad \forall s \in \boldsymbol{S}^{in}, i \in \mathbf{I}_s, m, m' < m \quad (21)$$

$$0 \leq \sum_{i' \in \mathbf{I}_s} \sum_{m'} \rho_{i's} CB_{imi'm's} + \rho_{i,s} b_{im} + \mathrm{In0}_s \leq ST_s^{max} \; \forall s \in \boldsymbol{S}^{in}, i \in \mathbf{I}_s, m \quad (22a,b)$$

$$0 \leq \sum_{i \in \mathbf{I}_s} \sum_{m} \rho_{is} \cdot b_{im} + \mathrm{In0}_s \leq ST_s^{max} \qquad \forall s \in \boldsymbol{S}^{in} \quad (23)$$

For the objective of minimizing makespan (*MS*), demand constraints are given in Eq.(24). Makespan must exceed the finish and transfer time of all batches.

$$\sum_{i \in \mathbf{I}_s^P} \rho_{is} \cdot \sum_m b_{im} \geq D_s \qquad \forall s \in \mathbf{S}^P \quad (24)$$

$$MS \geq T_{im}^{\mathrm{f}} \qquad \forall i, m \quad (25)$$

$$MS \geq T_{ims}^{\mathrm{s}} \qquad \forall s \in \boldsymbol{S}^{in}, i \in \mathbf{I}_s, m \quad (26)$$

We fix some variables for two batches of the same task by Eq.(27) because these two batches must be performed in sequence. The variables in Eq. (28) control the state inventory being positive and lower than storage capacity. As there are always abundant raw materials and infinite storage for products, the related variables are fixed as zero.

$$X_{imim'} = 1; X_{im'im} = 0; XS_{imim's} = 1; XS_{im'ims} = 0; CB_{imim's} = 0 \; \forall \mathrm{s}, i, m' > m \quad (27)$$

$$XS_{imim's} = 0, CB_{imim's} = 0, T_{ims}^{s} = 0 \qquad \forall s \in (\mathbf{S}^P \cup \boldsymbol{S}^R), i, m, i', m' \quad (28)$$

3. Genetic algorithm

A knowledge-guided GA is designed to generate good-quality solutions in short time frames. A three-part chromosome $c = (c^P, c^U, c^R)$ is constructed to represent a solution for a given problem. Elements $c_n^P \in [1, P]$ are ordinal numbers of products, thus the sequence in c^P indicates production sequence of products. c^U covers assigned units for batches of tasks that have multiple feasible units. Here, one position of c^U corresponds to

one batch of a task and the element $c_n^U \in [1, \mathbf{J}_n]$ decides the assigned unit j. $c_n^R \in [0,1]$ is used to determine if the latest processed task is repeated. c^R is introduced to influence the sequence of tasks that are processed to produce different batches of products.

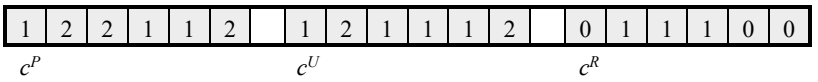


Fig. 1 An example of the designed chromosome

The fitness value of the chromosome is evaluated based on Makespan for the schedule, which is synthesized by decoding. The decoding algorithm works as follows. It steps iteratively through c^P to produce products until demand requirements are met. While producing a product for one batch, any involved task would be performed if its produced material is insufficient. Assigned unit for batch of task is determined by c^U, and latest performed task would be repeated when its corresponding element in c^R is 1. A heuristic rule 'earliest starting strategy' is adopted to start tasks as early as possible. As states are subject to finite or no intermediate storage, inventory level of storage must be monitored and checked at the start and finish times while processing tasks. In GA, the roulette wheel method is adopted to select parents who would be subjected to two-point crossover and two-point mutation. A knowledge-based search (Zheng and Wang, 2018) is incorporated to adjust sequence and assignment of child chromosome based on experiential possibility.

4. Hybrid algorithm

The proposed hybrid algorithm is illustrated in Figure 2. The number of feasible solutions transmitted from GA to MILP is $P_{size} \cdot 0.01$, where P_{size} is the population size of GA.

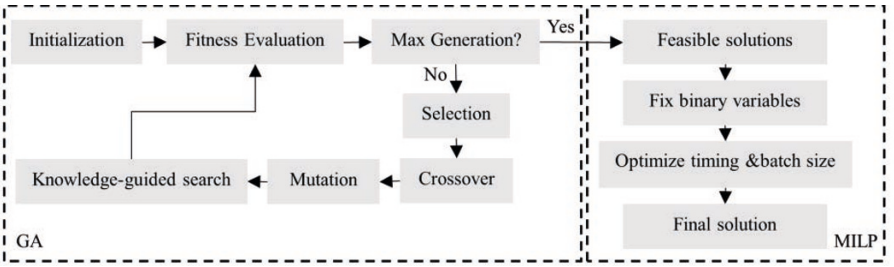


Fig. 2 The proposed hybrid solution algorithm

The strategies used to fix binary variables in the MILP model are described as follows. First, the batches of a task that can be processed in only one unit are fixed to be 1. That is $z_{im} = 1$. Second, tasks i and i' denote one task processing in different units ($j \in \mathbf{J}_i, j' \in \mathbf{J}_{i'}$), and batch m of task i is performed at T_{im}^{b} on unit j. If unit j' from time T_{im}^{b} to ($T_{im}^{\mathrm{b}} + a_i + \beta_i \cdot B_i^{max}$) is idle, batch m is potential to be divided into multiple batches performed on different units, implying z_{im} and $z_{i'm'}$, where batch m' of task i' is not performed in the solution of GA, would be optimized in MILP. Otherwise, $z_{im} = 1$. Third, the precedence between batches of tasks, whose batch information is fixed in the first two steps, on a processing unit in the solution of GA are used to fix $X_{imi'm'}$. $XS_{imi'm's}$ is partially fixed by enforcing batch m' of task $i' \in \mathbf{I}_s^C$ starts after batch m of task $i \in \mathbf{I}_s^P$ providing required state for batch m' (i.e., $XS_{imi'm's} = 1$). Also, batch m' of task i' finishes before batch m of task i whose produced state is consumed by batch m' to ensure inventory level of state s lower than maximum storage capacity (i.e., $XS_{i'm'ims} = 1$).

5. Computational Results

Seven examples from the literature are solved to illustrate the capability of the proposed hybrid algorithm. While Examples 1-3 are from He and Hui (2010), Examples 4-7 refer to the Kallrath example I9, I11, I14, and I15 from Vooradi and Shaik (2012). Examples 1, 2, and 3-7 are small-, medium- and industrial- scale examples, respectively. GA is implemented in MATLAB 2020 and the MILP model is solved using GAMS 33.2 on an AMD Ryzen™ 9 3900X 3.8 GHz, 48 GB RAM, running Windows 10. All examples are also solved using the MILP models of Vooradi and Shaik (2012) and Velez et al. (2015), denoted as VS2012 and VM2015.

The computational results are provided in Table 1, where N is the number of event point required, Gap is the relative gap, and H denotes the time horizon. A competitively efficient algorithm is perceived to find smaller Makespan (MS) or take shorter computational time (CPU). From Table 1, it can be seen that the proposed hybrid algorithm can obtain smaller MS using less CPU time for most examples, compared to VS2012. This can be attributed to the strong global search capacity of the first-stage GA in our model, finding good-quality solutions quickly for industrial-scale problems with a large number of binary variables. The proposed algorithm leads to generate the same global optimum for all examples, as those obtained by VM2015. In our work, sufficient population size and reproduction at the first stage and solutions transmitted to the second stage are required even for simple problems to explore global searching space and reach optimal areas. The CPU time thus is longer than that from VM2015 for small-scale Examples 1-2, but it is still accepted for industrial application. For industrial-scale examples 3-7, much less computational effort is required. The CPU time is reduced by an order of magnitude for Examples 3 and 5.

Table 1. Computational results from the proposed hybrid algorithm and the existing methods

	VS2012				VM2015			Hybrid algorithm	
Ex	N	MS	Gap	CPU (s)	H	MS	CPU (s)	MS	CPU (s)
1	22	37	-	45.2	60	37	0.8	37	3.1
2	65	109	1.4%	>3600	110	108	11.0	108	61.0
3	131	229	5.6%	>3600	219	217	3298.0	217	250.3
4	11	33	3.1%	>3600	60	32	13.2	32	12.1
5	12	40	-	1856.3	60	39	103.9	39	18.7
6	10	36	-	56.3	60	36	50.6	36	9.6
7	23	58	17.3%	>3600	60	52	93.9	52	58.2

The computational results from the hybrid algorithm with GA and the sequence-based MILP model are compared in Table 2. The average MS (denoted as Avg) and the standard deviation (denoted as SD) of MS in 50 runtimes are calculated. Although the hybrid algorithm and GA can find identical optima using similar computational time, the hybrid algorithm outperforms GA due to smaller average MS and more stability (i.e., smaller SD), implying the generation of the global optimum with higher probability. This can be attributed to the integration of the MILP model to further improve quality of feasible solutions from GA. The sequence-based MILP model obtains near-optimal solutions for Examples 4-5 and 7. However, it cannot find any feasible solution (denoted as NA) for

Examples 2-3 in 1 hour, whilst the hybrid algorithm can find the optimal solution within 5 minutes because the majority binary variables at the second stage have been fixed.

Table 2. Comparative results for hybrid algorithm and single algorithms

	GA				Sequence-based MILP			Hybrid algorithm			
Ex	MS	Avg	SD	CPU(s)	MS	Gap	CPU(s)	MS	Avg	SD	CPU(s)
1	37	37.1	0.2	2.6	37	0	1845	37	37.0	0	3.1
2	108	108.2	0.4	51.6	NA	-	>3600	108	108	0	61.0
3	217	217.3	0.5	169.1	NA	-	>3600	217	217.2	0.4	250.3
4	32	32.2	0.5	25.1	34	35.3%	>3600	32	32.1	0.3	12.1
5	39	39.2	0.4	53.5	40	30.0%	>3600	39	39.1	0.2	18.7
6	36	36.1	0.3	16.4	36	26.7%	>3600	36	36.0	0	9.6
7	52	52.4	0.7	87.3	60	33.3%	>3600	52	52.1	0.4	58.2

6. Conclusions

In this work, a hybrid algorithm combining GA and the sequence-based MILP formulation is proposed for scheduling multipurpose batch process. The computational results have demonstrated the capability of the proposed algorithm to solve large-size problems, showing that the same or better optimal solutions can be obtained with dramatically decreased computational time compared to the existing methods. In addition, the hybrid algorithm is superior to the sequence-based MILP model and GA.

Acknowledges

Dan Li appreciates financial support from UOM/CSC Joint Scholarship (201908130170). Jie Li would like to appreciate the financial support from EPSRC (EP/T03145X/1).

References

Y. Han, X. Gu, 2021, Improved multipopulation discrete differential evolution algorithm for the scheduling of multipurpose batch plants, Ind. Eng. Chem. Res., 60, 5530-5547.

I. Harjunkoski, C.T. Maravelias, P. Bongers, P.M. Castro, S. Engell, I.E. Grossmann, J. Hooker, C. Mendez, G. Sand, J. Wassick, 2014, Scope for industrial applications of production scheduling models and solution methods, Comput Chem Eng, 62,161-193.

Y. He, C.W. Hui, 2010, A binary coding genetic algorithm for multi-purpose process scheduling: A case study, Chem. Eng. Sci., 65(16), 4816-4828.

T. Nishi, Y. Hiranaka, M. Inuiguchi, 2010, Lagrangian relaxation with cut generation for hybrid flowshop scheduling problems to minimize the total weighted tardiness, Comput. Oper. Res., 37, 189-198

S. Velez, A.F. Merchan, C.T. Maravelias, 2015, On the solution of large-scale mixed integer programming scheduling models, Chem. Eng. Sci., 136(2), 139-157.

R. Vooradi, M.A. Shaik, 2012, Improved three-index unit-specific event-based model for short-term scheduling of batch plants, Comput Chem Eng., 43, 148-172.

M. Woolway, T. Majozi, 2018, A novel metaheuristic framework for the scheduling of multipurpose batch plants, Chem. Eng. Sci., 192, 678-687.

X. Zheng, L. Wang, 2016, A knowledge-guided fruit fly optimization algorithm for dual resource constrained flexible job-shop scheduling problem, Int. J. Prod. Res.54, 5554-5566.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50077-4

Scheduling of Electrical Power Systems under Uncertainty using Deep Reinforcement Learning

Akshay Ajagekar[a*], Fengqi You[a]

[a]*Cornell University, Ithaca, New York, 14853, USA*
asa273@cornell.edu

Abstract

In this work, a deep reinforcement learning-based solution approach for the unit commitment of power generation resources in energy systems with intermittent renewable energy resources and uncertain loads is presented. Real-world unit commitment problems are plagued with uncertain parameters introduced by the possibility of forecast errors or equipment failure that may negatively impact the power supply. It is imperative to develop a robust and computationally tractable framework to provide cost-effective commitment decisions. In the proposed solution technique, temporal and spatial correlational structures of uncertainties present in the system are captured with a neural network function approximator. The proposed solution technique is able to capture the temporal and spatial correlational structure of uncertainties present in the system. A causal policy is obtained which relies only on previously observed wind power and demand forecasts along with forecast errors. We conduct computational experiments on the IEEE 39-bus test case to demonstrate the effectiveness of the proposed solution strategy and improvement over existing unit commitment solution techniques. The proposed deep reinforcement learning-based solution strategy demonstrates effective computational performance and a reduction in operating costs over deterministic and stochastic approaches.

Keywords: Unit Commitment, Deep Reinforcement Learning, Machine Learning

1. Introduction

Unit commitment (UC) is one of the widely used optimization models in the power industry for scheduling and dispatch of electric power generation resources (Padhy, 2004). The UC problem is NP-hard and is challenging to solve as its size increases (Tseng, 1996). Real-world UC problems are plagued with uncertain parameters caused by the possibility of forecast errors or equipment failure (Håberg, 2019). Forecast uncertainty can affect solution quality and causes service interruptions (Ning et al., 2019, 2022). A more price-responsive demand and high penetration of wind power pose new challenges to the UC problem (Qiu et al., 2021), thus stressing the need for an effective methodology that produces robust UC decisions in the presence of real-time uncertainty (Shang et al., 2019). Several formulations of the UC problem have been previously proposed in terms of different uncertainty representations and solution techniques (Abujarad et al., 2017). Stochastic optimization techniques suffer from high computational costs while robust approaches may yield solutions that are too conservative (Zheng et al., 2014). Therefore, it is imperative to develop a more robust and computationally tractable framework as compared to other stochastic optimization approaches that provide more cost-effective commitment decisions.

Machine learning offers a powerful alternative to solving this scheduling optimization problem (Ning and You, 2019), especially on handling uncertainty. UC problems can also be formulated under a dynamic programming framework for decision-making in multiple stages (Pang & Chen, 1976). There have been attempts to solve stochastic UC problems with reinforcement learning (RL) techniques, including the use of deep neural networks as function approximators (Jasmin et al., 2016). In this paper, we propose a deep reinforcement learning (DRL) based technique for the effective solution of the UC problem under demand and wind power uncertainty. The proposed scheduling technique is capable of capturing the causal nature of uncertainties present in the system with deep neural networks. To demonstrate the applicability and efficiency of the proposed DRL-based solution approach, we conduct computational experiments with the IEEE 39-bus test case. The obtained solutions are also compared with UC solutions obtained using deterministic approaches that use point forecasts for demand load and wind power generation along with other stochastic approaches.

2. MDP Formulation and Safety

We formulate the UC problem as a Markov decision process (MDP). At any time step *t* on any given day, the day-ahead point forecast predictions for demand load and wind power are available. The net load, defined as the difference between total electrical load and total wind generation power, is denoted by $n_t = \sum_b d_{bt} - \sum_b w_{bt}$ and is considered for energy balance. The resulting net load forecasts are obtained accordingly. The uncertainty associated with net load is captured by ζ_t defined as the ratio of forecast error to net load forecast. Along with historical net load forecasts $n^f_{<t}$, historical uncertainty realizations $\zeta_{<t}$, and current forecast n^f_t, the commitment, startup, and shutdown decisions at the previous stage also constitute the system state. To decouple the state variables from multiple stages, additional state variables are introduced at each stage *t,* and are denoted by $u^\tau_{it-1}, \tau = 1,...,UT_i - 1$ and $v^\tau_{it-1}, \tau = 1,...,DT_i - 1$. These state variables store the historical account of startup and shutdown decisions for a duration of minimum up and down-times, respectively. The system's state for all generators $i \in G$ is then fully described by the tuple $s_t = \left(n^f_{<t}, \zeta_{<t}, n^f_t, z_{it-1}, u_{it-1}, v_{it-1}, u^\tau_{it-1}, v^\tau_{it-1}\right)$ The action space for the MDP includes the commitment decisions z_{it} accompanied by the startup and shutdown decisions, u_{it} and v_{it}, respectively. Prediction for the net load forecast error uncertainty denoted by $\hat{\zeta}_t$ is also considered as an action variable. The estimated net load can then be calculated as $n^f_t\left(1+\hat{\zeta}_t\right)$. The safe exploration of UC decisions is performed by solving the optimization problem denoted by $UC\left(\hat{\zeta}_t\right)$ in (1)-(7) which guarantees operational constraint satisfaction in UC. Minimum up and down-time constraints are reformulated to use the additional state variables u^τ_{it-1} and v^τ_{it-1}, and are given by Eq. (4) and (5).

$$\min \sum_{i \in G} \left[C_i\left(p_{it}\right) + C_i^{NL} z_{it} + C_i^{SU} u_{it} + C_i^{SD} v_{it} \right] \tag{1}$$

$$s.t. \quad z_{it} P_i^{\min} \le p_{it} \le z_{it} P_i^{\max} \quad \forall i \in G \tag{2}$$

$$\sum_{i \in G} p_{it} = n^f_t \left(1+\hat{\zeta}_t\right) \tag{3}$$

$$u_{it} + v_{it-1} + \sum_{\tau=1}^{DT_i-1} v_{it-1}^{\tau} \le 1 - z_{it-1} \quad \forall i \in G \tag{4}$$

$$v_{it} + u_{it-1} + \sum_{\tau=1}^{UT_i-1} u_{it-1}^{\tau} \le z_{it-1} \quad \forall i \in G \tag{5}$$

$$z_{it} - z_{it-1} = u_{it} - v_{it} \quad \forall i \in G \tag{6}$$

$$z_{it}, u_{it}, v_{it} \in \{0,1\}, \; p_{it} \ge 0, \; \forall i \in G \tag{7}$$

Selection of control actions is followed by the realization of forecast error ratio uncertainty ζ_t. Based on the observed net load, the commitment decisions can be updated to minimize incurred costs, which is consistent with UC as a multistage decision-making problem. To meet actual net load requirements, fast-start generators can be potentially turned ON or OFF, leading to additional startup or shutdown costs. The objective is to minimize such costs by selecting the control actions. To achieve this, we seek to maximize the reward in (8) with C_u defined in Eq. (9). The maximum achievable reward at any timestep *t* is zero. The optimal control actions required to meet actual net load requirements also constitute the next state s_{t+1}. The additional state variables are updated using transition dynamics shown in Eqs. (10) and (11).

$$r_t = -\left| C_u(\zeta_t) - C_u\left(\hat{\zeta}_t\right) \right| \tag{8}$$

$$\begin{aligned} & C_u(\zeta) = \sum_{i \in G} C_i^{NL} z_{it} + C_i^{SU} u_{it} + C_i^{SD} v_{it} \\ & s.t. \; (z_{it}, u_{it}, v_{it}) \in \arg\min UC(\zeta) \end{aligned} \tag{9}$$

$$u_{it}^{\tau} = u_{it-1}^{\tau-1}, \; \forall i \in G, \tau = 2, ..., UT_i - 1 \tag{10}$$

$$v_{it}^{\tau} = v_{it-1}^{\tau-1}, \; \forall i \in G, \tau = 2, ..., DT_i - 1 \tag{11}$$

3. Actor-Critic Method for Policy Learning

A deterministic policy that estimates net load forecast error ratio is considered here, which also serves to obtain control actions through safely exploring the constrained decision space. The forecast errors are independent of the system states like power dispatch and commitment decisions. So, the policy is considered to be a function of system states consisting of net load forecasts, current forecast, and previous uncertainty realizations, as shown in Eq. (12).

$$\hat{\zeta}_t = \pi\left(n_{<t}^f, n_t^f, \zeta_{<t}\right) \tag{12}$$

The DRL agent is trained using an off-policy actor-critic algorithm. Due to the deterministic nature of the policy and the continuous action space spanned by it, an actor-critic algorithm based on DDPG is used (Lillicrap et al., 2015). Both actor and critic are implemented using deep neural networks and can be parameterized by the networks' weights and biases. The architectures of the parameterized actor and critic can be denoted as $\pi_\theta\left(n_{<t}^f, n_t^f, \zeta_{<t}\right)$ and $Q_\phi(s_t, a_t)$, respectively. π_θ is a feed-forward deep neural network with two fully connected layers with 32 and 16 hidden units with a rectified linear unit (ReLu) activation. A linear activation is used for this final layer since the errors may have negative values. Similarly, the critic Q_ϕ uses the historical

UC decisions $\left\{u_{it-1}^{\tau}, v_{it-1}^{\tau}\right\}$, previous commitment decisions z_{it-1}, as input, along with the predicted forecast errors $\hat{\zeta}_t$ and time-state T_r. All the above state and action variables are concatenated and fed to a feed-forward deep neural network. The deep neural network consists of three fully connected layers with 32, 32, and 16 units, each following a ReLu activation. A single output is then obtained by adding a layer with a single neuron with linear activation that corresponds to the Q-value.

a)

Algorithm 1 Policy learning

Initialize actor $\pi_\theta\left(n_{<t}^f, n_t^f, \zeta_{<t}\right)$ and critic network $Q_\phi(s_t, a_t)$

Initialize target networks $\pi_\theta'\left(n_{<t}^f, n_t^f, \zeta_{<t}\right)$ and $Q_\phi'(s_t, a_t)$ with random weights

Initialize replay buffer R

for episode = 1:M

  Initialize states including previously observed forecast errors

  for t = 1:T

    Estimate forecast error ratio: $\xi \sim U(0,1)$

    $$\hat{\zeta}_t = \begin{cases} \hat{\zeta}_t \sim N\left(\mu_t, \sigma_t^2\right) & \text{if } \xi < \varepsilon \\ \pi_\theta\left(n_{<t}^f, n_t^f, \zeta_{<t}\right) & \text{otherwise} \end{cases}$$

    Determine control actions through safe exploration by solving $UC\left(\hat{\zeta}_t\right)$

    Solve $UC(\zeta_t)$ with observed net load and compute reward r_t

    Store transition (s_t, a_t, r_t, s_{t+1}) in R

    Sample random batch of transitions B from R

    Update actor : $\max J = \sum_{i \in B} Q_\phi\left(s_i, \pi_\theta(s_i)\right)$

    Update critic : $\min L = \sum_{i \in B}\left(y_i - Q_\phi(s_i, a_i)\right)^2$

    $$y_i = r_i + Q_\phi'\left(s_{i+1}, \pi_\theta'(s_{i+1})\right)$$

    Update target network parameters : $\theta' \leftarrow \tau\theta + (1-\tau)\theta$

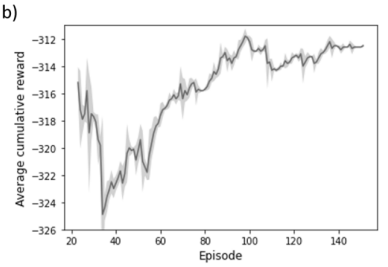
    $$\phi' \leftarrow \tau\phi + (1-\tau)\phi$$

  end

  Update exploration rate : $\varepsilon \leftarrow \varepsilon \cdot \delta$

end

b)



c)

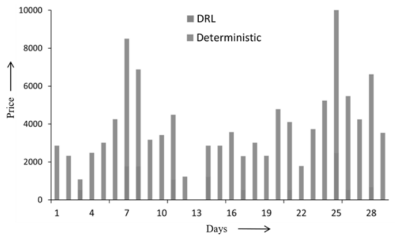


Fig 1. a) Actor-critic based algorithm for policy learning with conservative policy iteration to predict forecast errors, b) average cumulative returns obtained at each episode during policy learning and c) costs incurred by the system operator for starting additional generating units due to insufficient dispatch by the committed units

In order to perform soft updates in conservative policy iteration, we also initialize copies of actor and critic as targets denoted as $\pi_\theta'\left(n_{<t}^f, n_t^f, \zeta_{<t}\right)$ and $Q_\phi'(s_t, a_t)$. Typically, exploration in DDPG is performed by generating noise by sampling from a correlated normal distribution. Parameters of this normal distribution must be carefully chosen for effective performance. To bypass this restriction, an alternate approach based on the epsilon-greedy strategy is employed. In epsilon-greedy, forecast error ratio estimates are selected randomly with a probability of ε, also termed as exploration rate. $\hat{\zeta}_t$ is randomly sampled from a normal distribution $N\left(\mu_t, \sigma_t^2\right)$ where the parameters μ_t and σ_t are obtained by fitting historical values of forecast error ratios at t^{th} hour of the day to a normal distribution. The commitment and startup/shutdown control decisions are obtained through safely exploring the decision space without violating any constraint imposed by the estimated net load. This is followed by solving $UC(\zeta_t)$ with the observed net load. The transitions are stored in a replay buffer R of a fixed size. To train the actor and critic networks, we randomly sample a batch of transitions B from the replay buffer. The pseudo-code of the algorithm is provided in Fig. 1a.

4. Case Study: IEEE 39-Bus System

We consider an IEEE test system to demonstrate the applicability of the proposed DRL-based approach. We use historical demand data from NYISO for both learning and evaluation purposes. Both hour-ahead load forecasts and actual loads in various zones in New York are extracted from the historical data. The Eastern Wind Integration dataset provided by National Renewable Energy Laboratory (NREL) consists of hour-ahead wind power forecasts for simulated wind farms. A real-time dispatch process is simulated to validate the viability of the proposed DRL-based solution approach. The training process for the DRL agent is conducted in an episodic manner with episodes of length 24 hourly timesteps and converges in approximately 150 episodes. The length of the horizon for historical values that constitute the state is set to 12. The reward curves can be visualized in Fig. 1b, where the average cumulative returns over the length of the horizon are plotted. We conduct UC and economic dispatch simulations for each day of five consecutive months. The daily average costs are reported in Table 1 for each solution technique. The daily costs reported for the deterministic approach and DRL-based approach are the actual incurred real-time costs. Costs incurred with the DRL-based approach are considerably lower than the upper bounds provided by the SDDiP algorithm. SDDiP is a sampling-based variant of Bender's decomposition typically used for scenario decomposition techniques to solve the stochastic UC problem. Optimal costs of operation obtained with perfect uncertainty information can be used to compute the gap between the lower bound and the obtained solution. An average optimality gap of 6.65% is observed with the DRL-based solution for UC.

Table 1. Daily average costs obtained by different approaches for the IEEE 39-bus system

	Perfect knowledge ($)	***Deterministic approach ($)***	***SDDiP ($)***	***DRL ($)***
September	380,966	404,326	426,592	403,792
October	402,485	432,194	442,540	430,969
November	396,632	424,512	440,353	423,378
December	401,738	430,380	445,365	429,449
January	397,962	425,625	439,851	424,593

Since the goal of the DRL agent is to minimize excess penalty costs, the DRL approach is expected to produce minimum penalties by committing appropriate units capable of satisfying the observed load requirements. A representation of the penalties and their frequency for the month of January is shown in Fig. 1c. The costs incurred by the system operator for starting additional units to meet net load requirements are significantly higher than excess costs incurred by the DRL approach. Low penalties are eventually responsible for lower total operating costs with the DRL approach as compared to the deterministic approach. From the penalty costs, their frequency, and observed commitment status of generators, it can be inferred that the trained DRL agent is able to anticipate appropriate forecast error uncertainties and yields UC decisions that can satisfy net load requirements without starting additional units.

5. Conclusion

In this paper, we proposed a DRL-based scheduling approach for the UC problem under demand and wind power uncertainty. This involved formulating the UC problem as an MDP and maximizing cumulative rewards by the actor-critic algorithm. A trained

actor that predicts the net load forecast error ratios was obtained by training the DRL agent. Here, both actor and critic networks were parameterized by deep neural networks. Zero violation of any operational constraint in the UC problem is also guaranteed by our proposed approach. The efficiency of the proposed approach was evaluated on the IEEE 39-bus test system. Evaluation of the trained policy on the real-world load and wind power forecast data resulted in a reduction of penalty costs of commitment of additional units as well as a substantial overall cost saving with the proposed DRL-based scheduling approach.

References

S. Y. Abujarad, M. W. Mustafa, J. J. Jamian, 2017, Recent approaches of unit commitment in the presence of intermittent renewable resources: A review. Renewable and Sustainable Energy Reviews, 70, 215-223.

A. Ajagekar, F. You, 2021, Quantum computing based hybrid deep learning for fault diagnosis in electrical power systems. Applied Energy, 303, 117628.

M. Håberg, 2019, Fundamentals and recent developments in stochastic unit commitment. International Journal of Electrical Power & Energy Systems, 109, 38-48.

E. Jasmin, T. I. Ahamed, T. Remani, 2016, A function approximation approach to Reinforcement Learning for solving unit commitment problem with Photo voltaic sources. IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES).

T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, D. Wierstra, 2015, Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971.

C. Ning, F. You, 2018, Data-driven stochastic robust optimization: General computational framework and algorithm leveraging machine learning for optimization under uncertainty in the big data era. Computers & Chemical Engineering, 111, 115-133.

C. Ning, F. You, 2019, Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming. Computers & Chemical Engineering, 125, 434-448.

C. Ning, F. You, 2019, Data-Driven Adaptive Robust Unit Commitment Under Wind Power Uncertainty: A Bayesian Nonparametric Approach. IEEE Transactions on Power Systems, 34, 2409-2418.

C. Ning, F. You, 2022, Deep Learning Based Distributionally Robust Joint Chance Constrained Economic Dispatch Under Wind Power Uncertainty. IEEE Transactions on Power Systems, 37, 191-203.

N.P. Padhy, 2004, Unit commitment-a bibliographical survey. IEEE Transactions on Power Systems, 19, 1196-1205.

C. Pang, H. Chen, 1976, Optimal short-term thermal unit commitment. IEEE Transactions on Power Apparatus and Systems, 95(4), 1336-1346.

H. Qiu, W. Gu, F. You, 2021, Bilayer Distributed Optimization for Robust Microgrid Dispatch With Coupled Individual-Collective Profits. IEEE Transactions on Sustainable Energy, 12, 1525-1538.

H. Qiu, F. You, 2020, Decentralized-distributed robust electric power scheduling for multi-microgrid systems. Applied Energy, 269, 115146.

C. Shang, F. You, 2019, Data Analytics and Machine Learning for Smart Process Manufacturing: Recent Advances and Perspectives in the Big Data Era. Engineering, 5, 1010-1016.

C. L. Tseng, 1996, On power system generation unit commitment problems. University of California, Berkeley.

A. J. Wood, B. F. Wollenberg, G. B. Sheblé, 2013, Power generation, operation, and control. John Wiley & Sons.

Q. P. Zheng, J. Wang, A. L. Liu, 2014, Stochastic optimization for unit commitment—A review. IEEE Transactions on Power Systems, 30(4), 1913-1924.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50078-6

A Reinforcement Learning Approach to Online Scheduling of Single-Stage Batch Chemical Production Processes

Max Mowbray[a], Dongda Zhang[a], and Antonio Del Rio Chanona[b,*]

[a] *Centre for Process Integration, School of Chemical Engineering and Analytical Science, The University of Manchester, Manchester, M13 9PL, United Kingdom*
[b] *Centre for Process Systems Engineering (CPSE), Department of Chemical Engineering, Imperial College London, London, SW7 2AZ, United Kingdom*
a.del-rio-chanona@imperial.ac.uk

Abstract

The field of Reinforcement Learning (RL) has received a lot of attention for decision-making under uncertainty. Lately, much of this focus has been on the application of RL for combinatorial optimisation. Recent work has showcased the use of RL on a single-stage continuous chemical production scheduling problem. This work highlighted the potential of RL for optimal decision-making under uncertainty in the paradigm of (bio)chemical production scheduling. However, this novel approach is yet to be tested in the context of parallel unit operations and batch processing systems. In this work, we outline a framework for the use of RL to handle single-stage parallel, batch production. In particular, we incorporate elements such as uncertainties in the model data, limited batch size, sequencing constraints, and uncertainties in processing times and product demand, which make for a substantially harder problem. To handle the presence of precedence or succession constraints, by taking inspiration from approaches such as generalised disjunctive programming, we propose a novel methodology that identifies transformations of the control set available to the RL at each control interaction. Given that production typically operates under standard operating procedures, such transformations can be identified by logic. The efficacy of policy synthesis via evolutionary RL methods is benchmarked against mixed integer programming. The results of this study provide further support for the use of RL in online scheduling.

Keywords: Reinforcement Learning; Combinatorial Optimization; Production Scheduling; Machine Learning.

1. Introduction

The production scheduling of (bio)chemical processes is a major field of process systems engineering research. The foundational developments in the area from the 1980s – 2010s focused on the development of efficient, rigorous, finite dimensional mathematical models for use in optimization. The contributions made by Kondili *et al.* (1993) and Pantelides (1994), formed basis for the general description of production scheduling problems via discrete time and continuous time mixed integer programming (MIP) formulations on the basis of various network models of the scheduling problem. Additionally, due to the practical difficulties in solving mixed integer nonlinear models, typical MIP models that are developed are linear or are dependent upon linearizations of nonlinearities. Further drivers of the field include the development of robust optimization

models and innovative frameworks in the context of reactive (online) scheduling that aim to update the process schedule as uncertainties are realised. Despite the developments made to date in the field, in practice many production schedules are generated (in real plants) by teams of schedulers who rely upon their working knowledge and available heuristic rules, with relatively little reliance upon the rigorous mathematical models proposed by academia (Harjunkoski *et al.*, 2014) . This is due to a) the complexities of obtaining finite dimensional models robust to the underlying uncertainties that are computationally tractable online, and b) difficulties in accurately estimating suitable uncertain parameters (i.e. sets or probability distributions descriptive of them).

In an attempt to handle the challenges mentioned, recent works have investigated the use of Reinforcement Learning (RL) in the context of (bio)chemical production scheduling (Hubbs *et al.,* 2020a). RL promises to remedy the challenges previously discussed by a) providing reactive, uncertainty aware scheduling decisions via inference (i.e. prediction of scheduling decisions from a function) and shifting the computational load offline, and b) providing basis for the use of a greater range of models and descriptions of the underlying uncertainty. However, the application of RL to (bio)chemical production scheduling has been relatively limited. In the novel study provided by Hubbs *et al.* (2020a), the authors consider the sequencing of tasks on a single unit, in a single-stage continuous chemical production problem. The work demonstrates results that are competitive with stochastic and deterministic reactive MILP approaches.

Despite the promise of RL, it is yet to be demonstrated on case study with the type of complexity in decision-making seen in a real plant (globally). In this work, we consider the development of a methodology for the use of RL in a parallel, single-stage batch (bio)chemical production scheduling study with multiple units and various sources of uncertainty. We present a methodology based on a discrete time transcription of the production scheduling problem, although it is possible to use a continuous time (event-based) approach. Due to the presence of sequencing constraints derived from standard operational procedures (SOPs), the problem is complex. To mitigate the demands of learning a feasible policy through the reward function, we propose to aid the control selection by identifying nonlinear transformations of the prediction based on the SOPs stated in case study. This reduces the demands of learning through a reward signal alone. Such an approach has been previously studied, where transformations have instead been learned (Bamford and Ovalle, 2021). A similar idea is exploited in generalised disjunctive programming (GDP).

2. Methodology

In this work, we assume that there is a Markov decision process (MDP) that well represents the problem of scheduling single-stage batch operations in parallel in a (bio)chemical production plant. Specifically, we assume that there is: a set of states, $\boldsymbol{x} \in \mathbb{X} \subseteq \mathbb{R}^{n_x}$, that make the problem fully observable; a set of available control inputs $\boldsymbol{u} \in \mathbb{U} \subseteq \mathbb{Z}^{n_u}$ that may be selected; a reward function, $R: \mathbb{X} \times \mathbb{U} \times \mathbb{X} \to R_{t+1} \in \mathbb{R}$, that ranks process evolution with respect to control objectives; and, a probabilistic description of process evolution, such that:

$$\boldsymbol{x}_{t+1} = f(\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{s}_t) \tag{1}$$

where $t \in \{0, \dots, T\}$ is a discrete time index and the process is considered to evolve over discrete time horizon; and, $\boldsymbol{s} \in \mathbb{S} \subseteq \mathbb{R}^{n_s}$ is a realization of (general) process uncertainties.

Operationally, there is a constraint set, $\widehat{\mathbb{U}}_t \subset \mathbb{Z}^{n_u}$, that defines the available tasks or jobs, that may be scheduled in a given unit, $L \in \{1, \dots, n_u\}$, at any given time index. This can be derived from standard operating procedures (SOPs) that define the viable sequencing of operations in units, requirements for unit cleaning and maintenance periods, requirements for orders to be processed in campaigns (i.e. multiple batches consecutively if the order size is greater than maximum batch size of a unit) and that these batches must finish before another job or task is assigned to a given unit. Given the scheduling problem that we consider in this work adheres to a discrete time transcription, in essence, we are solving the following discrete time, finite horizon stochastic optimal control problem:

$$P(\pi_c) = \begin{cases} \max\limits_{\pi_c} \mathbb{E}_{\pi_c}\left[\sum\limits_{t=0}^{T-1} R_{t+1}\right] \\ X_0 \sim p(\boldsymbol{x}_0) \\ \boldsymbol{x}_{t+1} = f(\boldsymbol{x}_t, \widehat{\boldsymbol{u}}_t, \boldsymbol{s}_t) \\ \boldsymbol{u}_t = \pi_c(\boldsymbol{u}_t|\boldsymbol{x}_t) \\ \boldsymbol{u}_t \in \widehat{\mathbb{U}}_t \subseteq \mathbb{Z}^{n_u} \\ \forall\, t \in \{0, \dots, T\} \end{cases} \tag{2}$$

where $X_0 \in \mathbb{X}$ is a random variables described by the initial state distribution, $p(\boldsymbol{x}_0)$; and, $\pi_c(\boldsymbol{u}_t|\boldsymbol{x}_t)$ is a conditional probability mass function over controls, given the current state. In RL practice, the aim is to learn a functionalization of the policy, $\pi_c(\boldsymbol{u}_t|\boldsymbol{x}_t;\theta)$, where $\theta \in \mathbb{R}^{n_\theta}$. Conventionally, the functionalization is chosen to be nonlinear and suited to *end-to-end* learning, such that neural networks are often favoured. Selection of control inputs to the system (conditional to the state) are then provided by inference and learning of the optimal policy parameters, $\theta^* \in \mathbb{R}^{n_\theta}$, are learned through the reward function, R, and general policy iteration algorithms. Two points are worth noting here that provide basis for the methodology subsequently presented: a) the MDP framework does not naturally handle the hard constraints imposed by $\boldsymbol{u_t} \in \widehat{\mathbb{U}}_t$, and b) Eq. 2 formulates control inputs (decisions) as integer values that identify the allocation of a given task or job in a unit at a given time index. We explore how best to handle these two issues in the following.

Handling the constraints imposed on control selection in classical problems such as dynamic optimization of fed-batch processes with continuous control spaces (with upper and lower bounds) is often implicit (i.e. use of an activation function over the ANN output layer naturally places upper and lower bounds on control selection). In this class of problem the structure of the constraints on the control space is different and arises from SOPs. Given SOPs are typically defined logically (as in sequencing constraints), one can identify an additional transformation of the set of available controls, $\mathbb{U}$, at each control interaction, t, based on the current state of the plant and the SOPs themselves. This functional transformation is denoted, f_{SOP}: $\mathbb{U} \times \mathbb{X} \rightarrow \widehat{\mathbb{U}}$ and is assumed non-differentiable.

The conventional approach to select discrete control decisions from a function is either to a) predict the conditional probability density of each control explicitly in the output of the policy functionalisation, or b) predict the state-action value of each control in the output of the policy functionalization (this then enables the generation of a conditional probability mass function according to e.g. ϵ – greedy policies). However, these approaches scale poorly with the number of orders and units common to scheduling problems. A more intelligent approach is instead to predict a real value in a continuous

latent space, $w \in \mathbb{W}$, and then transform that prediction to a corresponding discrete control decision, $\boldsymbol{u}_t$. This is a common approach in recommender systems. The transformation could either be guided by the state-action value of the k nearest integer controls, or could be a deterministic rounding policy (i.e. the nearest integer function), denoted $f_r: \mathbb{W} \rightarrow \mathbb{U}$, as demonstrated (implicitly) in Hubbs *et al*. (2020b). Both transformations are non-differentiable.

Assuming the constraint set, $\hat{\mathbb{U}}$, at each control interaction can be identified, the rounding policy can be defined in this work as $f_r: \mathbb{W} \rightarrow \hat{\mathbb{U}}$, which enables selection of controls that explicitly satisfy the constraint set. In the case one is unable to identify $\hat{\mathbb{U}}$ absolutely via f_{SOP}, one can penalise violation of those constraints not handled innately by incorporating a deterministic expression for the constraint violation into a penalty function, $\varphi: \mathbb{X} \times \mathbb{U} \times \mathbb{X} \rightarrow \mathbb{R}$, (this is trivial if the constraint is neither subject to uncertain parameters, $\boldsymbol{s} \in \mathbb{S}$, nor soft). A figurative description of the algorithm proposed is provided by Figure 1.

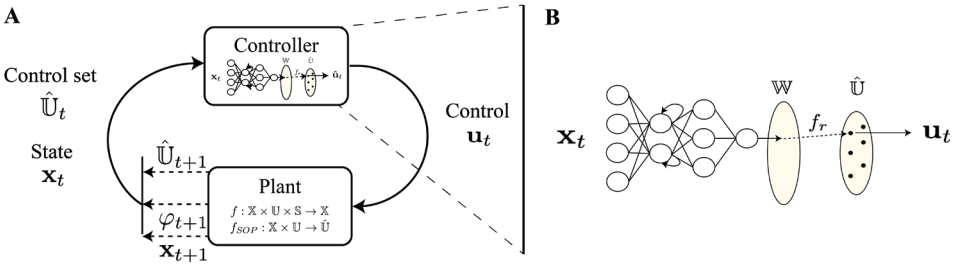


Figure 1: Figurative description of A) the handling of constraints on the control inputs via logical expression, as well as method of control selection, B) shows control selection in more detail.

Due to the problem posed, and the structure of the control space, this work uses evolutionary RL. Here, the exploration-exploitation paradigm is moved directly to the parameter space - removing reliance on first order gradients indicative of directions for policy improvement and mitigates the potential for policies to get stuck in local optima as well as instability in training. This is particularly likely using conventional policy optimization, given a) directions to improve policy parameters are estimated from the state distribution induced under the policy and b) in this work small changes in the policy parameters have potential to drastically alter the state distribution induced.

3. Case Study

To demonstrate the ideas presented in the methodology we work from the case study presented in Cerdá *et al.* (1997). In this work, the authors present a continuous time MILP model for the optimization of a single stage, parallel batch production environment (based on the state-task network representation). The problem definition is provided in Cerdá *et al.* (1997) and we refer the reader there for more details, however uncertain modifications are detailed by Table 1. It is of note that the methodology applies when uncertainty is derived from set based descriptions too. The problem is transcribed from a continuous time to a discrete time formulation (for RL) and to ensure the two are comparable, all processing times and cleaning times, are modified so that their greatest common factor is equivalent to the discrete time interval, Δt, used to define the time grid. In this work, $\Delta t = 0.5$ days and the modified case study data is provided by Table 2. Here we demonstrate the methodology with $J = 8$ orders and $T = 200$ (corresponding to 100 days). Due to the nature of the scheduling problem and the simulation-based methodology proposed, the

underlying simulation model used to generate the following results is both non-smooth and dependent on logic.

Table 1: Definition of the uncertain process parameters

Description	Descriptive distribution
A due date, τ_i for each customer order is uncertain, but has been estimated via a nominal value of $\bar{\tau}_i$ days from the start of the horizon. The variable τ_i is realized two days before delivery	$\tau_i \sim Pois(\bar{\tau}_i)$
The processing time, p_{il}, of task i in a unit l is subject to uncertainty but has a nominal value in days, $\bar{p}_{il}$	$p_{il} \sim U(p_{il}^{LB}, p_{il}^{UB})$ $p_{il}^{LB} = \max(0.5, \bar{p}_{il} - 0.5)$ $p_{il}^{UB} = \bar{p}_{il} + 0.5$

Table 2: Detail of the modified parameters from the original case study to enable comparison between the two time transcriptions. All other parameters are the same as the original study.

	Cleaning Time (days) between preceding task, i, and successor task, j								Nominal Processing Time (days) of task i			
	Task, j								Unit			
Task i	1	2	3	4	5	6	7	8	1	2	3	4
1	-	-	-	-	-	0.5	-	-	2.0	-	-	-
2	-	-	1.0	-	-	-	-	-	-	-	1.0	-
3	1.0	0.5	-	-	-	-	0.5	-	1.0	-	1.0	-
4	-	-	-	-	0.5	-	-	-	-	1.5	-	-
5	-	-	-	0.5	-	0.5	1.0	0.5	-	1.5	-	1.0
6	1.5	-	0.5	0.5	-	-	-	-	2.5	2.0	-	-
7	-	2.0	-	-	1.0	-	-	0.5	-	-	1.0	1.5
8	-	-	-	-	-	-	1.5	-	-	-	-	2.0

In practice, we are unable to explicitly identify $\hat{\mathbb{U}}_t$ explicitly. Instead, we are only able to identify $\bar{\mathbb{U}}_t$, where $\hat{\mathbb{U}}_t \subset \bar{\mathbb{U}}_t$. The constraint not handled innately through f_r is that a given task cannot be processed in more than one unit at the same time. Given, the objective of the scheduling function is to minimise tardiness in orders and makespan, we can declare the following penalty function (reward) and propose to use particle swarm optimization (PSO) to maximise it:

$$\varphi_{t+1} = \sum_{i=1}^{J} r_{t+1,i} + m - \kappa \|C\|_2 \tag{3}$$

where $C = [[c_1]^-, \dots, [c_I]^-] \in \mathbb{Z}^{n_u}$; $c_i = \sum_{l \in L} W_{ilt} - 1$, where $W_{ilt} \in \mathbb{Z}_2$ is a binary variable indicating whether task i is scheduled in unit l at time t; $[v]^- = \max(0, v)$; $\kappa = 250 \in \mathbb{R}^{++}$ is a penalty weight; $r_i = -1 \in \mathbb{R}^-$ is a penalty for the tardiness of an order (i.e. where $\tau_i < t_f^i$); and, $m = -1 \in \mathbb{R}^-$ is a penalty per discrete time step. Key performance indicators are $G_{\pi_c} = \mathbb{E}_{\pi_c}[\sum_{t=1}^{T} \varphi_t]$ and $\sigma_{\pi_c} = \Sigma[\sum_{t=1}^{T} \varphi_t]^{0.5}$.

4. Results and Discussion

To first benchmark the performance of the RL approach, we analyse the optimality of the solution policy found under a nominal model, which essentially corresponds to the generation of a schedule offline (i.e. for a deterministic reality). We find that the RL is able to obtain the same score under the objective function as the original MILP model for the case when there is no finite release time and when there is (i.e. $\varphi_{NoRT} = -62$ and $\varphi_{RT} = -65$). Now, we turn our attention to benchmarking the method when uncertainties

are present. The results are detailed by Table 3. We conduct 8 experiments investigating the potential of RL to handle process uncertainty and benchmark to a shrinking horizon MILP approach, which uses the nominal data (corresponding to $\bar{p}_{ij}$ and $\bar{\tau}_{ij}$).

Table 3: Results of online optimization of the production schedule for RL and MILP approaches. The more positive G_{π_c} the better (as the RL problem is posed as a maximization).

Exp No.	Due Date uncertainty	Processing time uncertainty	Release Times	RL (G_{π_c}+/- σ_{π_c})	MILP (online) (G_{π_c}+/- σ_{π_c})
1	False	True	False	**-61.9** +/- 4.4	-63.3 +/- 4.4
2	False	True	True	**-66.0** +/- 4.9	-66.3 +/- 4.9
3	True	False	False	**-66.8** +/- 8.7	-70.1 +/- 9.6
4	True	False	True	-73.8 +/- 10.7	**-73.6** +/- 10.3
5	True	True	False	**-67.4** +/- 10.9	-71.6 +/- 11.3
6	True	True	True	-75.3 +/- 11.5	**-75.1** +/- 11.7

From Table 3, it is clear that the RL approach proposed is competitive with the MILP benchmark proposed in Cerdá *et al.* (1997). In fact, the RL outperforms the MILP approach in 4 out of the 6 experiments. It is thought the benefits arise from the ability of RL to explicitly consider uncertainty. Furthermore, in this study the RL approach is 150 times computationally cheaper to identify a reactive scheduling decision online.

5. Conclusions

In alignment with the drivers for production to become more distributed, flexible and reactive to realisations of real world uncertainty, we have demonstrated the application of RL for the online optimization of single-stage, parallel batch (bio)chemical production scheduling. We show that the approach is competitive with online MILP approaches, but has the benefit of significant computational savings online. Future work will consider: the application to a larger plant; transcription to a continuous time model; application to multi-stage processes with resource constraints; and, integration into a framework that provides certainty for operators.

References

C.C. Pantelides, Unified frameworks for optimal process planning and scheduling, Proc. Second Conf. on Foundations of Computer Aided Operations, 253-274, (1994)

E. Kondili, C.C. Pantelides, R.W.H. Sargent, A general algorithm for short-term scheduling of batch operations – I. MILP formulation,, Comput. Chem. Eng. 17, (1993)

I. Harkunkoski, C.T. Maravelias, P.Bongers, P.M. Castro, S. Engell, I.E. Grossmann, J. Hooker, C. Mendez, G. Sand, J. Wassick, Scope for industrial applications of production scheduling models and solution methods, Comput. Chem. Eng. 62, (2014)

C. D. Hubbs, C. Li, N.V. Sahinidis, I.E. Grossmann, J.M. Wassick, A deep reinforcement learning approach for chemical production scheduling, Comput. Chem. Eng., 141 (2020a)

C. Bamford, A. Ovalle, Generalising Discrete Action Spaces with Conditional Action Trees, arXiv preprint arXiv:2104.07294v1 (2021)

C.D. Hubbs, H.D. Perez, O.Sarwar, N.V. Sahinidis, I.E. Grossmann, J.M. Wassick, OR-Gym: A Reinforcement Learning Library for Operations Research Problems, arxiv:2008.06319 (2020b)

J. Cerdá, G. P. Henning, I.E. Grossmann, A Mixed-Integer Linear Programming Model for Short-Term Scheduling of Single-Stage Multiproduct Batch Plants with Parallel Lines, Ind. Eng. Chem. Res. 36, 5, (1997)

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50079-8

An adaptive multi-cut decomposition based algorithm for integrated closed loop scheduling and control

Ilias Mitrai[a], Prodromos Daoutidis[a,*]

[a]*Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, 55455, USA*
daout001@umn.edu

Abstract

The integration of scheduling and control leads to large scale optimization problems whose monolithic solution is challenging. In this paper we propose an adaptive muti-cut algorithm to solve the integrated optimization problem to global optimality in reduced time. The reduction of the CPU time is achieved via the addition of multiple cuts and the adaptive addition/removal of cuts based on the executed schedule. We apply the proposed approach to a continuous system and analyze its computational performance.

Keywords: Control; Optimization; Operation

1. Main Text

Fast changing economic environments render the traditional sequential decision-making strategy suboptimal. The integration of scheduling and control is considered a promising avenue to improve the economic performance of process systems by considering simultaneously decisions at both time scales [Daoutidis et al., 2018]. Closed loop scheduling, i.e. modification of the schedule in response to production changes or disturbances is also essential to guarantee its feasibility and optimality during real time operation [Zhuge and Ierapetritou, 2012, Risbeck et al., 2019]. The implementation of such a closed loop approach is limited by the computational complexity of the problem. The key difficulty arises due to the nonlinear behavior of process systems, which leads to nonconvex, large scale optimization problems whose real time solution is challenging. Different approaches have been proposed to address this issue. In one approach, surrogate models have been used to approximate the dynamic behavior of the system [Pattison et al., 2017] and the cost associated with the execution of processing tasks [Charitopoulos et al., 2019, Chu and You, 2013]. The solution time can also be improved by exploiting the structure of the problem using decomposition based solution algorithms [Chu and You, 2015]. However, these methods cannot always guarantee global optimality due to the nonconvexity of the problem. In this work we propose an adaptive multi-cut decomposition based algorithm for integrated closed loop scheduling and control for multiproduct continuous systems. The proposed algorithm is based on a hybrid multi-cut Generalized Benders Decomposition (GBD) algorithm proposed by Mitrai and Daoutidis (2021). In this approach the integrated problem considers simultaneously all the transitions between the products for all the slots and the cost associated with the dynamic transitions between the products is approximated using cuts. The solution of this problem provides the production sequence and the state and manipulated variable profiles to be implemented. The integrated problem is resolved to

compensate for updated process information (the values of the state variables of the system, the time horizon, the inventory level, product demand and price, etc.). In order to accelerate the solution of the integrated problem at different time points, we propose the adaptive addition/removal of cuts. Specifically, the cuts added at previous time points are incorporated directly in the solution of the problem, only for the transitions that can possible occur. This adaptive approach leads to a reduction in computational time and thus enables fast rescheduling as necessary. We apply this approach to a continuous stirred tank reactor and analyze the ability of the proposed approach to handle disturbances at both the scheduling and control level.

2. Problem formulation and decomposition

2.1. Scheduling problem

We will assume that N_p products must be produced over a time horizon H which is discretized into N_s slots. We define variable $W_{ik} \in \{0,1\}$ which is equal to 1 if product i is produced in slot k and zero otherwise, and variable $Z_{ijk} \in \{0,1\}$ which is equal to 1 if a transition occurs between products i and j in slot k and 0 otherwise. The logic constraints are:

$$\sum_i W_{ik} = 1 \; \forall \, k \tag{1}$$

$$Z_{ijk} \geq W_{ik} + W_{j,k+1} - 1 \; \forall \, i,j,k \neq N_s \tag{2}$$

The starting and ending time in slot k are T_k^s ($T_1^s = T_0$) and T_k^e ($T_{N_s}^e = H$) respectively. The production time of product i in slot k is Θ_{ik}, the transition time in slot k is θ_k^t. The timing constraints are the following:

$$T_k^e = T_k^s + \sum_i \Theta_{ik} + \theta_k^t \;\; \forall k \neq N_s \tag{3}$$

$$T_{k+1}^s = T_k^e \;\; \forall k \neq N_s \tag{4}$$

$$\Theta_{ik} \leq W_{ik} H \; \forall i,k \tag{5}$$

The amount of product i manufactured and stored in slot k is q_{ik} and S_{ik} respectively. The demand of product i is d_i, the production rate is r_i and the amount of product i sold in slot k is S_{ik}. The production constraints are

$$\begin{aligned} &I_{ik} = I_{ik-1} + \mathrm{r}_i \, \Theta_{\mathrm{ik}} - S_{ik} \;\; \forall i,k,k > 1 \\ &S_{iN_s} \geq d_i \; \forall \, i. \end{aligned} \tag{6}$$

2.2. Dynamic model

The dynamic behaviour of the system is described by a set of ordinary differential equations $\dot{x} = F(x,u)$, where x are the state variables, u are the manipulated variables and F are vector functions. These equations are discretized using the method of orthogonal collocation on finite elements. We consider simultaneously all the transitions and define x_{ijfck}^n and u_{ijfck}^m as the value of state n and manipulated variable m for a transition from product i to product j in slot k, finite element f and collocation point c. Finally, we define θ_{ijk} as the transition time for a transition from product i to j in slot k, and the discretized equations are

$$x_{ijfck} = F_d\left(\dot{x}_{ijfck}, u_{ijfck}, x0_{ijfk}, \theta_{ijk}\right) \forall\, n, i, j, f, c, k$$
$$x0_{ij1k} = x_i^{ss},\ x_{ijN_{fe}N_{cp}k} = x_i^{ss} \forall\, i, j, k \tag{7}$$
$$u_{ij11k} = u_i^{ss},\ u_{ijN_fN_ck} = u_j^{ss} \forall\, i, j, k$$

where x_i^{ss}, u_i^{ss} are the steady state values of the state and manipulated variables for product i and F_d denote the discretized equations.

2.3. Integrated problem

The objective function of the integrated optimization problem is $\Phi_1 - \Phi_2$, where

$$\Phi_1 = \sum_{ik}\left(P_{ik}S_{ik} - C_{ik}^{op}q_{ik} - C^{inv}I_{ik}\right) - \sum_{ijk} C_{ij}^{tr}Z_{ijk}$$

$$\Phi_2 = \sum_{ijk} Z_{ijk}a_u\left(\sum_{fc} N_{fe}^{-1}t_{ijfck}^{d}\Lambda_{cN_c}\left(u_{ijfck} - u_j^{ss}\right)^2\right) = \sum_{ijk} Z_{ijk}a_u f_{ijk}^{dyn}.$$

P_i, C_i^{op} are the price and operating cost of product i, C^{inv} is the inventory cost, C_{ij}^{tr} is the transition cost from product i to j and a_u is a weight coefficient. Finally, the transition time for each slot and period depends on the transitions that occur and we define θ_{ijk} as the transition time from product i to j in slot k (the lower bound is the minimum transition time $\theta_{ij}^{\min}$) and the following equations are added:

$$\theta_k^t = \sum_{i,j} \theta_{ijk}Z_{ijk} \ \forall\, i, j, k \neq N_s \tag{8}$$

The goal of the optimization problem is to maximize $\Phi_1 - \Phi_2$ subject to Eq. 1-8.

3. Decomposition based solution algorithm

In this section we present the hybrid multi-cut GBD algorithm [Mitrai and Daoutidis, 2021]. Analysis of the structure of the problem via Stochastic Blockmodeling [Mitrai et al., 2021] reveals a hybrid core-community structure. The scheduling constraints/ variables form the core and the variables/constraints associated with the dynamic optimization problems are assigned in communities. The core and the communities are connected via the transition times θ_{ijk}. Given the structure of the problem, we define ϕ_{ijk} as the value function of a transition from product i to j in slot k, and the dynamic optimization problem for this transition can be written as:

$$\phi_{ijk}\left(\theta_{ijk}\right) = \text{minimize } f_{ijk}^{dyn} \text{ subject to } \ g_{ijk}^{dyn} \leq 0\ (Eq.\ 7),\ \hat{\theta}_{ijk} = \theta_{ijk} : \lambda_{ijk} \tag{9}$$

where λ_{ijk} is the Lagrange multiplier and is equal to the negative of the subgradient of ϕ_{ijk} for $\theta_{ijk} = \bar{\theta}_{ijk}$. The optimization problem can be written as [Geoffrion, 1970]

$$\begin{aligned}&\text{maximize } \Phi_1 - \textstyle\sum_{ijk} Z_{ijk}\eta_{ijk}\\ &\text{subject to } Eq.\ 1-6,\ 8,\ \eta_{ijk} \geq \phi_{ijk}^{v} - \lambda_{ijk}^{v}\left(\theta_{ijk} - \bar{\theta}_{ijk}^{v}\right)\ \forall i, j, k, v \in \mathcal{V} \text{ (Eq. 11)}\end{aligned} \tag{10}$$

We will follow a hybrid multicut GBD approach to solve this problem. The master problem is a Mixed Integer Nonlinear Program solved with Gurobi [Gurobi, 2021] and the subproblems, which are nonlinear programs solved with IPOPT [Wachter and Biegler, 2006], are the dynamic optimization problems only for the transitions that occur. Since η_{ijk} approximates the transition from product i to j in slot k, this approximation will also be valid for other slots. Hence in each iteration, Eq. 11 for a

given i, j is added for all slots. We refer the reader to [Mitrai and Daoutidis, 2021] for a detailed explanation of the algorithm.

4. Adaptive multicut algorithm

The solution of the above problem will provide the production sequence, production times and dynamic transition profiles of the states and manipulated variables. We will assume that at some time t a disturbance affects the system and the value of the state variable is $\bar{x}$ (the predicted value from the initial schedule is x), the demand is d and the inventory of product i is I_i^0. At this point a modified integrated problem must be solved. Specifically, the time horizon is $H - t$ since the system was following the initial schedule during the first t hours. Also, at time t different amounts of each product have been produced, hence the initial inventory if product i, I_i^0, can be nonzero. Finally, in the first slot two transitions can occur. The first is a transition from $\bar{x}$ to the steady state of the new product i (x_i^{ss}) that will be manufactured in the first slot. Once $x = x_i^{ss}$, product i will be produced and then a transition will occur between product i produced in slot 1 and product j produced in slot 2. In order to model this problem, we will define a binary variable $\hat{Z}_i$ which is equal to 1 if a transition occurs from the intermediate state $\bar{x}$ to product i and 0 otherwise. We also define $\hat{\theta}_i$ as the transition time for the aforementioned transition. In order to model the transition in this slot we add the following constraint:

$$\hat{Z}_i = W_{i1} \; \forall i \tag{12}$$

Based on the above constraint a transition from the intermediate state to the steady state of product i is performed only if product i is manufactured in the first slot. We also define $\hat{\eta}_i$ as the approximation of the value function $\hat{\phi}_i$ for the transition from the intermediate state to the steady state of product i. Given these variables, the transition time in the first slot is given by the following constraint:

$$\theta_1^t = \sum_i \hat{Z}_i \, \hat{\theta}_i + \sum_{ij} Z_{ij1} \theta_{ij1} \tag{13}$$

Overall the optimization problem is:

$$\text{maximize } \Phi_1 - \sum_{ijk} Z_{ijk} \eta_{ijk} - \sum_i \hat{Z}_i \, \hat{\eta}_i \tag{14}$$

$$\text{subject to } Eq.1-6,8,12,13, \; \hat{\eta}_i \geq \hat{\phi}_i^v - \hat{\lambda}_i^l \left(\hat{\theta}_i^l - \bar{\hat{\theta}}_i^l \right) \; \forall l \in \mathcal{L}$$

where $\hat{\lambda}_i^l$ is the optimal Lagrangean multiplier for the equality constraint $\hat{\theta}_i = \bar{\hat{\theta}}_i$ and l denotes the iteration number. The hybrid multicut GBD can be used to solve the above problem once a disturbance affects the system. In order to reduce the CPU time further we propose an adaptive solution approach, where for the solution of the integrated problem at time t we add all the cuts evaluated in the previous iterations for the transitions that can occur. This strategy leads to a reduction in the CPU time, since fewer iterations are necessary. However, we must note that the cost associated with the transition from the intermediate state must be approximated every time the integrated problem is solved.

5. Case study

We will assume that the system is an isothermal CSTR where an irreversible reaction occurs $3A \rightarrow B$, and the dynamic behaviour is described by the following equation $\frac{dc}{dt} = \frac{Q}{V}\left(c_{feed} - c(t)\right) - k\, c(t)^3$, where $c\ (mol/L)$ is the concentration, $Q\ (L/hr)$ is the inlet flowrate (manipulated variable) and V, c_{feed}, k are the reactor volume, inlet concentration and reaction constant respectively. First we solve the integrated problem (Eq. 10) to obtain the initial schedule. The optimality gap tolerance is set to 1%. The hybrid multi-cut GBD algorithm solves the problem in 13 CPU seconds and the production sequence is $2 \rightarrow 1 \rightarrow 3 \rightarrow 4 \rightarrow 5$, the value of the objective function is 7.5 10^5 and the production results are presented in Fig. 2.

Table 1 Operating conditions and economic data, scheduling horizon $H = 24$

Product	c^{ss}	Q^{ss}	Prod. rate	Demand	Price	C^{op}	$C^{tr}/10$				
							1	2	3	4	5
1	0.24	200	150	600	200	13	0	10	6	12	15
2	0.2	100	80	550	160	22	15	0	5	8	10
3	0.3	400	278	600	130	35	20	15	0	10	15
4	0.39	1000	607	1200	110	29	90	10	12	0	10
5	0.5	2500	1250	2500	140	25	15	10	15	14	0

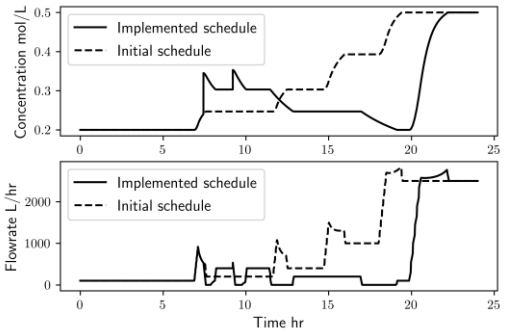


Figure 1 Concentration and inlet flowrate profiles for the nominal and implemented schedule

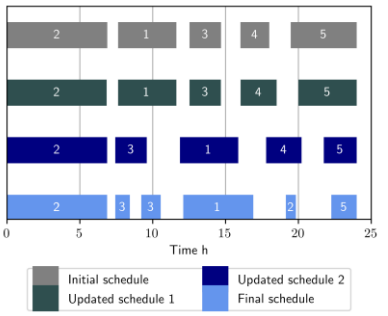


Figure 2 Gantt chart for the initial and updated schedules

First we will consider a case where after 4 hours of operation, the demand of product 4 changes from 1200 to 1500. At this point, 320 mol of product 2 have been produced and the length of the time horizon is 20 hours. Using the adaptive algorithm the integrated problem (Eq. 14) is solved in 1.8 CPU seconds, the value of the objective function is 7.1 10^5 and the updated schedule is presented in Fig. 2 (Update 1). In this case, the value of the objective function is lower, compared to the initial schedule, since more time is dedicated to the production of product 4. Solving the problem hybrid multi-cut GBD algorithm requires 20 CPU seconds. In this case the adaptive algorithm reduces the CPU time by 91 %.

Next, we will assume that after 7.4 hours, a disturbance in the inlet concentration causes the concentration in the reactor to be equal to 0.33 mol/l. At this point, the demand of

product 2 is satisfied and the associated cuts are not considered. The integrated problem is solved in 17 CPU seconds using the adaptive algorithm, the optimal sequence is $3 \rightarrow 1 \rightarrow 4 \rightarrow 5$ (Fig. 2 Update 2) and the value of the objective function is 4.2 10^5. In this case the CPU time is higher than the previous case since the transition from the intermediate state to the different products must be approximated. Also, the value of the objective function is reduced since more time is spent in transitions. Solving the problem with the hybrid GBD multi-cut algorithm requires 33 CPU seconds. Finally, after 9.1 hours of operation the concentration in the reactor is 0.35 (the nominal value is 0.3), the order of product 4 is cancelled, the demand of product 3 changes to 650 and additional 50 mol of product 2 are ordered. The adaptive algorithm solves the problem in 9.6 CPU seconds and the production sequence is $3 \rightarrow 1 \rightarrow 2 \rightarrow 5$ (Fig. 2 Final schedule). The hybrid GBD algorithm requires 20 CPU seconds. The profiles of the concentration and inlet flowrate for the initial and final schedule are presented in Fig. 1.

6. Conclusions

The real time solution of integrated optimization problems is computationally challenging. In this paper, we propose an adaptive multi-cut algorithm which can solve the integrated optimization problem in reduced computational time via the adaptive addition/removal of cuts, which approximate the cost associated with dynamic transitions between products. We consider disturbances in both the control and scheduling and we show that the proposed approach can update the schedule in order to guarantee optimality and feasibility. In the future we intend to apply this algorithm to more complicated continuous systems and batch systems.

References

Charitopoulos, V.M., Papageorgiou, L.G. and Dua, V., 2019, Closed-loop integration of planning, scheduling and multi-parametric nonlinear control. Computers & Chemical Engineering, 122, pp.172-192

Chu, Y. and You, F., 2015, Model-based integration of control and operations: Overview, challenges, advances, and opportunities. Computers & Chemical Engineering, 83, pp.2-20

Daoutidis, P., Lee, J.H., Harjunkoski, I., Skogestad, S., Baldea, M. and Georgakis, C., 2018, Integrating operations and control: A perspective and roadmap for future research. Computers & Chemical Engineering, 115, pp.179-184

Geoffrion, A.M., 1970. Elements of large-scale mathematical programming Part I: Concepts. Management Science, 16(11), pp.652-675.

Risbeck, M.J., Maravelias, C.T. and Rawlings, J.B., 2019. Unification of closed-loop scheduling and control: State-space formulations, terminal constraints, and nominal theoretical properties. Computers & Chemical Engineering, 129, p.106496.

Gurobi Optimization, 2021, Gurobi optimizer reference manual. URLhttp://www.gurobi.com

Mitrai, I., and Daoutidis, P., 2021, A multicut Generalized Benders Decomposition approach for the integration of process operations and dynamic optimization for continuous systems, under review

Mitrai, I., Tang, W. and Daoutidis, P., 2021. Stochastic blockmodeling for learning the structure of optimization problems. AIChE Journal, p.e17415.

Pattison, R.C., Touretzky, C.R., Harjunkoski, I. and Baldea, M., 2017, Moving horizon closed-loop production scheduling using dynamic process models. AIChE Journal, 63(2), pp.639-651

Wächter, A., & Biegler, L. T., 2006, On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Mathematical programming, 106(1), 25-57

Zhuge, J. and Ierapetritou, M.G., 2012, Integration of scheduling and control with closed loop implementation. Industrial & Engineering Chemistry Research, 51(25), pp.8550-8565

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50080-4

Uncertainty Evaluation of Biorefinery Supply Chain's Economic and Environmental Performance Using Stochastic Programming

Yuqing Luo[a], Marianthi Ierapetritou[a*]

[a]*Department of Chemical and Biomolecular Engineering, University of Delaware, 150 Academy Street, Newark, Delaware, 19716, USA*

mgi@udel.edu

Abstract

Life cycle assessment (LCA) and technoeconomic analysis (TEA) are essential tools for evaluating biorefinery performance and designing cost-effective and environmentally friendly supply chains. However, biorefinery operations often suffer from significant temporal and spatial uncertainties, including raw material supply and product demands. This work uses stochastic programming and multi-period planning to design a cost-efficient modular biorefinery supply chain under uncertain demand and material supply. Next, the proposed model is used to design and evaluate modular biorefinery performance in the Baltimore-Wilmington-Philadelphia region. Moreover, the optimization result illustrates the seasonal variability of biomass-based product emission due to demand/supply uncertainty, which cannot be captured by the conventional LCA uncertainty analysis.

Keywords: Biorefinery, modular production, supply chain optimization, stochastic programming, life cycle assessment

1. Introduction

The use of cheap and abundant biomass feedstocks in chemical production is established as a promising alternative to cut greenhouse gas emissions of the chemical industry [Ulonska *et al.*, 2018]. However, biomass feedstocks are often complex mixtures with a considerable amount of lignin, cellulose, and hemicellulose. Thus, the biomass conversion facilities often adopt the so-called biorefinery strategy to use a combination of different reaction units and generate multiple products from each feedstock component. A superstructure optimization framework is commonly used to select the appropriate feedstocks, operating conditions, conversion technologies, and facility locations from the numerous alternatives [You and Wang, 2011].

Nevertheless, parameters used in the biorefinery design often come with considerable uncertainties, such as availability of feedstocks, volatile prices, and experimental yields with intrinsic variations [Baral *et al.*, 2019]. Ignoring such uncertainties often leads to suboptimal or infeasible design [Li *et al.*, 2011]. On the other hand, uncertainty analysis also plays an essential role in LCA. However, most of the LCA uncertainty analyses are limited to sensitivity analysis or the semi-quantitative Pedigree method. The LCA Pedigree approach starts with rating the data reliability, completeness, temporal correlation, geometric correlation, and further technological correlation using indicator scores from 1 to 5. These scores are then transformed to uncertainty factors between 1 to

2 before being assigned as the geometric standard deviation for uncertain parameters, such as the raw material usage [Ciroth *et al.*, 2016]. Although the Pedigree method is a good indicator of data quality, encoding qualitative assessment descriptions into probability distributions inevitably suffers from subjectivity [Henriksson *et al.*, 2015]. Recently, technology choices under parameter uncertainties have been modeled in consequential LCA with optimization tools when multiple technologies exist for manufacturing the same product [Kätelhön *et al.*, 2016]. Nevertheless, this model uses the arithmetic mean of each linear programming problem's solution for LCA calculation, which is less effective than the stochastic programming with recourse actions in capturing the actual supply chain behavior and corrective actions when facing uncertainties [Sahinidis, 2004].

As a promising strategy for biomass supply chain design, modular manufacturing has demonstrated excellent cost reduction potential and extra supply chain flexibility under uncertainties of biomass feedstock availability [Allman *et al.*, 2021]. It has also been shown to benefit from the economy of numbers that reduce capital investment [Bhosekar *et al.*, 2021].

This work utilizes the two-stage stochastic programming and rolling horizon formulation to design a distributed biorefinery supply chain under demand and supply uncertainties. To enhance the process's flexibility, the expansion and movement of modular biomass conversion units are permitted after the initial installation at each production site [Allman *et al.*, 2021]. In addition to the optimal design, the proposed stochastic programming model also provides quantitative insights into the uncertainties of economic and environmental performance using only historical data, which has the potential to replace the Pedigree methods for LCA uncertainty evaluation.

2. Modular Biorefinery Supply Chain Model

2.1. Stochastic programming formulation of distributed biorefinery supply chain

The expansion and module's movement at each site are modelled by the following conservation equation (1):

$$\boldsymbol{n_{j,m,t} = n_{j,m,t-1} + z_{j,m,t} + \sum_{j' \in J'} (v_{j',j,m,t-1} - v_{j,j',m,t-1})} \quad (1)$$

where m is the module types for process units; $n_{j,m,t}$ is the number of unit m at production site j during time period t; $z_{j,m,t}$ is the newly purchased m units at time t at the same site; $v_{j,j',m,t-1}$ is the number of modules moved from site j to j' at time t-1. The material flow in and out of the process site follows the mass balance equation (2).

$$\boldsymbol{\sum_{w \in W} Q_{j,w,p,t} = \sum_{s \in S} \sum_{f \in F} conv_{f,p} \cdot Q_{s,j,f,t}} \quad (2)$$

where $Q_{s,j,f,t}$ is the flowrate of feedstock f from supplier s to site j during time t; $conv_{j,p}$ is the conversion of product p using feedstock f; $Q_{j,w,p,t}$ the flowrates of product p from site j to warehouse w. The production activity cannot exceed the total installed capacity at site j, which is shown in equation (3).

$$\boldsymbol{\sum_{s \in S} \sum_{j \in J} Q_{s,j,f,t} \leq \sum_{j \in J} \sum_{m \in J} c_m \cdot n_{j,m,t}} \quad (3)$$

where c_m is the maximum capacity of unit m. The inventory balance is equation (4).

$$\boldsymbol{I_{w,p,t} = I_{w,p,t-1} + \sum_{j\in J} Q_{j,w,p,t} - \sum_{r\in R} Q_{w,r,p,t}} \quad (4)$$

where $I_{w,p,t}$ is the inventory of product p at warehouse w during time t; $Q_{w,r,p,t}$ is the flowrates of products from warehouse w to market r during time period t, respectively. The objective function is the total expected costs in equation (5):

$$\boldsymbol{\zeta = \sum_{j\in J}\sum_{m\in M} q_m \cdot n_{j,m,0} + \mathbb{E}[\sum_{t\in T}\sum_{m\in M}\sum_{j\in J}(\sum_{j'\in J'} r_{j,j',m}\cdot v_{j,j',m,t} + o_m \cdot x_{j,m,t} + q_m \cdot z_{j,m,t}) + \sum_{t\in T}\sum_{f\in F}\sum_{s\in S}\sum_{j\in J} h_{j,s}\cdot Q_{s,j,f,t} + \sum_{t\in T}\sum_{p\in P}\sum_{r\in R} b_{r,p}\cdot B_{r,p,t} + \sum_{t\in T}\sum_{p\in P}\sum_{w\in W}(\sum_{j\in J} h_{j,w}\cdot Q_{j,w,p,t} + \sum_{r\in R} h_{w,r}\cdot Q_{w,r,p,t} + g_{w,p}\cdot I_{w,p,t})\,]} \quad (5)$$

where q_m is the capital cost and o_m is the operating cost of module m; $r_{j,j',m}$ is the cost of moving unit m from site j to j'; $h_{j,s}$, $h_{j,w}$, and $h_{w,r}$ are the transportation costs of feedstocks from supply s to site j, product from site j to warehouse w and to market r; $g_{w,p}$ is the inventory holding cost of product p at warehouse w and $b_{r,p}$ is the backorder cost for product p at market r; $B_{r,p,t}$ is the unmet demand for product p at market r during time t, which is calculated by equation (6).

$$\boldsymbol{B_{r,p,t} = \delta_{r,p,t} - \sum_{w\in W} Q_{w,r,p,t}} \quad (6)$$

where $\delta_{r,p,t}$ is the uncertain demand of product p at market r during t.

2.2. Life cycle assessment for each scenario of the stochastic programming

Using a cradle-to-gate LCA system boundary, carbon sequestration during plant growth, the emission of production activity, transportation, and upstream emission of backorder are included. Since multiple products are often generated in the biorefinery, the "avoided burden" approach is adopted to account for the credits of by-products by calculating the emissions of their production in standalone processes and deducting them from the initial emission [Anastasopoulou *et al.*, 2020]. The total emission of the biorefinery supply chain is given by equation (7).

$$\boldsymbol{emission = \sum_{t\in T}\sum_{m\in M}\sum_{j\in J}(\sum_{j'\in J'}\alpha_{j,j',m}\cdot v_{j,j',m,t} + \beta_m\cdot x_{j,m,t}) + \sum_{t\in T}\sum_{f\in F}\sum_{s\in S}\sum_{j\in J}\theta_{s,j}\cdot Q_{s,j,f,t} + \sum_{t\in T}\sum_{p\in P}\sum_{r\in R}\eta_{r,p}\cdot B_{r,p,t} + \sum_{t\in T}\sum_{p\in P}\sum_{w\in W}(\sum_{j\in J}\theta_{j,w}\cdot Q_{j,w,p,t} + \sum_{r\in R}\theta_{w,r}\cdot Q_{w,r,p,t}) - \sum_{t\in T}\sum_{p'\in P'}\sum_{r\in R}\eta_{r,p'}\cdot\delta_{r,p',t}} \quad (7)$$

where $\alpha_{j,j',m}$ is the emission of moving module m from j to j'; β_m is the gate-to-gate emission of modular m occurring at production stage; $\theta_{s,j}$, $\theta_{j,w}$, $\theta_{w,r}$ are the emissions when transporting feedstocks f or product p from supplier s to site j, from site j to warehouse w, and from warehouse w to market r; $\eta_{r,p'}$ is the cradle-to-gate emission of the by-product p' bought from market r to satisfy the unmet demand.

2.3. Case study of distributed modular biorefinery operation

The aforementioned model is then applied to a case study of designing distributed biorefinery supply chain in the Baltimore-Wilmington-Philadelphia area, which covers 6 counties in Pennsylvania, 4 counties in Maryland and the New Castle County in Delaware. Figure 1 listed the structure and location of the supply chain containing 7 supply regions, 7 processing sites, 3 warehouses, and 3 market locations.

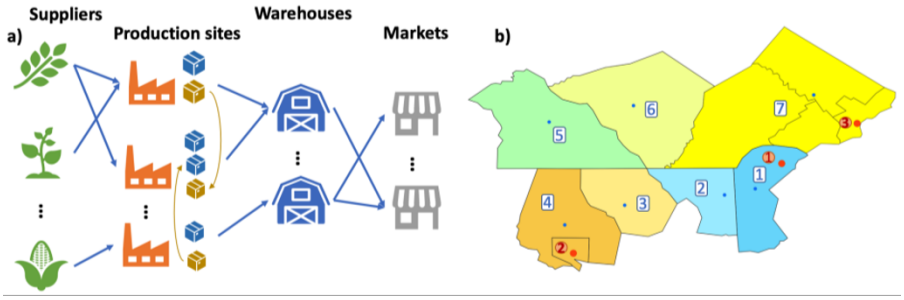


Figure 1. a) structure of modular biorefinery supply chain. b) suppliers, production sites (blue), warehouses, and markets (red) for the case study.

The supply of corn stover, poplar and willow are based on the regional biomass supply data [Langholtz *et al.*, 2016]. The temporal variation in biomass supply is also considered as corn is harvested only from August to November in these three states, while poplar and willow are more stable sources of feedstocks throughout the year. As for the biomass conversion systems, three scales of process units (1.2 kt/y, 2.4 kt/y, and 6 kt/y) for two biomass conversion technologies are available. The molten salt hydrate (MSH) units convert biomass feedstocks to furfural, 5-hydroxymethylfurfural (HMF), and lignin, while the reductive catalytic fractionation (RCF) units further utilize lignin to produce pressure-sensitive adhesives (PSA). The yields of the above technologies are extracted from the Aspen Plus simulation (Aspen Tech, Burlington, MA).

The functional unit of LCA is chosen as 1 kg of PSA supplied to the market. Then, background data of transportation and upstream emissions come from the Ecoinvent v3.3 database and literature results [Athaley *et al.*, 2019, Wernet *et al.*, 2016]. The capital and operating costs are based on the Aspen Plus simulation and technoeconomic analysis for MSH and RCF technologies [Athaley *et al.*, 2019, Bhosekar *et al.*, 2021]. The optimization model is implemented in GAMS 33.1 and cplex 12.10 solver on a computer with Intel Xeon E-2274G CPU @ 4.00GHz 32 GB RAM.

3. Results and discussion

When the weights of conversion units are high, the number of modules at each site increases monotonically over time through capacity expansion (left of Figure 2). This increase is more pronounced after September when the demand for chemicals rises and the supply of corn stover emerges. However, the movement of process unit between processing sites is not observed since the moving cost of heavy modules is relatively high s. On the other hand, when the weights of MSH units are low (right of Figure 2), they are moved around frequently, which accounts for the decrease of Cecil County's MSH units in May and October.

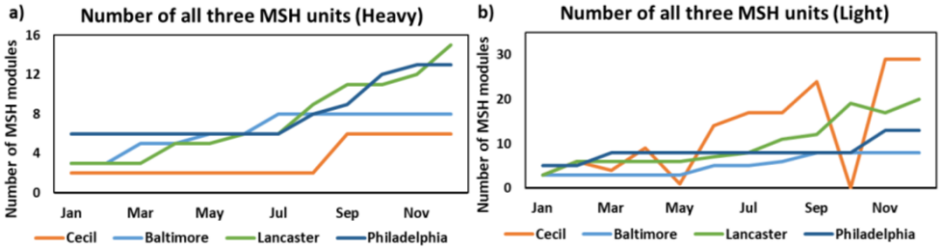


Figure 2. Number of all MSH units in four sites [a) heavy units, b) light units].

The rolling horizon approach implements the first-stage decisions (module moving, expansion, production, transportation, and inventory management) to minimize the expected supply chain cost of all generated future scenarios [Bhosekar *et al.*, 2021]. The predicted costs and GWPs of these possible scenarios in the next period (the shaded area in Figure 3) are compared to the actual cost and GWP of the implemented action (solid lines in Figure 3). The rolling horizon approach underestimated the unmet demand in the next stage, especially before September, when the supply of biomass feedstocks is tight. Thus, the actual realization of the cost and GWP of purchasing PSA from the market (solid lines in Figure 3) is higher than the center of the predicted uncertainty ranges.

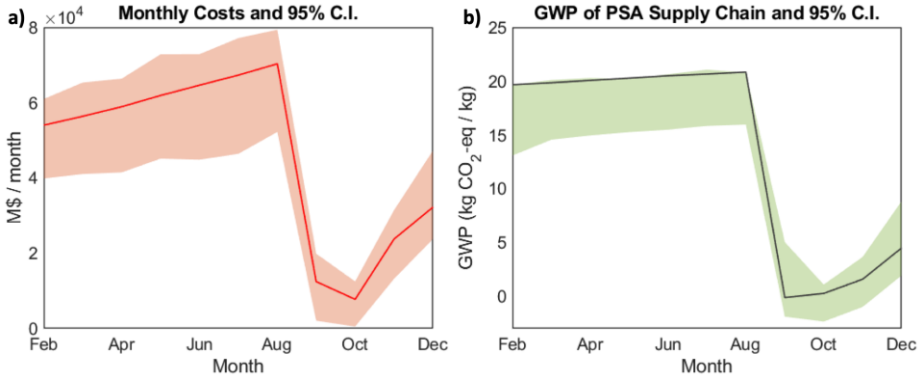


Figure 3. a) predicted 95% confidence interval of uncertain scenarios' cost (orange) and actual cost after uncertainty realization (red line); b) predicted 95% confidence interval of GWP (green) and actual GWP after uncertainty realization (black line).

In September, a large quantity of corn stover supply appears as the corn stover is harvested. More units are added and moved to the sites near the corn stover supply so that as much demand is met as possible. Since buying PSA from conventional oil-based chemical plants for the unmet demand is both expensive (backorder cost) and has high greenhouse gas emission, there is a sudden drop of costs and GWPs when most of the demand is fulfilled with increased production capacities in September. This result demonstrates the flexibility of the distributed modular biorefinery to address the uncertain and shifting supply through gradual expansion and reallocation, rather than overdesigning the capacity in the first place.

Moreover, the traditional Pedigree method applied to LCA is not capable of translating the demand and supply uncertainties into probability distributions because they only indirectly affect the material flows. Therefore, it cannot illustrate the GWP changes throughout the year due to differences in biomass supply. Nevertheless, the proposed stochastic programming method naturally chooses the appropriate supply chain decisions under each scenario, accurately reflecting GWP's response to supply and demand's seasonal variation. Consequently, the empirical distribution of bio-based PSA production emission could then be used as the uncertain input for other LCA studies.

4. Conclusions

In this work, a two-stage stochastic programming model is used to design a distributed biorefinery supply chain that produces value-added chemicals from poplar, willow, and corn stover. Historical supply data are utilized to build the multiperiod scenarios in a case study that contains 3 biomass feedstocks, 7 supply regions, and 3 markets. Next, rolling

horizon approach is utilized to design the supply chain with minimal expected cost. Not only is this model able to design and evaluate the economic and environmental performance of the modular biorefinery supply chain, but also the emission in each scenario of the stochastic programming provides valuable LCA uncertainty information. The LCA uncertainty evaluated in this manner uses actual historical data and represents the rational selection of suppliers and technologies, providing the empirical foundation of uncertainty that the traditional Pedigree method is weak in [Ciroth *et al.*, 2016].

Acknowledgements: The authors are grateful for financial support from the National Science Foundation (NSF GCR CMMI 1934887) and U.S. Department of Energy's RAPID Manufacturing Institute for Process Intensification (DE-EE000788-7.6).

References

Allman, A., Lee, C., Martín, M. and Zhang, Q., (2021) Biomass waste-to-energy supply chain optimization with mobile production modules, *Comput. Chem. Eng.,* **150,** 107326

Anastasopoulou, A., Keijzer, R., Patil, B., Lang, J., van Rooij, G. and Hessel, V., (2020) Environmental impact assessment of plasma-assisted and conventional ammonia synthesis routes, *Journal of Industrial Ecology,* **24,** 1171– 1185

Athaley, A., Annam, P., Saha, B. and Ierapetritou, M., (2019) Techno-economic and life cycle analysis of different types of hydrolysis process for the production of p-Xylene, *Comput. Chem. Eng.,* **121,** 685-695

Baral, N. R., Davis, R. and Bradley, T. H., (2019) Supply and value chain analysis of mixed biomass feedstock supply system for lignocellulosic sugar production, *Biofuels, Bioprod. Bioref.,* **13,** 3, 635-659

Bhosekar, A., Athaley, A. and Ierapetritou, M., (2021) Multiobjective Modular Biorefinery Configuration under Uncertainty, *Ind. Eng. Chem. Res.,* **60,** 35, 12956-12969

Bhosekar, A., Badejo, O. and Ierapetritou, M., (2021) Modular supply chain optimization considering demand uncertainty to manage risk, *AIChE Journal,* **n/a,** n/a, e17367

Ciroth, A., Muller, S., Weidema, B. and Lesage, P., (2016) Empirically based uncertainty factors for the pedigree matrix in ecoinvent, *Int. J. Life Cycle Assess.,* **21,** 9, 1338-1348

Henriksson, P. J. G., Heijungs, R., Dao, H. M., Phan, L. T., de Snoo, G. R. and Guinée, J. B., (2015) Product Carbon Footprints and Their Uncertainties in Comparative Decision Contexts, *PLoS One,* **10,** 3, e0121221

Kätelhön, A., Bardow, A. and Suh, S., (2016) Stochastic Technology Choice Model for Consequential Life Cycle Assessment, *Environ. Sci. Technol.,* **50,** 23, 12575-12583

Langholtz, M. H., Stokes, B. J. and Eaton, L. M., (2016) 2016 Billion-ton report: Advancing domestic resources for a thriving bioeconomy, Volume 1: Economic availability of feedstock, *Oak Ridge National Laboratory, Oak Ridge, Tennessee, managed by UT-Battelle, LLC for the US Department of Energy,* **2016,** 1-411

Li, Z., Ding, R. and Floudas, C. A., (2011) A Comparative Theoretical and Computational Study on Robust Counterpart Optimization: I. Robust Linear Optimization and Robust Mixed Integer Linear Optimization, *Ind. Eng. Chem. Res.,* **50,** 18, 10567-10603

Sahinidis, N. V., (2004) Optimization under uncertainty: state-of-the-art and opportunities, *Comput. Chem. Eng.,* **28,** 6, 971-983

Ulonska, K., König, A., Klatt, M., Mitsos, A. and Viell, J., (2018) Optimization of Multiproduct Biorefinery Processes under Consideration of Biomass Supply Chain Management and Market Developments, *Ind. Eng. Chem. Res.,* **57,** 20, 6980-6991

Wernet, G., Bauer, C., Steubing, B., Reinhard, J., Moreno-Ruiz, E. and Weidema, B., (2016) The ecoinvent database version 3 (part I): overview and methodology, *Int. J. Life Cycle Assess.,* **21,** 9, 1218-1230

You, F. and Wang, B., (2011) Life Cycle Optimization of Biomass-to-Liquid Supply Chains with Distributed–Centralized Processing Networks, *Ind. Eng. Chem. Res.,* **50,** 17, 10102-10127

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50081-6

An Improved Optimization Model for Scheduling of an Industrial Formulation Plant based on Integer Linear Programming

Vassilios Yfantis[a*], Alexander Babskiy[a], Christian Klanke[b], Martin Ruskowksi[a], Sebastian Engell[b]

[a]*Chair of Machine Tools and Control Systems, Technische Universität Kaiserslautern, Gottlieb-Daimler-Str. 42, 67663 Kaiserslautern, Germany*
[b]*Process Dynamics and Operations Group, TU Dortmund University, Emil-Figge-Str. 70, 44227 Dortmund, Germany*

vassilios.yfantis@mv.uni-kl.de

Abstract

This contribution deals with the development of an integer linear programming (ILP) model and a solution strategy for a two-stage industrial formulation plant with parallel production units for crop protection chemicals. Optimal scheduling of this plant is difficult, due to the number of units and operations that must be scheduled while at the same time a high degree of coupling between the operations is present. The problem is further complicated by the presence of optional intermediate storage that leads to alternative branches in the processing sequence of the products. The presented approach is compared to previous ones, namely a mixed-integer linear programming- and a constraint programming-based one. The ILP-based approach exhibits vastly superior computational performance, while still achieving the same solution quality.

Keywords: Batch Process Scheduling, Integer Linear Programming, Decomposition

1. Introduction

The increasing competition on the global market in addition to varying customer demands necessitates an increase in the efficiency and flexibility of production processes. Batch processes offer this kind of flexibility in the case of demand-driven production. A key component to the efficiency of such batch processes is optimal scheduling, i.e., the allocation of limited resources to manufacture several products over a given time horizon. Schedules should be generated in a fast and reliable manner to adapt to varying customer demands. Furthermore, schedules should try to optimize some criterion, e.g., minimizing production time or maximizing profit. These requirements for scheduling can be addressed by optimization methods, like integer and mixed-integer programming. Optimization models can include various constraints that describe the production process while simultaneously optimizing a scheduling objective. The main bottleneck of most optimization models in production scheduling is the computation time. This issue can be handled by applying decomposition techniques, where the scheduling problem is solved in an iterative manner (Elkamel et al., 1997). A straightforward decomposition approach is the iterative scheduling of batches or orders. The realization of the decomposition then mainly depends on the

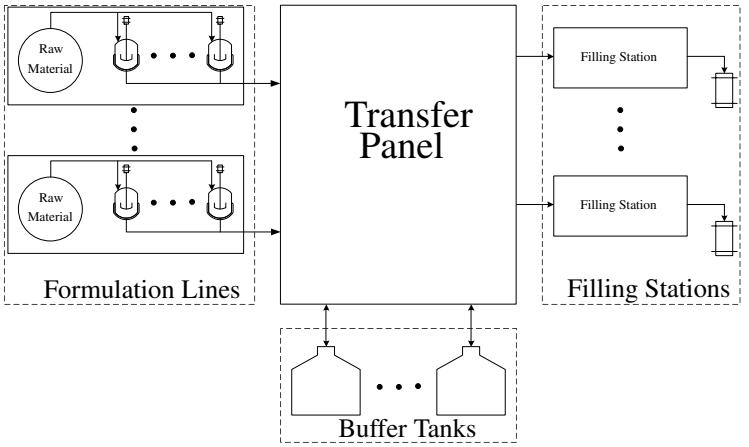


Figure 1: Schematic representation of the industrial formlation plant (Yfantis et al., 2019).

model structure, e.g., whether the model represents time through a time grid (Yfantis et al., 2019) or through precedence relations (Elekidis et al., 2019). In this contribution an efficient integer linear programming model for an industrial formulation plant is presented. A decomposition approach is employed, where orders are scheduled in an iterative fashion, while considering decisions from previous iterations. The solution approach is evaluated on an industrial-scale case study.

2. Indutrial Formulation Plant

The industrial formulation plant is schematically depicted in Figure 1. It can be divided into three parts, the formulation lines, the filling stations, and the buffer tanks. The plant operates in a sequential manner. Intermediate products are produced in the formulation lines and then filled into their final containers by the filling stations. The buffer tanks serve to decouple the two production stages. All sections of the plant are connected by a transfer panel. Each formulation line consists of a raw material pre-processing line, in which the preparation of active ingredients and solvents takes place, and several identical standardization tanks. After the pre-processing, a batch mixing operation takes place in one of the standardization tanks. The standardization tanks are always utilized to their full capacity, i.e., overproduction can occur. As a single pre-processing line feeds multiple standardization tanks, only one batch can start processing in each formulation line at each time point. Furthermore, each order can only be processed on a subset of available formulation lines. After a minimum standardization time, the intermediate product can be filled by a filling station. However, intermediate storage in the standardization tank or in an available buffer tank is also possible. The filling stations operate in a continuous manner, i.e., without an internal storage. A connected standardization or buffer tank is continuously drained by the filling station with an order and station dependent flowrate. Each filling station can only process a subset of available orders. After an operation finishes in any piece of equipment, a sequence dependent changeover time must elapse before the start of the next operation. The filling stations constitute a bottleneck of the process, as they cannot operate during

the night shift, unlike the formulation lines, which operate continuously during the entire time horizon. The scheduling task consists of allocating the batches of the different orders to the standardization tanks and the subsequent filling operations to the filling stations while minimizing the total production time of the schedule. The buffer tanks can be used to decouple the two production stages, while accounting for constraints on the maximum capacity of the tanks. The case study has been investigated by Yfantis et al. (2019) and Klanke et al. (2021b). In the former work, mixed-integer linear programming (MILP) was employed together with a decomposition strategy, and a problem instance identical to the one examined in this paper was solved for a scheduling horizon of one week. In Klanke et al. (2021b) the same problem instance was solved by combining constraint programming (CP) and a moving-horizon strategy, outperforming the previous MILP formulation. Furthermore, different case studies for the same formulation plant were solved in Klanke et al. (2021a) using a heuristics-assisted genetic algorithm.

3. Solution Approach

3.1. Integer Linear Programming Model

In this section, the proposed integer linear programming model is presented. Since the model is very complex, this sections only focuses on some key constraints and variables, as well as on the objective function. The goal is to schedule the set of orders $\mathcal{I}$ on the available machines J. The machines are divided into the standardization tanks of the formulation lines $\mathcal{J}^{\text{FL}}$ and the filling stations $\mathcal{J}^{\text{FS}}$. The machines that can process order i are denoted by $\mathcal{J}_i$. The available buffer tanks are modeled by the set $\mathcal{B}$. The time horizon is discretized into equidistant time points $\mathcal{T}$. Some of the key constraints are shown in Eq. (1) - (6). The binary variable R_{ijt} indicates that a batch of order i is released from standardization tank j at time t. Eq. (1) guarantees the satisfaction of demand D_i, where cap_j is the batch size in tank j. The binary variable is set to one once the tank has been emptied. This is modeled by Eq. (2), where $E^{\text{fill}}_{ijj't}$ is a binary variable indicating the end of filling of a batch of order i from standardization tank j by filling station j' at time t and $Ref_{ijbj't}$ is a binary variable indicating a refilling of this batch into buffer tank b, in order to later be filled by filling station j'. A batch can be stored inside a standardization tank prior to its release. Intermediate storage of a batch of order i in standardization tank j at time t is indicated by the binary variable L_{ijt}. This variable is updated by Eq. (3), where E_{ijt} is a binary variable representing the end of a standardization operation, $S^{\text{fill}}_{ijj't}$ models the start of a filling operation from standardization tank j by filling station j'. When a standardization tank j is processing a batch of order i at time t the binary variable X_{ijt} is active. It is updated through the starting (S_{ijt}) and ending (E_{ijt}) binary variables in Eq. (4). Processing in the filling station is modeled by similar constraints. An important aspect of the scheduling problem is the modeling of the buffer balances. Instead of modeling stored quantities in the buffer tanks, Eq. (5) models the time intervals necessary to empty buffer tank b, containing order i by filling station j, if filling starts at time t through the integer variable I_{ibjt}. This variable is updated at every time step, using the parameter $p_{ijj'}$, which is equal to the number of time points needed to fill a batch of order i from standardization tank j' by filling station j and the binary variable Y_{ibjt}, indicating that an order i is filled from buffer b by filling station j at time t. Eq.

(6) ensures that the buffer level does not exceed its maximum capacity by considering an upper bound on the required filling time. Further constraints include the changeovers in the different machines, modeled in a similar fashion to Eq. (4), also using binary variables for their start, end, and processing. The processing times are modeled by linking the binary variables for the start and end of an operation through their time indices. The objective of the optimization problem is modeled by Eq. (7). In the first term the starting and end times of the filling operations are minimized. The remaining terms serve to minimize idle times, which occur in a makespan minimization due to multiple symmetric solutions. The second term penalizes the use of the buffer tanks. The scaling parameter w_{ij} is equal to the mean filling time of batches of order i by filling station j. The third term discourages intermediate storage in the standardization tanks if it is unnecessary.

$$\sum_{t\in\mathcal{T}}\sum_{j\in\mathcal{J}_i^{\text{FL}}} R_{ijt}\cdot cap_j \geq D_i,\ \forall i\in\mathcal{I} \tag{1}$$

$$R_{ij,t+1} = \sum_{j'\in\mathcal{J}_i^{\text{FS}}} E^{\text{fill}}_{ijj't} + \sum_{b\in\mathcal{B}}\sum_{j'\in\mathcal{J}_i^{\text{FS}}} Ref_{ijbj',t+1},\ \forall i\in\mathcal{I}, j\in\mathcal{J}_i^{\text{FL}}, t\in\mathcal{T}\backslash\{|\mathcal{T}|\} \tag{2}$$

$$L_{ij,t+1} = L_{ijt} + E_{ijt} - \sum_{j'\in\mathcal{J}_i^{\text{FS}}} S^{\text{fill}}_{ijj',t+1} - \sum_{b\in\mathcal{B}}\sum_{j'\in\mathcal{J}_i^{\text{FS}}} Ref_{ijbj',t+1}, \quad \forall i\in\mathcal{I}, j\in\mathcal{J}_i^{\text{FL}}, t\in\mathcal{T}\backslash\{|\mathcal{T}|\} \tag{3}$$

$$X_{ij,t+1} = X_{ij,t} + S_{ij,t+1} - E_{ijt},\ \forall i\in\mathcal{I}, j\in\mathcal{J}_i^{\text{FL}}, t\in\mathcal{T}\backslash\{|\mathcal{T}|\} \tag{4}$$

$$I_{ibj,t+1} = I_{ibjt} + \sum_{j'\in\mathcal{J}_i^{\text{FL}}} Ref_{ij'bj,t+1}\cdot p_{ijj'} - Y_{ibjt},\ \forall i\in\mathcal{I}, b\in\mathcal{B}, j\in\mathcal{J}_i^{\text{FS}} \tag{5}$$

$$I_{ibjt} \leq cap_{ibj},\ \forall i\in\mathcal{I}, b\in\mathcal{B}, j\in\mathcal{J}_i^{\text{FS}}, t\in\mathcal{T} \tag{6}$$

$$\min\ \frac{1}{2}\cdot\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}_i^{\text{FL}}}\sum_{j'\in\mathcal{J}_i^{\text{FS}}}\sum_{t\in\mathcal{T}}\left(E^{\text{fill}}_{ijj't} + S^{\text{fill}}_{ijj't}\right) + \sum_{i\in\mathcal{I}}\sum_{b\in\mathcal{B}}\sum_{j\in\mathcal{J}_i^{\text{FS}}}\sum_{t\in\mathcal{T}}\frac{1}{w_{ij}}\cdot Y_{ibjt}\cdot t + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}_i^{\text{FL}}}\sum_{t\in\mathcal{T}} L_{ijt} \tag{7}$$

3.2. Decomposition

Due to its size and complexity the model cannot be solved in a monolithic fashion. To generate feasible schedules in a time efficient manner an order decomposition approach, similar to Yfantis et al. (2019), is employed. The orders are scheduled iteratively while preventing overlaps through constraints. These infeasible allocations can easily be identified since each machine possesses an active binary variable at each time point where an operation is being performed, instead of just using a single binary variable for the start of an operation. The night shifts of the filling stations are modeled in a similar way. In contrast to the approaches in Yfantis et al. (2019) and Klanke et al. (2021b) no batch decomposition is needed, as the model can schedule orders that consist of a large number of

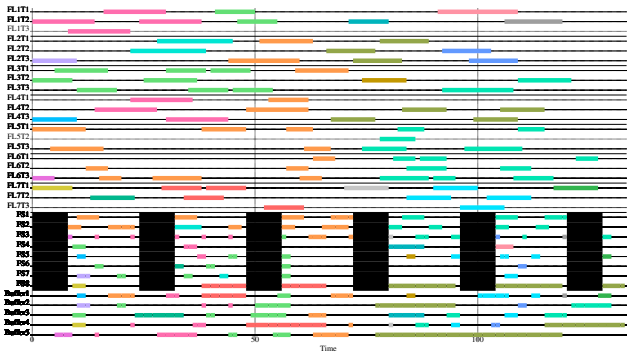


Figure 2: Gantt chart of the schedule generated with the proposed ILP-based approach.

batches efficiently. Furthermore, no two-step optimization approach is necessary, as the operations are already shifted to the left through the chosen objective function. Lastly, the time horizon is shifted to the end of the next day from the current makespan. If the subproblem is infeasible the time horizon is shifted by an additional day until a solution is found.

4. Results

The presented ILP-based solution approach was evaluated on the case study presented in Yfantis et al. (2019) and Klanke et al. (2021b). The setup consists of 7 formulation lines, each containing 3 standardization tanks, 8 filling stations and 5 buffer tanks. In total, 20 orders of different quantities, resulting in 78 batches are scheduled. A time horizon of one week, divided into 1-hour intervals, is considered. The solution approach was implemented in the programming language Julia (Bezanson et al., 2017). The ILP was solved using Gurobi on a Desktop PC (AMD Ryzen 5 3600 6-Core Processor @3.6 GHz). The subproblems were all solved to a 0 % optimality gap. The generated Gantt chart is depicted in Figure 2. It represents the batches on each standardization tank of the formulation lines, separated by the black solid lines, the filling stations, and the buffer tanks. Furthermore, the night shifts of the filling stations are illustrated as black regions. A makespan of 133 h is obtained, which is equal to the results obtained by the previous solution approaches. However, the benefit of the proposed ILP-based approach can be seen in the required computation time (cf. Table 1). The superior performance of the ILP model is further underlined by the fact, that no batch-based decomposition is needed. Instead, only an order decomposition is performed, so that a single subproblem can require scheduling a large number of batches, which would render it intractable for the previous approaches. The computation time is further enhanced by the lack of a two-step optimization approach, due to the chosen objective function, which results in fewer idle times than a makespan minimization. The superior performance can be attributed to the multiple active binary variables for a given schedule. In the MILP-based approach of Yfantis et al. (2019) binary variables only indicate the start of an operation, resulting in far less active binary variables. The tightly constrained active binary variables of the ILP aid the search procedure of the solver.

Table 1: Comparison between different solution approaches for scheduling of the industrial formulation plant.

Model	MILP (Yfantis et al., 2019)	CP (Klanke et al., 2021b)	ILP
Makespan	133 h	133 h	133 h
Computation Time	38 min	23 min	51 s

5. Conclusion and Outlook

This work presented a novel ILP-based formulation for the scheduling of an industrial formulation plant. In contrast to previous approaches, the model only employs integer variables, which greatly enhances its computational performance. Instead of minimizing the makespan, an objective function that discourages idle times is formulated, eliminating the need for a two-step optimization approach. The structure of the model enables a monolithic optimization without running into memory limitation issues. However, then the solution times are prohibitive for a real application. Nevertheless, in future work a monolithic optimization can be performed on specialized hardware to provide a reference for the decomposition approaches and other solution methods.

Acknowledgements

This work was partially funded by the European Regional Development Fund (ERDF) in the context of the project OptiProd.NRW (https://www.optiprod.nrw/en).

References

J. Bezanson, A. Edelman, S. Karpinski, V. B. Shah, 2017. Julia: A fresh approach to numerical computing. SIAM Review 59 (1), pp. 65–98.

A.P. Elekidis, F. Corominas, M.C. Georgiadis, 2019. Production Scheduling of Consumer Goods Industries. Industrial & Engineering Chemistry Research 58, pp. 23261-23275.

A. Elkamel, M. Zentner, J.F. Pekny, G.V. Reklaitis, 1997. A Decomposition Heuristic for Scheduling the General Batch Chemical Plant. Engineering Optimization 28 (4), pp. 299-330.

C. Klanke, D. Bleidorn, C. Koslowski, C. Sonntag, S. Engell, 2021a. Simulation-based scheduling of a large-scale industrial formulation plant using a heuristics-assisted genetic algorithm. GECCO '21: Proceedings of the Genetic and Evolutionary Computation Conference Companion, pp. 1587–1595.

C. Klanke, D. Bleidorn, V. Yfantis, S. Engell, 2021b. Combining Constraint Programming and Temporal Decomposition Approaches - Scheduling of an Industrial Formulation Plant. Lecture Notes in Computer Science, Vol. 12735, pp. 133-148.

V. Yfantis, T. Siwczyk, M. Lampe, N. Kloye, M. Remelhe, S. Engell, 2019. Iterative Medium-Term Production Scheduling of an Industrial Formulation Plant. In: Computer Aided Chemical Engineering, Vol. 46, pp. 19–24.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50082-8

Optimal Sourcing, Supply and Development of Carbon Dioxide Networks for Enhanced Oil Recovery in CCUS Systems

Demian J. Presser[a,b], Vanina G. Cafaro[a,b], Diego C. Cafaro[a,b,*]

[a]*INTEC (UNL-CONICET), Guemes 3450, 3000 Santa Fe, Argentina*
[b]*Fac. de Ing. Química (UNL), Santiago del Estero 2829, 3000 Santa Fe, Argentina*
dcafaro@fiq.unl.edu.ar

Abstract

Carbon capture, utilization and storage (CCUS) is one of the most promising technologies for mitigating anthropogenic CO_2 emissions. The deployment of CUUS typically requires heavy capital investments that need to be offset by the benefits of carbon utilization. Carbon dioxide enhanced oil recovery (EOR-CO_2) consists on injecting large quantities of CO_2 into mature oil reservoirs to boost hydrocarbon recovery. It is one of the most effective ways to bring economic viability to CCUS projects, also providing the means for the geological sequestration of CO_2. The planning of CCUS coupled to EOR comprises interrelated decisions aiming to maximize oil production and CO_2 sequestration. In this work, we propose a novel optimization approach to allocate CO_2 from capture sources to oil fields according to their potential, and determine how these reservoirs should be developed over time. To this end, we seek for the optimal design of pipeline networks, as well as the injection plan in each reservoir according to the CO_2 availability. The results show that the coordinated operation of EOR-CO_2 in several oil reservoirs is crucial to the success of a CCUS project. An illustrative case study of 3 reservoirs and 2 sources (power plants) is presented. The production strategy yields up to 40% reduction in carbon emissions from the power plants and a positive net present value of 74 million USD in five years.

Keywords: EOR, Supply Chain, Carbon Dioxide, Optimization, CCUS, MINLP

1. Introduction

Greenhouse gas emissions (GHG) are expected to have their second largest increase in history due to global economies recovering from the COVID pandemic (IEA, 2021). Reducing CO_2 emissions is an indispensable requirement to achieve the goals of the Paris Agreement (2015), and carbon capture, utilization and storage (CCUS) systems are the most promising option to meet the targets. CCUS refers to the set of techniques to capture large amounts of CO_2 emissions from flue gas of fossil power plants and industrial processes, to subsequently transport CO_2 to utilization points, and finally guarantee its permanent sequestering. Although CCUS systems have been recognized for decades as one of the most promising technologies in the pursuit of net-zero emissions, progress has been relatively slow due to economic drawbacks (Nuñez and Moskal, 2019). Carbon dioxide enhanced oil recovery (EOR-CO_2) is a production technique consisting on injecting large quantities of CO_2 into mature oil reservoirs to extend their economic lifespan. CO_2 facilitates the displacement of the residual oil that remains trapped after primary and secondary production by reducing interfacial and

surface tension, swelling oil, decreasing viscosity and improving mobility ratio (Lake, 1989). Moreover, EOR-CO_2 provides an efficient path for permanently sequestering massive amounts of GHG. Numerous studies conclude on the capability of EOR- CO_2 to achieve net zero (and even negative) emissions (Cuéllar and Azapagic, 2015).

So far, EOR-CO_2 has allowed large-scale, economic deployment of CCUS (Mavar et al., 2021). Nevertheless, proper planning of CCUS-EOR projects is required to address the allocation of CO_2 over time, the design of CO_2 pipeline networks, and the balance between maximizing oil recovery and GHG sequestration. Several contributions related to CCUS design and EOR planning have been published in recent years, although none of them tackle both problems in an integrated manner (Tapia, 2018). Turk et al. (1987) present one of the first formulations for the optimal allocation of CO_2, assuming a fixed economic value for its use. Middleton and Bielicki (2009) propose a mixed integer linear model (MILP) for the design of CO_2 capture and sequestration networks, setting a target value for the amount of CO_2 to store. Tan et al. (2013) present a multiperiod MILP to find the best matches between sources and sinks, accounting for injection rates and time windows. Tapia et al. (2016) solve the optimal allocation of CO_2 between a power plant and multiple reservoirs by means of a general scheduling framework. A fixed decreasing production yield is assumed, with no more details on the reservoir depletion behavior. On the other hand, forecasting production is a key feature for the development of optimization models. Capacitance Resistance Models (CRM) predict the fraction of the flow injected into a well that is conveyed to another well, from historical data (Yousef et al., 2005). Coupling CRM and fractional flow models allows high-level optimization of the reservoir development and preliminary assessment of the field production. Eshraghi et al. (2016) propose different heuristic approaches to establish the best injection strategy for a set of wells in a reservoir. Tao and Bryant (2015) take advantage of CRM to optimize CO_2 sequestering in an aquifer accounting for different injection rates. In 2021, Presser et al. combine CRM and a fractional flow approach to optimize polymer flooding production strategies in mature oil fields.

In this work, we propose the first mathematical programming approach to optimally plan the design and development of CCUS-EOR projects in an integrated fashion. Decisions addressed by the model include the allocation of CO_2 from sources to EOR reservoirs, pipeline network design and oil field development strategies. An illustrative case study is presented to show the potential of the tool and draw conclusions.

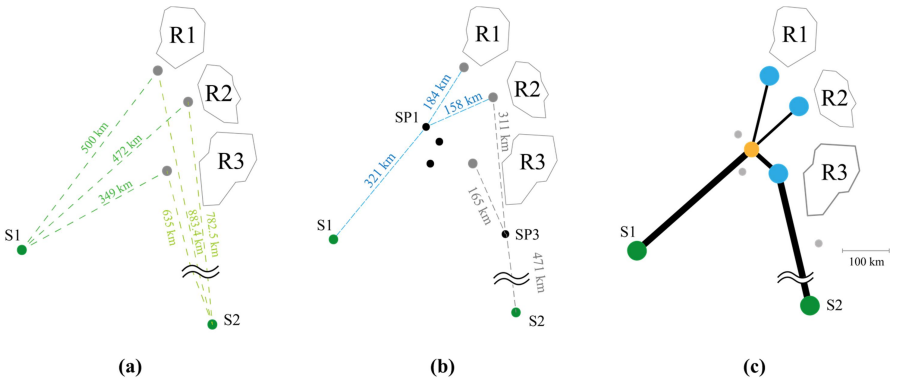


Figure 1 – (a) Layout and distances between sources S and reservoirs R. (b) Distances between nodes and intermediate points SP. (c) Best configuration found.

2. Problem Statement

The problem addressed in this work can be stated as follows: Given n CO_2 sources (e.g., power plants) venting flue gas at a certain rate, m depleted reservoirs to be developed through CO_2-EOR, potential locations for CO_2 pipelines (see Figure 1), reservoirs characterization in terms of selectable wells to be operated, forecasted decline curves and connectivities between the wells; we aim to optimally determine: (a) the allocation of CO_2 flows between sources and reservoirs, (b) the pipeline network design, (c) the wells to be operated as injectors and producers in each field, (d) the timing for operating each of these wells, and (e) the CO_2 injection rates in order to maximize the net present value of the CCUS-EOR project. The objective function accounts for the benefits from crude oil sales, as well as from CO_2 sequestration.

3. Mathematical Formulation

The formulation integrates two well-known reservoir prediction models: CRM and Gentil fractional flow (GFF) (Gentil, 2005). CRM allows for the characterization of the reservoir by assigning connectivities and time constants to every pair of wells based on history matching, while GFF assesses the production decline for each producer with respect to the cumulative injection of CO_2 reaching its drainage volume through a semi-empirical power-law function. We assume that connectivities and time constants have been inferred from secondary production, and the decline curve for each producing well is also known. The mathematical formulation is based on the set $t \in T$ representing time periods (typically semesters or years). Let $s \in S$ stand for CO_2 sources (e.g., power plants) with a known maximum supply rate $co2r_{s,t}$ (Mt/y), and $r \in R$ be reservoirs for EOR-CO_2 exploitation. Elements $p \in P$ represent pipes of different diameters and flow capacities, and $sp \in SP$ stand for intermediate points between s and r where pipelines can be joined or branched. Finally, $sc \in SC$ account for production schemes, comprising subsets of active injection wells i and producers j in reservoir r.

Eqs. (1) to (4) calculate the volume of CO_2 received by well j from the injection in well i, according to the selected production scheme and connectivities. Parameter $fs_{i,j,sc}$ stands for the connectivity between i and j under production scheme sc. The 0-1 variable $xsc_{r,sc,t}$ equals 1 if sc is the scheme selected for time t in reservoir r (0 otherwise).

$$QRE_{i,j,t} = \sum_{sc \in SC_{i,j}} QINJ'_{i,t,sc}\, fs_{i,j,sc} \quad \forall r, i \in I_r, j \in J_r, t \tag{1}$$

$$QINJ'_{i,t,sc} \leq \sum_s co2r_{s,t}\ \ xsc_{r,sc,t} \qquad \forall r, sc, i \in I_r \cap I_{sc}, t \tag{2}$$

$$\sum_{sc \in SC_i} QINJ'_{i,t,sc} = QINJ_{i,t} \qquad \forall r, i \in I_r, t \tag{3}$$

$$\sum_{sc \in SC} xsc_{r,sc,t} \leq 1 \qquad \forall r, t \tag{4}$$

For simplicity, if under the production scheme sc a producing well is inactive, flows are proportionally redistributed among the remaining wells, as in Eq. (5). Connectivity factors are defined in advance and can be adjusted following any other criterion.

$$fs_{i,j,sc} = f_{i,j} / \sum_{j \in J_{sc}} f_{i,j} \qquad \forall r, i \in I_r, j \in J_r \tag{5}$$

According to DFF, the productivity of each producing well decreases as a function of the cumulative volume of CO_2 received in its drainage volume, as modelled by Eqs. (6) and (7). $CGIA_{i,t}$ is a continuous variable accounting for the cumulative amount of CO_2 that has reached the drainage volume of j up to time t, $F^o_{j,t}$ stands for the fraction of oil in the production flow, and $Q^o_{j,t}$ is the oil production rate from j during time step t.

$$QTR_{j,t} = \sum_{i \in I_r} QRE_{i,j,t} \;; \qquad CGIA_{j,t} \geq CGIA_{j,t-1} + QTR_{j,t} \qquad \forall r, j \in J_r, t \qquad (6)$$

$$F^o_{j,t} \leq \frac{1}{1+\alpha_j \; CGIA_{j,t}{}^{\beta_j}} \;; \qquad Q^o_{j,t} \leq F^o_{j,t} \; QTR_{j,t} \qquad \forall r, j \in J_r, t \qquad (7)$$

Eq. (8) identifies the conversion of well i from production to injection mode at time t through the binary $ycv_{i,t}$. Note that the values of $xpr_{i,t}$ and $xinj_{i,t}$ can be directly calculated from $xsc_{r,sc,t}$ (the selected production scheme).

$$ycv_{i,t} \geq xinj_{i,t} - xpr_{i,t-1} \qquad \forall r, i \in I_r, t \qquad (8)$$

The supply of CO_2 to active reservoirs depends on injection decisions. Eq. (9) shows how CO_2 demand can be met from the selected sources ($PCO2_{r,t}$) or from recycling ($RCO2_{r,t}$), as in Eq. (10). In these volume balances, rec_r computes the proportion of the non-oil production stream that can be reconditioned and reinjected.

$$\sum_{i \in I_r} QINJ_{i,t} \leq PCO2_{r,t} + RCO2_{r,t} \qquad \forall r, t \qquad (9)$$

$$RCO2_{r,t} \leq rec_r \left(\sum_{j \in J_r} QTR_{j,t} - Q^o_{j,t} \right) \qquad \forall r, t \qquad (10)$$

Connecting sources with reservoirs through pipelines is a model decision, allowing for the allocation of CO_2 flows. The reservoirs can be fed directly from the sources or through intermediate nodes, as in Eq. (11). Eqs. (12) to (14) stand for the selection of pipeline diameters/capacities (psl_p) for the links *s-r*, *sp-r* and *s-sp*, respectively. Finally, Eq. (15) imposes the volume balance at the intermediate nodes.

$$PCO2_{r,t} = \sum_{s \in S_r} DCO2_{s,r,t} + \sum_{sp \in SP_r} ICO2_{sp,r,t} \qquad \forall r, t \qquad (11)$$

$$DCO2_{s,r,t} \leq \sum_{p \in P} psl_p \; ysr_{s,r,p} \qquad \forall s, r, t \qquad (12)$$

$$ICO2_{sp,r,t} \leq \sum_{p \in P} psl_p \; yspr_{sp,r,p} \qquad \forall sp, r, t \qquad (13)$$

$$FCO2_{s,sp,t} \leq \sum_{p \in P} psl_p \; yssp_{s,sp,p} \qquad \forall s, sp, t \qquad (14)$$

$$\sum_{r \in R} ICO2_{sp,r,t} = \sum_{s \in S} FCO2_{s,sp,t} \qquad \forall sp, t \qquad (15)$$

Where $ysr_{s,r,p}$, $yspr_{sp,r,p}$ and $yssp_{s,sp,p}$ are 0-1 variables. On the other hand, Eq. (16) estimates the amount of CO_2 produced in the sources that is not used for EOR, and therefore is economically penalized in the objective function.

$$\sum_r DCO2_{s,r,t} + \sum_{sp} FCO2_{s,sp,t} + CO2V_{s,t} = co2r_{s,t} \qquad \forall s, t \qquad (16)$$

The objective of this mixed integer nonlinear programming (MINLP) model seeks to maximize the net present value (Eq. 17), where r is the discount rate. Incomes (INC_t) are determined by the oil prices times the predicted production over t, adding credits for CO_2 sequestering. In turn, capital investments ($CAPEX_t$) comprise drilling and completion costs for new wells, conversion and shut-in charges. We also include a fixed term ($SCAPEX$) for pipeline and EOR facilities construction at the initial time. Operating expenditures ($OPEX_t$) involve injection and production costs for active wells, CO_2 acquisition, pumping, conditioning and recycling, and produced flows carrying and processing charges. Finally, we add a penalty cost term for not using CO_2 from sources ($SPEX_t$). Note that by Eqs. (6) and (7), the MINLP yields a nonconvex relaxation.

$$\text{Max } z = \sum_t \; (1+r)^{t-1} (INC_t - CAPEX_t - OPEX_t - SPEX_t) - SCAPEX \qquad (17)$$

4. Results

An illustrative case study is proposed to show the capabilities of the model. Two sources, three reservoirs and four potential split points are addressed, as shown in Figure 1. For clarity, only some distances are presented although all other connections are also possible. The arrangement of wells in each reservoir and their connectivities are displayed in Figure 2. Each of the potential producing wells has a pair of parameters α_j and β_j defining how fast oil productivity declines. Regarding CO_2 sourcing and network design, three pipeline sections are assessed, with transportation capacities ranging from 2 to 8 kt/day. The time horizon is discretized into 10 semesters. Oil price is assumed to increase over time and CO_2 acquisition costs are expected to decrease due to scalability and expertise in the capture and conditioning process. CO_2 recycling rates are set to 50%, 60% and 45% for reservoir 1, 2 and 3, respectively.

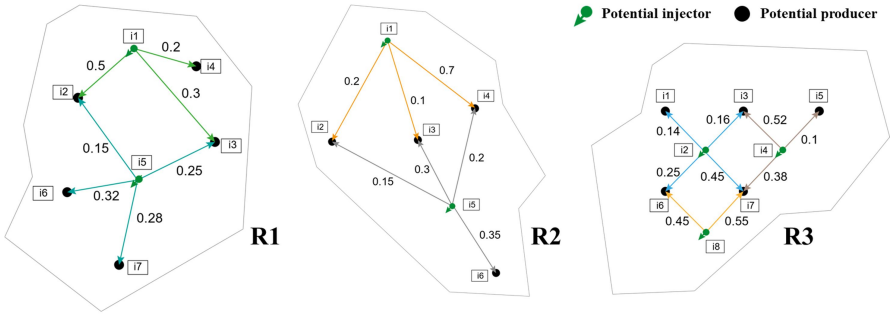


Figure 2 – Arrangement of wells in each reservoir and connectivity coefficients.

The nonconvex MINLP model is implemented in GAMS and solved through DICOPT, using CONOPT4 and CPLEX for NLP and MIP subproblems respectively. The model comprises 4,561 eqs, 1,417 0-1 vars and 2,231 cont. vars. The algorithm reaches a solution amounting to 74 MMUSD as NPV in 25 minutes of CPU. Given that DICOPT does not guarantee the global optimality, a tailored MILP relaxation is proposed to estimate the optimality gap. This relaxation yields results 10% higher than the solution obtained with DICOPT. The results of the source allocation and supply chain design are shown at the right of Figure 1. The best-found configuration suggests that S1 must supply the three reservoirs simultaneously through the splitting point SP2, while S2 should only feed R3 through a mainline of capacity *pl3*. Another mainline of capacity *pl3* connects S1 with SP2, while two pipes *pl1* connect SP2 with R1 and R2, and a pipe of capacity *pl2* feeds R3. In addition, the development strategies for injectors and producers for each of the reservoirs are presented in Figure 3. This figure shows that the most promising production schemes are exploited earlier, requiring less CO_2 for high production, while the wells with high potential but low initial connectivity (e.g., I3 in R2) are isolated to increase flow to them over the end of the horizon. Other schemes with good potential but higher CO_2 requirements are also tapped in later periods.

5. Conclusions

A novel MINLP formulation has been developed for the integration of CO_2 supply and EOR production planning decisions for the optimization of CCUS-EOR initiatives. Results suggest that the coordinated planning of several reservoirs is critical for the economic viability of these projects. The simultaneous optimization of production

strategies may allow exploiting the most promising regions of each reservoir with no need to delay the start of EOR projects by restricting themselves to the need for rigid supplies. The model also facilitates the evaluation of large-scale CCUS systems involving multiple sources and sinks in a generalized framework, providing further guidance on the environmental impacts of the initiatives. Results for an illustrative case show an economic benefit of 74 MMUSD, using and finally sequestering 40% of CO_2 emissions from the sources (43 Mt of CO_2 over 5 years). Finally, accounting for the possibility of delaying investments in processing facilities in future works may bring further economic benefits to the strategies. In addition, addressing uncertainties of oil and CO_2 prices and the productive behaviour of the wells is another possible extension.

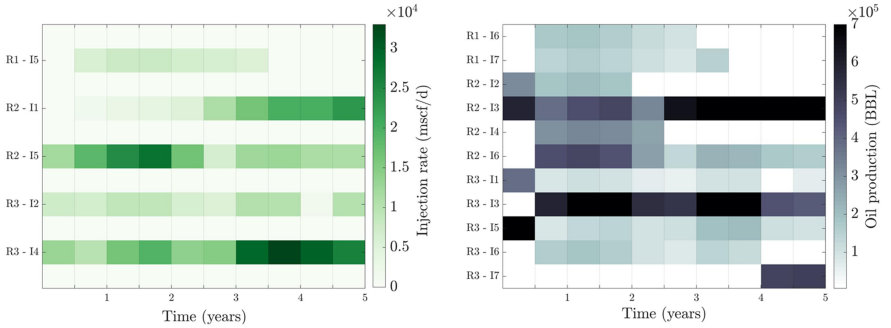


Figure 3 – Injection rate strategy and expected production in the best solution found.

6. References

R. M. Cuéllar-Franca, A. Azapagic, 2015, Carbon capture, storage and utilisation technologies: A critical analysis and comparison of their life cycle envir. impacts, J. of CO_2 Util., 9, 82–102.

S. E. Eshraghi, M. Rasaei, S. Zendehboudi, 2016, Opt. of mis. CO_2 EOR and strg. using heur. meth. comb. w. cap/res and Gentil fract. flow models, J. of Nat, Gas Sci & Eng, 32, 304–318.

P. H. Gentil, 2005, The Use of Multilinear Regression Models in Patterned Waterfloods: Physical Meaning of the Regression Coefficients, MSc thesis, University of Texas, Austin, TX, US.

L. W. Lake, 1989, Enhanced Oil Recovery, Prentice Hall: Englewood Cliffs, NJ, US.

R. S. Middleton, J. M. Bielicki, 2009, A compr. carbon capt. and strg. infrast. model, Energy Proc., 1(1), 1611–1616.

K. Novak Mavar, N. Gaurina-Međimurec, L. Hrnčević, 2021, Significance of Enhanced Oil Recovery in Carbon Dioxide Emission Reduction. Sust., 13(4), 1800.

V. Núñez-López, E. Moskal, 2019, Pot. of CO_2-EOR for Near-Term Decarb., Front. Clim., https://doi.org/10.3389/fclim.2019.00005.

D. J. Presser, V. G. Cafaro, D. C. Cafaro, 2021, Optimal Prod. Strategies for the Development of Mature Oil Fields through Polymer Flooding, Ind. & Eng. Chem. Res., 60(28), 10235–10253.

R. R. Tan, K. B. Aviso, S. Bandyopadhyay, D. K. Ng, 2013, Opt. sour.-sink match. in carb. capt. and strg. sys. with time, inj. rate, and cap. const., Envir. Prog. & Sust. Energy, 32(2), 411–416.

Q. Tao, S. L. Bryant, 2015, Optimizing Carbon Sequestration With the Capacitance/Resistance Model, SPE Journal, 20(05), 1094–1102.

J. F. D. Tapia, J. Y. Lee, R. E. Ooi, D. C. Foo, R. R. Tan, 2016, Optimal CO_2 allocation and scheduling in enhanced oil recovery (EOR) operations. Applied Energy, 184, 337–345.

J. F. D. Tapia, J. Y. Lee, R. E. Ooi, D. C. Foo, R. R. Tan, 2018, A rev. of opt. and dec.-making models for the plan. of CO2 capt., util. and strg. (CCUS) sys., Sust. Prod. and Cons., 13, 1–15.

G. A.Turk, T. B. Cobb, D. J. Jankowski, A. M. Wolsky, F. T. Sparrow, 1987, CO_2 Transport: A new application of the assignment problem, Energy, 12(2), 123-130.

A. A. Yousef, P. H. Gentil, J. L. Jensen, L. W. Lake, 2005, A Capacit. Model To Infer Interwell Connect. From Prod. and Inj. Rate Fluct., SPE Annual Tec. Conf. and Exhib., Dallas, TX, US.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50083-X

Production scheduling in multiproduct multistage semicontinuous processes. A constraint programming approach

Lautaro D. Marcolini[a], Franco M. Novara[a], Gabriela P. Henning[a,b,*].

[a] Facultad de Ingeniería Química, Universidad Nacional del Litoral, Santa Fe 3000, Argentina
[b] INTEC (Universidad Nacional del Litoral, CONICET), Santa Fe 3000, Argentina

[*] *Corresponding Author's E-mail: ghenning@intec.unl.edu.ar*

Abstract

An expressive constraint programming (CP) formulation has been proposed to address the scheduling problem of a make-and-pack process. The resulting CP model is able to consider the typical processing stages of food industries in an integrated fashion, while capturing many features found in the industrial practice. The proposed CP formulation has been extensively tested and compared with other existing approaches. The results have shown that the model can efficiently solve medium and large-scale problem instances with multiple constraining features. The examples that have been solved show that the proposed formulation is computationally efficient

Keywords: Production scheduling; Constraint programming; Multiproduct Multistage Semicontinuous Processes; Make-and-Pack; Food industry.

1. Introduction

The production scheduling of multiproduct multistage semicontinuous facilities is addressed in this contribution. Many food production processes have three main stages: (i) processing of raw materials into intermediate products, (ii) storage of these intermediate products, which may require an additional operation (e.g., fermentation, aging), and (iii) packing of the final products. In consequence, in such processing facilities, batch and continuous operations interplay in the manufacturing route, leading to a semicontinuous production mode. In particular, the short term scheduling problem of a real-world multistage food process previously studied by other authors (Kopanos et al., 2011; 2012) is considered in this work. The process being tackled is derived from a real-world ice-cream production facility, which was originally introduced by Bongers and Bakker (2006). In many food processing plants, scheduling just focuses on the packing units for which an efficient schedule is sought. Once such agenda is obtained, it is propagated upstream to the other processing stages. However, such approach is only appropriate when there is a unique bottleneck in the process and it is associated to the packing stage units, independently of the product mix. Unfortunately, this assumption is not always valid. Therefore, it is necessary to address the scheduling problem of all the processing stages in an integrated fashion, leading to a defying problem, whose combinatorial complexity increases significantly with the number of products and their demands (higher number of batches), as well as the consideration of changeover times.

2. Methodology

Constraint Programming (CP) techniques have been successfully applied to scheduling problems by the Process Systems Engineering (PSE) community. Most approaches have been devoted to the scheduling of multiproduct multistage batch plants (Novara et al., 2016), leaving aside semicontinuous processes. An expressive constraint programming formulation has been proposed to address the challenging problem described in the previous paragraphs. The resulting CP model is able to consider the typical processing stages of food industries in an integrated fashion, while capturing many features found in the industrial practice. The proposal is based on the ILOG-IBM OPL language and the CP Optimizer, which are embedded within the CPLEX Optimization Studio (IBM ILOG, 2013).

3. CP model

Sets/Indexes. *B/b:* batches to be produced. *Bp/-:* batches of products *p*. C_p*/c:* possible campaigns of product *p*. *P/p:* products to be manufactured. *S/s:* processing stages. *Sa/-:* subset of aging/storage stages. *Sp/-:* subset of production stages. *U/u:* equipment units. U_s*/us:* set of units belonging to stage *s, s = Card(S)*

Parameters. $changeOverTime_{u,p,p'}$*:* changeover time between products *p* and *p'* in unit *u*. $maxAgingT_p$*:* maximum aging time for product *p*. *cleaningTime:* final cleaning time. $pt_{p,u}$*:* processing time required by a batch of product *p* in unit *u*.

Variables. $campaign_{p,us,c}$*:* interval variable that spans over all the processing tasks that belong to a campaign *c* of product *p* carried out in unit *us* that belongs to the packing stage. $campaignSeq_{us}$*:* sequence variable defined for each unit *us* belonging to packing stage. It represents an ordering of campaign interval variables associated with *u*. Each interval variable in this sequence is characterized by a type that is equal to the campaign product. $stTask_{b,s}$*:* interval variable that represents the processing of batch *b* at stage *s*. $task_{b,u}$*:* interval variable representing the execution of batch *b* in unit *u*. $taskSeq_u$*:* sequence variable capturing the ordering of the $task_{b,u}$ activities that take place in unit *u.*

Constraints. Constraint (1) prescribes that each batch must be assigned to just one processing unit at each stage; i.e. just one instance of $task_{b,u}$ will be part of the schedule and its interval will be the same of $stTask_{b,s}$. Constraint (2) works in s similar way regarding the tasks of the last stage, which are executed under a campaign mode.

$$Alternative\left(stTask_{b,s}, all(u \in Us)\ task_{b,u}\right), \qquad \forall b \in B, \forall s \in S \tag{1}$$

$$Alternative\left(stTask_{b,s}, all\left(u \in Us, c \in C_p\right)\ task_{b,u,c}\right), \quad \forall b \in B, s = card(S) \tag{2}$$

Constraints (3) and (4) establish appropriate timing relationships between the first two adjacent tasks associated with any batch *b*. They synchronize the start of the first manufacturing stage, which is a continuous one, with the beginning of the second stage, which is a batch aging and storage activity.

$$startAtStart\left(task_{b,u}, task_{b,u'}\right), \quad \forall b \in B, \forall u \in U_S, \forall u' \in U_{S'}, s \in Sp, s' \in Sa, s' = s + 1 \tag{3}$$

$$startAtStart\left(stTask_{b,s}, stTask_{b,s'}\right), \forall b \in B, \forall s \in Sp, s' \in Sa, s' = s + 1 \quad (4)$$

Similarly, constraints (5) and (6) synchronize the finishing of the storage/aging activity and the packing one, prescribing that both must end at the same time, i.e. when the packaging task has already consumed all the material.

$$endAtEnd\left(task_{b,u}, task_{b,u',c}\right), \forall b \in B, \forall u \in U_S, \forall u' \in U_{S'}, \forall s \in Sa, \\ c \in C_p, s' = card(S) \quad (5)$$

$$endAtEnd\left(stTask_{b,s}, stTask_{b,s'}\right), \forall b \in B, \forall s \in Sa, s' = card(S) \quad (6)$$

Constraint (7) prescribes that the duration of each processing task depends on the unit assigned to it.

$$sizeOf\left(task_{b,u}\right) = pt_{p,u} \cdot presenceOf\left(task_{b,u}\right), \\ \forall p \in P, \forall b \in B_P, \forall u \in U_S, \forall s \in Sp \quad (7)$$

For storage/aging tasks, limits on their duration must be imposed. A batch should remain in aging vessels a minimum processing/aging time, captured by Expression (8), and no longer than its corresponding shelf-life, which is represented by constraint (9).

$$sizeOf\left(task_{b,u}\right) \geq pt_{p,u} \cdot presenceOf\left(task_{b,u}\right), \\ \forall p \in P, \forall b \in B_P, \forall u \in U_S, \forall s \in Sa \quad (8)$$

$$sizeOf\left(task_{b,u}\right) \leq maxAgingT_p \cdot presenceOf\left(task_{b,u}\right), \\ \forall p \in P, \forall b \in B_P, \forall u \in U_S, \forall s \in Sa \quad (9)$$

Constraints (10) to (13) capture the campaign operation mode of the last stage, where any packing activity $task_{b,us,c}$ must be part of a campaign. The variable $campaign_{p,us,c}$ represents a campaign *c* associated with product *p* in unit *us* belonging to the packing stage. By resorting to the *span* CP construct, expression (10) ensures that each packing task associated with a campaign of a certain product *p* takes place within the spanning interval of such campaign.

$$span\left(campaign_{p,us,c}, all(b \in B_P)\ task_{b,us,c}\right), \\ \forall p \in P, \forall c \in C_p, \forall us \in U_s, s = Card(S) \quad (10)$$

Expression (11) enforces all the campaign variables associated with a given unit not to overlap with each other.

$$noOverlap(campaignSeq_{us}), \qquad \forall us \in Us, s = Card(S) \quad (11)$$

In addition, constraint (12) prescribes that if the interval variable $task_{b,us,c}$ representing the packing task of batch *b* in unit *us*, associated with the campaign *c*, is included in the solution, the corresponding interval variable representing the packing campaign has to be included too.

$$presenceOf\left(task_{b,us,c}\right) \geq presenceOf\left(campaign_{p,us,c}\right), \\ \forall p \in P, \forall c \in C_p, \forall b \in B_P, \forall us \in Us, s\ = Card(S) \quad (12)$$

Constraint (13) avoids overlapping the execution of tasks in any unit u and simultaneously inserts the corresponding changeover time between consecutive tasks assigned to the unit.

$$noOverlap\left(taskSeq_u, changeOverTime_{p,p',u}\right), \qquad \forall u \in U, \forall p, p' \in P \tag{13}$$

Constraints (14) and (15) reduce the search space and improve the computational performance of the formulation by removing mathematical symmetries. For each unit u, if more than one batch that belongs to a given product p, is assigned to it, those batches must be processed following an increasing *id* number sequence.

$$startBeforeStart\left(task_{b,u}, task_{b',u}\right), \quad \forall u \in U_S, \forall s \in Sa, \forall b, b' \in B_P, b' = b + 1, b \neq card(b) \tag{14}$$

$$startBeforeStart\left(stTask_{b,s}, stTask_{b',s}\right), \quad \forall s \in Sa, \forall b, b' \in B_P, b' = b + 1, b \neq card(b) \tag{15}$$

Expression (16) represents the objective function to be minimized, which is makespan. The expression adds a final cleaning time which must be performed in the packing lines.

$$max\left(endOf\left(stTask_{b,s}\right)\right) + cleaningTime, \forall b \in B, \forall s \in S \tag{16}$$

4. Results

The proposed model has been tested by means of the well-known case-study originally introduced by Bongers and Bakker (2006). In addition, a comparison with the results reported by Kopanos et al. (2012) has been made.

The process corresponds to an ice-cream production facility, which manufactures eight different products, named A to H. The plant layout is depicted in Fig. 1. As seen, it has three stages: (i) processing of raw materials into intermediate products, (ii) storage and aging of these intermediate products, and (iii) packing of the final products. At the first stage only one manufacturing line is available. The second stage has six vessels and the last one has two packing lines. At stage 1 and 3 sequence dependent changeover activities must be considered. Changeovers at stage 2 are negligible; however, minimum and maximum aging times must be considered at this stage. Finally, to improve the efficiency of the last packing stage, a campaign operation mode must be enforced in it.

In order to test the CP formulation 20 different problems instances (P.01 to P.20) of this case study have been solved, varying the number of batches needed to fulfil increasing product demands. In fact, from problems P.01 to P.20, the number of batches raises from 70 to 180. The examples were solved on a computer having 16 GB of RAM memory and AMD Ryzen 3 3200G processor. Optimal solutions were reached in 13 out the 20 instances with low computational effort (29 to 734 seconds of CPU time) and only 7 good quality suboptimal solutions were reached. A limit of 3600 s of CPU time was imposed.

Table 1 presents the values of the objective function that were obtained by means of this proposal and the ones that were reported by Kopanos et al. (2012). It can be seen that in nine instances better values have been reached and in two cases the same values have been obtained. The worst quality solution that was reached has a makespan value that is only 0.23% greater that the corresponding optimal solution.

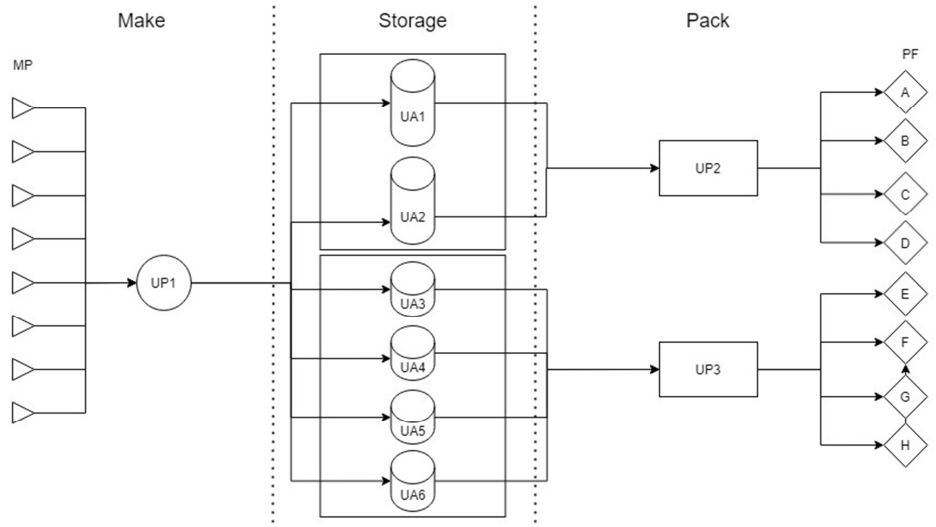


Fig. 1. Manufacturing process that corresponds to the case study under consideration.

Table 1. Objective function values obtained by means of this proposal and the ones of Kopanos et al. (2012).

	Solution Approach		
Instance	**Proposed CP Model**	**MIP-R***	**MIP BasB***
P.01	**120.28**	120.33	120.33
P.02	119.48	**118.17**	**118.17**
P.03	131.62	**131.48**	**131.48**
P.04	**142.07**	142.10	142.10
P.05	**149.65**	149.66	149.66
P.06	152.88	**152.34**	**152.34**
P.07	162.50	**161.47**	**161.47**
P.08	**171.35**	171.37	171.37
P.09	176.23	**175.82**	**175.82**
P.10	**187.75**	**187.75**	**187.75**
P.11	**191.18**	191.25	191.25
P.12	**206.42**	**206.42**	**206.42**
P.13	202.67	**201.76**	**201.76**
P.14	**223.55**	223.56	223.56
P.15	**224.68**	224.71	224.71
P.16	222.58	**222.06**	**222.06**
P.17	238.48	**238.04**	**238.04**
P.18	251.98	**251.49**	**251.49**
P.19	**260.45**	260.52	260.52
P.20	**291.72**	291.75	291.75

Bold numbers represent the best value of the objective function corresponding to each instance.

Fig. 2 depicts the Gantt diagram corresponding to the largest problem instance, having 180 batches.

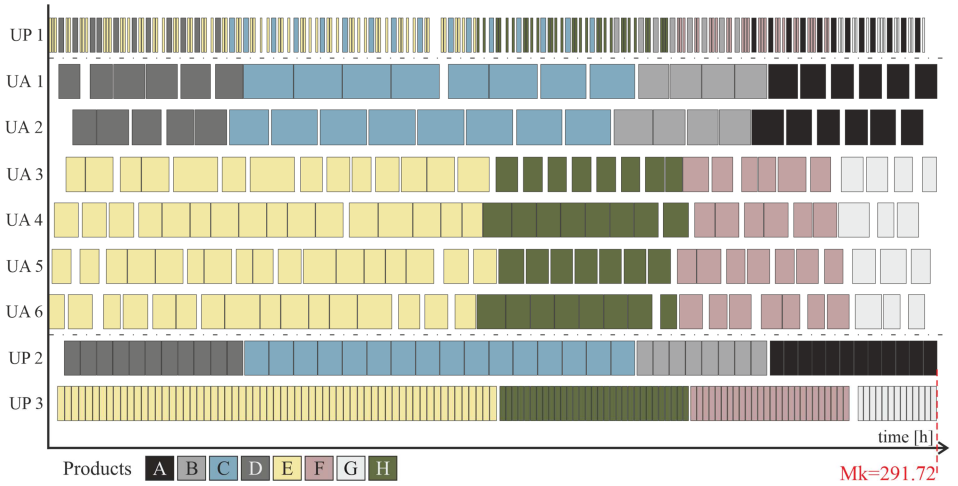


Fig. 2. Gantt diagram corresponding to problem instance P.20.

5. Conclusions and future work

Until now, most constraint programming models addressing industrial scheduling problems have focused on batch plants. In this contribution, an efficient CP model was developed to address the scheduling problem of make-and-pack plants, in which batch groupings into campaigns are considered in the last packing stage. It was applied to twenty instances of a case study, corresponding to an ice cream producer company, ranging from 70 to 180 batches. Despite the significant increase in the number of batches (157%), the performance of the model did not degrade because the rise of the number of variables was limited (138%). A comparison with the results reported by Kopanos et al. (2012) has been made, which allows concluding that the model is competitive. Future work will include the extension of the proposal to consider more complex industrial processes. In addition, the stochastic nature of certain variables, such as the processing rates/times, will be taken into account.

References

Bongers P.M.M., Bakker B.H., Application of multi-stage scheduling. In: Marquardt W., Pantelides C.C. (Eds.), 16th European Symposium on Computer Aided Process Engineering and 9th International Symposium on Process Systems Engineering, Elsevier. 1917-1922 (2006).

IBM ILOG, IBM ILOG CPLEX Optimization Studio. http://www-01.ibm.com/software/integration/optimization/cplex-optimization-studio/ (2013).

Kopanos G.M., Puigjaner L., Georgiadis M.C., Production scheduling in multiproduct multistage semicontinuous food processes. *Industrial Engineering & Chemistry Research*, 50, 6316-6324, (2011).

Kopanos G.M., Puigjaner L., Georgiadis M.C., Efficient mathematical frameworks for detailed production scheduling in food processing industries, *Computers and Chemical Enginnering,* 42, 206–216 (2012).

Novara, F.M., Novas, J.M., Henning, G.P., A novel constraint programming model for large-scale scheduling problems in multiproduct multistage batch plants: Limited resources and campaign-based operation, *Computers and Chemical Enginnering,* 93, 101–117 (2016).

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50084-1

Maintenance scheduling optimization for decaying performance nonlinear dynamic processes

Bogdan Dorneanu[a], Vassilios S. Vassiliadis[b], Harvey Arellano-Garcia[a,*]

[a]*LS Prozess- und Anlagentechnik, Brandenburgische Teschnische Universität Cottbus-Senftenberg, Cottbus, D-03044, Germany*
[b]*Cambridge Simulation Solutions Ltd., Larnaca 7550, Cyprus*
arellano@b-tu.de

Abstract

A first contribution of this paper is an overview of the research efforts and contributions over several decades in the area of scheduling maintenance optimization for decaying performance dynamic processes. Following breakthrough ideas and implementation in the area of heat exchanger networks for optimal scheduling of cleaning actions subject to exchanger surface fouling, these concepts were transferred successfully to the area of scheduling catalyst replacement actions in catalytic reactor networks. This necessary overview leads to the main, second contribution aimed with this work: its application to restorative maintenance scheduling in the area of RON regeneration actions planning, as well as point to new areas where this approach can be fruitfully applied to and extended into in the near future – particularly enhancing model descriptions that include general types of planning uncertainty. The effectiveness and efficacy of the approach is demonstrated computationally in this work.

Keywords: maintenance scheduling optimization, decaying performance processes, multistage optimal control, bang-bang optimal control, reverse osmosis networks.

1. Introduction

Modern engineering systems and manufacturing processes are nowadays very complex, with the demand for integration and multitasking processing being an ever-increasing trend so as to facilitate flexible manufacturing over multiple products, increase efficiency, reduce costs and environmental impact, as well as to secure safe operation. Production facilities thus involve numerous interactions and dependencies between components, and operate in highly dynamic environments. The operation of processes with decaying performance over time gives rise to challenging modelling and optimization problems. As the performance degrades over time, process shutdown for unit cleaning (reverse osmosis networks (Saif et al. 2019), heat exchanger networks (Al Ismaili et al. 2019)) or catalyst changeovers (catalytic processes) (Adloor & Vassiliadis, 2021) must be planned to restore it.

In order to avoid this, parallel processing lines are used to manufacture the products. This set up can improve the flexibility of the production process by allowing the shut-down of one unit for cleaning purposes, while the remaining units continue to meet the products' demand. While this maintenance action does improve the product yield, there are negative impacts associated with this operation, such as loss of production time, or energy and labour costs to restore the performance (Adloor & Vassiliadis, 2020). This leads to a trade-off to be addressed for each unit: while frequently cleaning results in high

production rates, large maintenance costs and loss in production occur. This trade-off can be optimally managed by developing maintenance schedules that specify the optimum units to be used and the optimal use time of each unit in the parallel set up, over a fixed time horizon. The schedule may also be required to fulfil a constraint that no two units undergo cleaning action at the same time due to production requirements or labour and equipment availability (Al Ismaili et al. 2018).

Additionally, it is also necessary to identify the optimal operating conditions, as well as to ensure that the maintenance schedule and the process operation are tailored to produce an adequate inventory of product to effectively meet varying demand across the time horizon, while also avoiding excessively high storage costs. An integrated execution of all these decisions in an optimal manner can greatly minimise the negative effects of the performance decaying process, and thereby maximise the profit (Adloor & Vassiliadis, 2020).

The following sections present an overview of research efforts and contributions over several decades in the area of scheduling optimization for decaying performance dynamic processes, with particular focus on RON regeneration, as well as new areas of application and extension.

2. Maintenance scheduling of decaying performance processes

Two approaches are commonly employed in dealing with the maintenance scheduling (Santamaria & Macchietto, 2018):

a) *Optimal scheduling problem*, with binary decision variables associated with the operating states of the units (cleaning/operating) and the timing and sequencing of the task. The problem is combinatorial in nature and it is typically addressed using (pseudo-)steady-state models.

b) *Dynamic optimal scheduling problem* involving differential-algebraic equations (DAEs). In this case, the result is a (mixed-integer) nonlinear programming problem, but offers the flexibility of accommodating various types of models (Assis et al., 2015).

Furthermore, the accuracy of the process models used is of paramount importance. Rigorous models, capturing the full representation of the physical phenomena can be computationally expensive for a large-scale scheduling problem. Yet, inadequately describing the physics of the process may affect the validity of the obtained maintenance schedules, and the result may end up being useless for practical application (Van Horenbeek et al. 2010).

In the following, the underlying scheduling problem is reformulated as a dynamic multistage optimization (optimal control) model, and cast in a form that promotes bang-bang type solutions for the control variables associated with restorative action periods. This bang-bang behaviour is entirely equivalent to having a Boolean variable (integer, binary) within an otherwise smoothly represented dynamic optimisation model.

This approach has been successfully applied for solving maintenance scheduling problems for HENs (Al Ismaili et al., 2018) and catalytic reactor networks (Adloor & Vassiliadis, 2020). Furthermore, it has enabled reliable inclusion of process uncertainty to be included realistically in the resulting models (Al Ismaili et al., 2019; Adloor &

Vassiliadis, 2021). In the following, the application of the approach on the maintenance scheduling of a RON will be demonstrated as an original contribution of this work.

3. Maintenance scheduling of reverse osmosis networks (RONs)

Reverse osmosis (RO) is a well-established technology for water desalination. A commercial RO desalination system consists of seawater intake, seawater pre-treatment, main RO separation and post-treatment sections, which include several RO passes with auxiliary equipment, e.g., high-pressure pumps, energy recovery, etc. (Saif et al. 2019).

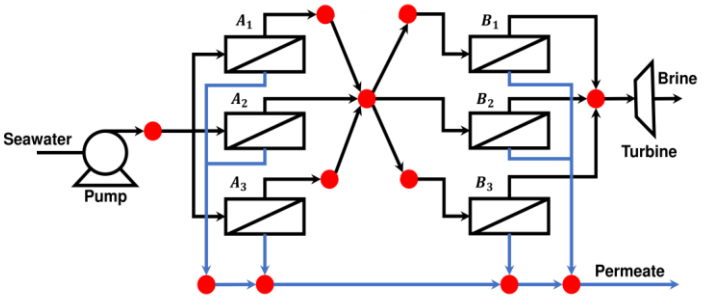


Figure 1: Reverse osmosis network considered

For the case study, a RON with 2 stages of 3 individual modules, with a total of 6 RO units, illustrated in Figure 2, is considered. Each RO unit has a membrane area, A of 152 m^2 and a permeate recovery ratio of 0.65. The RON processes a total flowrate of 5,004 m^3/h of sea water, with a salt content of 34,800 ppm. Other parameters for the RON operation (*e.g.,* membrane permeability decay constant – γ, solute transport parameters, initial water permeability, etc.) are taken from See et al. 2004.

3.1. Mathematical model

The decay in the unit's performance is defined as:

$$-\frac{dK_i}{dt} = Y_i \cdot \frac{K_i}{\gamma} \qquad i = 1, \dots, 6 \qquad (1)$$

Where K = the membrane permeability [kg m^{-2}], Y = a binary variable, equal to 1 if the RO unit is in operation or 0 if the unit is in cleaning action, and i = the RO module number.

The flowrates of the RO module inlet streams, F are determined from the total flowrate, F_{total}, as follows:

$$F_i = \frac{1}{3} \cdot Y_i \cdot F_{total} \qquad i = 1, \dots, 3 \qquad (2)$$

$$F_j = \frac{1}{3} \cdot Y_j \cdot R_k \qquad j = 4, \dots, 6 \qquad k = 1, \dots, 3 \qquad (3)$$

Where R = reject flowrate [m^3 day^{-1}].

The permeate flowrate, P_i is determined based on the permeate recovery ratio from:

$$P_i = Y_i \cdot \alpha \cdot F_i \quad (4)$$

Where α = the permeate recovery ratio.

The concentration of the permeate, $C_{P,i}$ is calculated based on the solute transport parameter, D, the pressure drop, ΔP and the osmotic pressure, $\Delta\pi$ as:

$$C_{P,i} = \frac{D \cdot C_{F,i}}{\Gamma \cdot K_i \cdot (\Delta P_i - \Delta\pi_i)} \quad (5)$$

Where Γ = the membrane geometry correction factor, and C_F = the concentration of the RO module inlet stream.

The pressure drop over a RO module is determined as:

$$\Delta P_i = Y_i \cdot \left(\frac{P_i}{A \cdot \Gamma \cdot K_i} + \Delta\pi_i\right) \qquad i = 1, \dots, 6 \quad (6)$$

The scheduling of the maintenance actions is defined as an optimisation problem having as decision variables the binary variables Y_i. The objective function to be optimised is the total cost of operating the RON, calculated as:

$$J = Income_{Sale} - Cost_{Cleaning} - Cost_{Energy} \quad (7)$$

With the income from permeate sales determined as:

$$Income_{Sale} = cost_{permeate} \cdot \sum_i P_i \qquad i = 1, \dots, 6 \quad (8)$$

The cost of cleaning as:

$$Cost_{Cleaning} = cost_{cleaning\ action} \cdot N_{CleaningPeriods} \quad (9)$$

And the energy cost:

$$Cost_{Energy} = cost_{electricity} \cdot \frac{\sum_i F_i \cdot \Delta P_i}{\eta_{pump}} \quad (10)$$

Where η_{pump} = pump energy efficiency.

The cost of electricity is assumed equal to 0.30 € kWh^{-1}, the cost of the cleaning action is 100 € $unit^{-1}$ cleaning $action^{-1}$, while the permeate is sold at 0.48 € m^{-3}. The dynamic model of the RON presented above is implemented as a multiperiod simulation model, with a planning horizon equal to 26 weeks. This long planning horizon has been selected to even out economic effects. The optimization is implemented using a heuristic penalty scheme, to enforce binary values or the controls. A multiple start policy with 50 cycles is considered in order to determine the spread of the local solutions.

3.2. Results and discussion

The optimisation problem is implemented in Python v3.8 and solved on an Intel Core i7-8550U CPU @1.80GHz, 16.0 GB RAM. The *minimize* solver from *scipy.optimize* is used. Both linear and nonlinear constraints are implemented. The CPU time is 163.97 minutes. Out of the 50 multiple start cycles considered, 13 were successful (an optimal solution is found), with the value of the objective between €7.275 and €7.565 million. From the results in Figure 2 it can be observed that each of these solutions is being obtained more

than once, hence the advantage of using the multiple start policy in identifying a better solution.

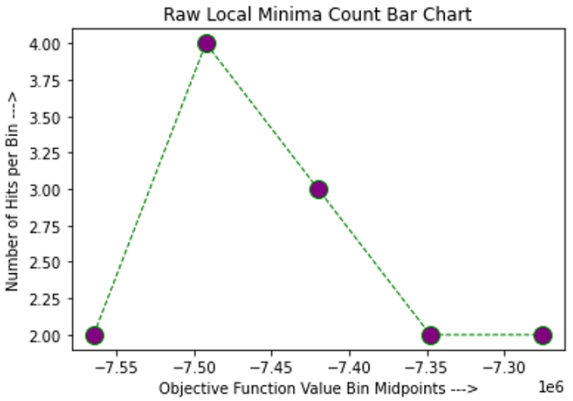


Figure 2: Local minima for the successful optimization cycles

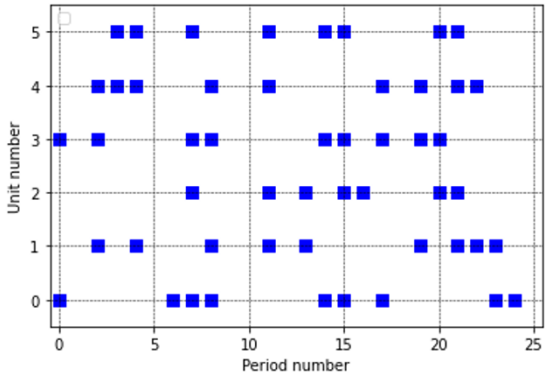


Figure 3: Cleaning schedule for the RON

The best point, with a value of the objective function of €7.565 million. The resulting maintenance schedule is presented in Figure 3. For most of the heat exchangers, there is no cleaning action near the start and at the end of the operating horizon, a similar behaviour observed by Al Ismaili et al. (2018) for HENs. This is because there is very little incentive to increase the cleaning cost further.

The number of cleaning actions varies between 7 (for RO module 3) and 9 (RO modules 1,2, 4, and 5). This could indicate that these modules are more important in the network.

4. Future directions for maintenance scheduling optimization

The maintenance scheduling of RONs can be further investigated to account for uncertainty in the sensor data regarding equipment degradation. Furthermore, comparisons with similar studies using mixed-integer formulations should be considered. The optimal control approach for scheduling maintenance and production can be further applied to other similar processes with decaying performance, such as biopharmaceutical manufacturing under performance decay (Liu et al. 2014) or chromatography-based processes (Vieira et al. 2017). Parallel production lines able to process multiple feeds and

produce multiple products should be investigated as well, to overcome the combinatorial nature of the mixed-integer methodology (Adloor & Vassiliadis, 2020).

5. Conclusions

This work provides an overview of the optimal control approach for scheduling maintenance of decaying performance processes. An original case study of RONs is presented to illustrate the application of this procedure, based on the formulation of the scheduling task as a multistage mixed-integer optimal control problem, considering a dynamic model of the process. The requirement for an integer variable handling solution algorithm (*e.g.,* Branch-and-Bound) is not needed, although it is not always possible to ensure the necessary condition that each resulting case will exhibit such bang-bang behaviour.

Nonetheless, for significant portions of the observed solution profiles, over many applications over the course of years, it has been observed that mild penalization heuristic schemes work sufficiently well and very reliably over a multitude of published case study solutions.

References

S. Adloor & V.S. Vassiliadis, 2020, An optimal control approach to scheduling maintenance and production in parallel lines of reactors using decaying catalysts, Computers and Chemical Engineering 142, 107025

S. Adloor & V.S. Vassiliadis, 2021, An optimal control approach to considering uncertainties in kinetic parameters in the maintenance scheduling and production of a process using decaying catalysts, Computers and Chemical Engineering 149, 107277

R. Al Ismaili et al., 2018, Heat exchanger network cleaning scheduling: From optimal control to mixed-integer decision making, Computer and Chemical Engineering 111, pp. 1-15

R. Al Ismaili et al., 2019, Optimisation of heat exchanger network cleaning schedules: Incorporating uncertainty in fouling and cleaning model parameters, Computers and Chemical Engineering 121, pp. 409-421

B.C.G. Assis et al., 2015, Dynamic optmization of the flow rate distribution in heat exchanger networks for fouling mitigation, I&EC Research 54, pp. 6497-6507

S. Liu et al., 2014, Optimal production and maintenance planning of biopharmaceutical manufacturing under performance decay, I&EC Research 53, pp. 17075-17091

Y. Saif et al., 2019, MINLP model for reverse osmosis network design under time-variant operation constraints, I&EC Research 58, pp. 22315-22323

F. L. Santamaria & S. Macchietto, 2018, Integration of optimal cleaning scheduling and control of heat exchanger network undergoing fouling: Model and formulation, I&EC Research 57, pp. 12842-12860

H.J. See et al., 2004, Design of reverse osmosis (RO) water treatment networks subject to fouling, Water Science and Technology 49 (2), pp. 263-270

A. Van Horenbeek, et al., 2010, Maintenance optimization models and criteria, Int. J. Syst. Assur. Eng. Manag. 1(3), pp. 189-200

M. Vieira et al., 2017, Production and maintenance planning optimization in biopharmaceutical processes under performance decay using a continuous-time formulation: A multi-objective approach, Computers and Chemical Engineering 107, pp. 111-139

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50085-3

Cleaning schedule for heat exchanger networks subjected to maintenance constraints

Parag Patil[a], Babji Srinivasan[b,d], Rajagopalan Srinivasan[c,d*]

[a]*Department of Chemical Engineering, Indian Institute of Technology Gandhinagar, Gandhinagar382355, India*
[b]*Department of Applied Mechanics, Indian Institute of Technology Madras, Chennai, 600036, India*
[c]*Department of Chemical Engineering, Indian Institute of Technology Madras, Chennai, 600036, India*
[d]*American Express Lab for Data Analytics, Risk and Technology, Indian Institute of Technology Madras,Chennai Tamil Nadu, 600036, India*
**raj@iitm.ac.in*

Abstract

Fouling degrades the overall efficiency of the heat exchanger networks (HENs), which results in a significant economic loss. The mitigation of fouling in an operational HEN is carried out by optimizing the cleaning schedules of the heat exchangers. Although such approach can save costs, it is subjected to the exact implementation of the optimal cleaning schedule. Usually, the small and medium-scale process industries face difficulties in implementing such solutions due to limited resources, which forces them to rely on suboptimal cleaning schedules, such as postponing or avoiding few cleaning tasks. This work addresses this gap by optimizing the cleaning schedule considering the maintenance resource limitation. Our approach considers a mixed-integer linear programming (MILP) based optimization considering groupings of heat exchangers based on their spatial locations for ease of maintenance The proposed formulation is applied on a HEN with linear and asymptotic fouling, with and without cleaning cost. The results show that the approach can prevent a considerable economic loss, which would incur due to suboptimal cleaning schedules due to resource limitations.

Keywords: Heat exchanger network, fouling, cleaning schedule, MILP, maintenance constraints

1. Introduction

The HENs are present in almost all process industries, such as oil refineries, pulp, paper mills, sugar factories, etc. (Trafczynski et al., 2021). Such networks are mainly employed to recover the waste energy from hot process outlet streams to the cold process inlet streams in the process plant. As the operation progresses, the foulant in the streams starts depositing in the heat exchangers. The thermal conductivities of the foulant are usually lower than the material of construction of the heat exchangers; thus, their deposition results in reduced heat transfer rates. Moreover, the cross-sectional areas in the heat exchangers also reduce due to foulant deposition, increasing the pressure drops. Overall, the fouling affects thermal and hydraulic efficiencies of the HENs. Extra pumping power and utility consumption are required to compensate for this efficiency reduction, resulting in substantial economic loss. In 2015, the total cost of fouling in

preheat trains in US refineries was reported to be about $2.26 billion (Coletti et al., 2015). Thus, fouling mitigation is essential to prevent such huge economic losses.

Generally, fouling in HENs is mitigated at two levels: (1) at the design or retrofit stage, where the heat exchanger geometries and HEN structures are optimized to minimize foulant deposition; (2) At the operational stage, where mass flowrates and heat exchanger cleaning schedules are optimized to reduce the additional utility and cleaning costs. The current work is focused on the existing networks; hence falls under the latter type. Several studies in this category formulate it as an optimization problem, considering the operating variables and cleaning activities as continuous and binary variables, respectively. The independent and simultaneous consideration of the variables in the optimization framework results in nonlinear programming (NLP) and mixed-integer nonlinear/linear programming problems(MINLP/MILP). Several authors have formulated the optimization of cleaning schedules as MINLP and MILP problems (Georgiadis et al., 2000, Smaili et al., 2002). Optimization based on simultaneous consideration of cleaning schedules and flow distribution has also been studied (Santamaria and Macchietto, 2020). Although these studies have shown the potential to save costs due to fouling, the saving is only possible with effective implementation of the obtained solutions. However, the small and medium-scale process industries usually face difficulties in implementing optimal solutions due to limited resources such as low maintenance budgets (Wang, 2016). Therefore, they tend to follow a suboptimal cleaning schedule by skipping or postponing the cleaning of heat exchangers, which results in higher utility consumption. Thus, obtaining the optimal cleaning schedules considering maintenance resource constraints is necessary. To our knowledge, no fouling mitigation studies have evaluated these limitations.

This work addresses the described gap by modelling the cleaning schedule of HENs with maintenance resource limitations. The heat exchangers in HENs can be either grouped based on similarities of the type of cleaning required, such as mechanical and chemical methods, or based on their spatial locations for ease of maintenance. This work proposes an MILP formulation for cleaning of HENs considering heat exchanger groups. The grouping is modelled using linear constraints. Next, we describe the MILP formulation used.

2. MILP problem formulation description

The HEN is modelled by the digraph method, where the edges are streams (s), and vertices are process equipment (E). The set of streams(s) is divided in the subset of cold streams (c_s) and hot streams (h_s) . The set of equipment is divided into subsets of supply units (Su_E), demand units (Du_E), heat exchangers (Hx_E), mixers (Mx_E), splitters (Sp_E) and an additional heater (He_E) . The overall operating time is discretized into sub-periods, denoted by τ. Following is a brief discussion about the constraints and objective functions:

a) *Mass balance constraints*: The mass flow rates in each stream (m_s) are assumed to be constant during the operational. The mass balance in the network is simulated based on the Equations (18) to (23) from Assis et al. (2013).
b) *Energy balance constraints*:
    - The energy balance across all the equipment is simulated as per Equations (26) to (31) from Assis et al. (2013).
    - The heat transfer across the heat exchangers are modelled by lumped-parametric model of $P - NTU$ method as follows:

$$P_{hx,\tau}\left(1-y_{hx,\tau}\right)T_{hx,\tau}^{c,i}+\left(1-P_{hx,\tau}\right)\left(1-y_{hx,\tau}\right)T_{hx,\tau}^{h,i}-T_{hx,\tau}^{h,o}=0 \quad (1)$$

$$T_{hx,\tau}^{c,i}-T_{hx,\tau}^{c,o}-CR_{hx,\tau}\left(T_{hx,\tau}^{h,i}-T_{hx,\tau}^{h,o}\right)=0 \quad (2)$$

where, $P_{hx,\tau}$ is the heat exchanger effectiveness parameter in the interval (τ), based on the number of transfer units $\left(NTU_{hxv,\tau}\right)$. $CR_{hx,\tau}$ is ratio of heat capacity of the flowrates. $y_{hx,\tau}$ is binary variable, representing state of the heat exchanger: $y_{hx,\tau}$=1denotes cleaning and $y_{hx,\tau}$=0 denotes under operation. $T_{hx,\tau}^{c,i}$, $T_{hx,\tau}^{c,o}$ and $T_{hx,\tau}^{h,i}$, $T_{hx,\tau}^{h,o}$ are inlet and outlet temperatures of cold and hot streams in the heat exchangers (hx) in τ interval, respectively. It can be observed that equation (1) contains the terms $y_{hx,\tau}T_{hx,\tau}^{c,i}$ and $y_{hx,\tau}T_{hx,\tau}^{h,i}$, which are bilinear, which are linearized using a set of linear inequalities (Floudas, 1995). Also, $P_{hx,\tau}$ is a nonlinear function of thermal resistance due to fouling $\left(R_{f_{hx,\tau}}\right)$. Thus, $P_{hx,\tau}$ introduces fouling in the energy balance through Equation (1). It is to be noted that, both the cold and hot streams across the heat exchangers are bypassed during their cleaning.

c) *Incorporation of fouling*: We consider linear and asymptotic nature of fouling based on the following respective equations:

$$R_{f_{hx,\tau}}=K_{hx}*\tau \quad (3)$$

$$R_{f_{hx,\tau}}=R_{fhx}^{\infty}(1-\exp-\tau/\Gamma) \quad (4)$$

where, K_{hx}, R_{fhx}^{∞} and Γ are the fouling parameters with appropriate units.

d) *Objective function*: It is sum of the additional utility cost and cleaning cost of each heat exchanger in network, as follows:

$$fobj=\sum_{\tau=1}^{\tau_f}\frac{m_s Cp_s}{\zeta}\left(T_{targ}-T_{\tau,s'}\right)+\sum_{hx=1}^{n}\sum_{\tau=1}^{\tau_f}C_{hx}y_{hx,\tau} \quad (5)$$

where T_{targ} is the target temperature, s' is the target stream and C_{hx} is the cleaning cost of each heat exchanger. ζ is the efficiency of the heater. $fobj$ is the overall cost, which has to be minimized by the optimization formulation by providing an optimal cleaning schedule using the proposed MILP formulation. Following section demonstrates a case study based on proposed MILP approach.

3. Case study

In this section, we apply MILP optimization formulation explained in section (2) on a HEN, shown in Fig.1 (Assis et al. 2013). It consists of four supply units, four demand units, six heat exchangers, four flow splitters, four mixers and a fired heater. All the heat exchangers have heat transfer area of 400 m^2 and the overall heat transfer coefficient of

253 W/m^2 in the clean condition. The overall heat transfer rates in each heat exchangers

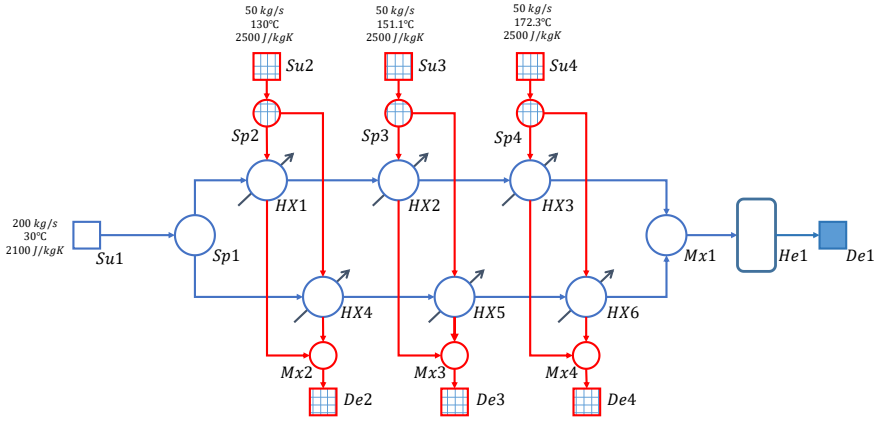


Figure 1: Schematic of the HEN used in the case study (adopted from Assis et al., 2013)

is 4.44 MW. The cold and hot fluids are flowing through the tubes and shell sides, respectively. The split fractions in each of the splitter is 0.5. The nominal temperature of the stream entering fired heater is 97.06℃. Heater provides additional energy in case if this temperature falls below its nominal value due to fouling or if any heat exchanger is taken offline for cleaning. The cost of utility, supplied to the heater is assumed to be 0.01 £/kWh with 75% efficiency.

Two cases of linear and asymptotic fouling depositions based on Equations (3) and (4) are considered here. Further, two scenarios in each of the fouling cases are considered. In one, the cleaning cost of the heat exchanger is neglected, while in another, it is considered £ 4,000 per cleaning. Thus, four scenarios are generated as follows:

a) *Scenario 1*: Linear fouling with cleaning cost
b) *Scenario 2*: Linear fouling without cleaning costs
c) *Scenario 3*: Asymptotic fouling with cleaning cost
d) *Scenario 4*: Asymptotic fouling without cleaning cost

An operational period of 12 months is considered in each scenario. The optimization formulation described in the section (2) is applied in all the scenarios. The obtained schedules for three scenarios (except scenario 3) are shown in Fig.2. Scenario 3 does not have the cleaning schedule for any heat exchanger. The overall costs incurred in each scenario are tabulated in Table 1. It can be observed that the number of cleanings in the scenarios considering cleaning costs is less than the scenarios without cleaning costs. Also, the costs in the scenarios with linear fouling are lower than those with asymptotic fouling. It is to be noted that, the optimization in scenario 3 denotes no cleaning of any heat exchanger.

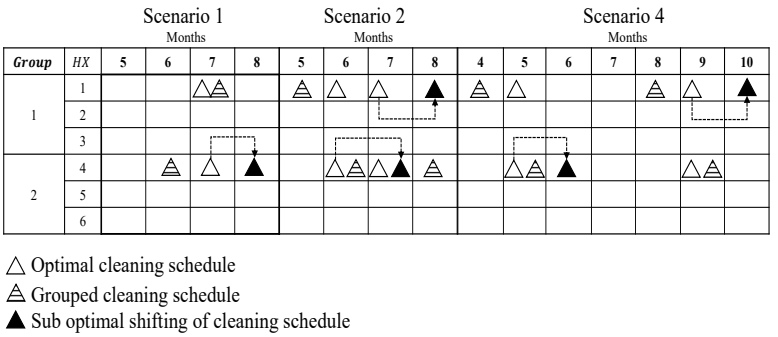


Figure 2: Cleaning schedule for scenarios 1, 2 and 4 without grouping, with grouping and with resource constraints.

Table 1: Costs in various scenarios in optimal and suboptimal conditions

Scenarios	Cost with optimal schedule without group (£)	Cost with suboptimal schedule (£)	Cost with optimal schedule with group (£)	Saving, %
1	90,643	91,110	90,716	84.4
2	80,500	82,145	81,433	43.3
3	1,49,560	1,49,560	1,49,560	-
4	1,44,400	1,44,450	1,44,850	88.9

As discussed in section (2), the maintenance teams in the process industries may not consistently implement the obtained optimal cleaning schedules due to resource limitations. They may end up skipping or postponing the cleaning of a few heat exchangers, resulting in suboptimal cleaning schedules resulting in higher costs. One of the approaches to tackle this problem is to group the heat exchangers based on the similarity of type of foulant being deposited in the heat exchangers or the locations of the heat exchangers. By grouping the heat exchangers, the maintenance team can easily perform the cleaning activities, following the optimal schedule obtained. For the considered network, we assume that the heat exchangers $HX1$ to $HX3$ and $HX4$ to $HX6$ are spatially located near each other. Thus, two groups based on their locations are defined as follows:

a) Group 1: $HX1 - HX3$
b) Group 2: $HX4 - HX6$.

The maintenance team can perform cleaning of multiple heat exchangers if they belong to a same group. Whereas, the cleaning activities would get postponed or skipped if the multiple heat exchangers from different groups are in the cleaning schedule. It can be seen from Fig. 2 that $HX1$ and $HX4$ have to be cleaned simultaneously in 7^{th} month for scenario 1. However, they belong to different groups. If cleaning of these heat exchangers is postponed by a month, it may result in higher cost of £ 91,110 due to sub-optimality (The cost for optimal cleaning schedule is £ 90,643). Whereas, if optimization is carried out with defined groups, the obtained schedule allows the maintenance team to clean the heat exchangers based on their limitations. The cost in this case is £ 90,716, which reduces the economic loss by 84.4 %, comparing the suboptimal schedule. All the other scenarios show a similar observation.

4. Conclusion and future work

Fouling in operational HENs is tackled by optimizing the mass flow distribution and cleaning schedule of the heat exchangers using MINLP or MILP frameworks. However, implementations of obtained optimal solutions is dependent on the limitations of the maintenance teams, particularly in the small and medium scale industries. This paper takes into account the maintenance resource limitations while considering the HEN cleaning optimization problem.

We formulate the problem using the MILP approach. Further, we obtained the optimal cleaning schedule in a HEN with six heat exchangers under various scenarios of linear and asymptotic fouling, with and without cleaning costs. We assume additional constraints of grouping the heat exchangers as per their spatial locations. Our results reveal that the proposed approach can prevent a considerable economic loss due to possible sub-optimal cleaning schedule, which are commonly followed by the maintenance teams due to resource limitations.

In future, we are planning to extend the proposed methodology to larger networks, where multiple grouping is possible. Apart from locations, the grouping will be also carried out based on fouling type, as different foulants require different cleaning methods.

Acknowledgement

This work is partially funded by American Express Lab for Data Analytics, Risk and Technology, Indian Institute of Technology Madras, Chennai, India.

References

Assis, B. C., Lemos, J. C., Queiroz, E. M., Pessoa, F. L., Liporace, F. S., Oliveira, S. G., & Costa, A. L. (2013). Optimal allocation of cleanings in heat exchanger networks. *Applied thermal engineering*, *58*(1-2), 605-614.

Coletti, F., Joshi, H. M., Macchietto, S., & Hewitt, G. F. (2015). Introduction. In Crude Oil Fouling: Deposit Characterization, Measurements, and Modeling (pp. 1–22). Elsevier Inc.

Floudas, C. A. (1995). *Nonlinear and mixed-integer optimization: fundamentals and applications*. Oxford University Press.

Georgiadis, Michael C., Lazaros G. Papageorgiou, and Sandro Macchietto. "Optimal cleaning policies in heat exchanger networks under rapid fouling." Industrial & engineering chemistry research 39.2 (2000): 441-454.

Ishiyama, E. M., Heins, A. V., Paterson, W. R., Spinelli, L., & Wilson, D. I. (2010). Scheduling cleaning in a crude oil preheat train subject to fouling: Incorporating desalter control. Applied Thermal Engineering, 30(13), 1852–1862.

Santamaria, F. L., & Macchietto, S. (2020). Online Integration of Optimal Cleaning Scheduling and Control of Heat Exchanger Networks under Fouling. Industrial and Engineering Chemistry Research, 59(6), 2471–2490.

Trafczynski, M., Markowski, M., Urbaniec, K., Trzcinski, P., Alabrudzinski, S., & Suchecki, W. (2021). Estimation of thermal effects of fouling growth for application in the scheduling of heat exchangers cleaning. *Applied Thermal Engineering*, *182*, 116103.

Wang, Y. (2016). What are the biggest obstacles to growth of SMEs in developing countries? – An empirical evidence from an enterprise survey. *Borsa Istanbul Review*, *16*(3), 167–176.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50086-5

Estimating Energy Market Schedules using Historical Price Data

Nicole Cortes[a], Xian Gao[a], Bernard Knueven[b], and Alexander W. Dowling[a*]

[a]*Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN 46556, USA*
[b]*Computational Science Center, National Renewable Energy Laboratory, Golden, CO 80401, USA*
**adowling@nd.edu*

Abstract

The global climate crisis is expected to reshape the energy generation landscape in the coming decades. Increasing integration of non-dispatchable renewable energy resources into energy infrastructures and markets creates uncertainty as well as new opportunities for flexible energy systems. To conduct proper economic evaluation of flexible energy systems, such as integrated energy systems (IES), advancements in modelling of market interactions, such as bidding, is crucial. This work presents a shortcut algorithm which uses two mixed integer linear programs to compute dispatch schedules (e.g., hourly power production targets) that are constrained by the resource's bid information and characteristics (e.g., minimum up and down times) based on historical locational marginal price (LMP) data. The proposed algorithm is approximately 100 times faster and uses orders of magnitude less data than a full production cost model (PCM). We find the shortcut simulator recapitulates generator dispatch signals for the Prescient PCM with approximately 4% error for the RTS-GMLC test system.

Keywords: Electricity Generation, Energy Markets, Integrated Energy Systems, Multiscale Simulation

1. Introduction

Governments around the world have pledged to lower their carbon emissions in response to climate change. Incorporating more variable renewable energy (VRE) sources, such as wind and solar, into power systems is critical to meet these goals. While VRE resources have many benefits such as low to zero emissions and operating costs, their unpredictable nature is challenging for electric grid operations. They increase price variability (Seel et al. 2018) and create strong incentives for more flexible energy generation and consumption. Using historical market price data, Dowling et al. 2017a showed that energy systems can more than double their profits by participating in faster market timescales. Many new promising technologies, including integrated energy systems (IES) which exploit the synergy between multiple technologies (e.g., renewables, nuclear, fossil-based with CO_2 capture, energy storage) by tightly coupling them into single systems (Arent et al. 2021) can provide flexibility to enhance grid reliability and resilience with high VRE utilization. But properly valuing the flexibility of these new technology concepts requires analysis that directly considers interactions between IESs

and energy markets. Traditional energy system value metrics, such as levelized cost of electricity (LCOE), do not capture the value created in the market (Dowling et al. 2017b).

Wholesale energy markets coordinate the generation and consumption of electricity from an increasingly diverse set of technologies. The markets set energy prices in a two-settlement system: a day-ahead market (DAM) to meet forecasted demand and a balancing real-time market (RTM) for fast adjustments. Market participants, providing energy generation or ancillary services (various reserves or frequency regulation), can interact with the market via self-scheduling or bidding. A resource that self-schedules creates its own power generation schedule over its preferred planning horizon and is subject to the cleared market price. In contrast, bidding requires participants to submit a set of power-price pairs to the independent service operator (ISO). The power-price pairs reflect the resource's marginal costs and generation flexibility to the ISO. With all the submitted bids, generation is scheduled by optimizing the bids and clearing the market in order of cost. Once enough generation has been scheduled to meet forecasted energy demand for the considered horizon (following day for DAM or following hours for RTM), the locational marginal price (LMP), or price per MWh produced, is set by the highest cost resource to clear the market. Ela et al. 2014 found self-scheduling, although popular for market-based technoeconomic analysis, results in lower profits than bidding. Despite this fact, much of the current technoeconomic analysis of novel, more flexible energy concepts are done via self-schedule and their value may not be fully estimated.

Bids submitted by generators enable flexibility in the system's power output and schedule, and with more flexibility, the market has more options to meet ever-increasing demand. Therefore, for the technoeconomic analysis of flexible energy system concepts, simulating their market performance while bidding is essential. But this evaluation requires models to predict energy dispatch calculated from resource bids. Unfortunately, Production Cost Models (PCMs), which mimic market clearing by ISOs, are 'data-hungry'; they require knowledge of all generation resources in the grid, network topology, demand, and renewables forecasts, etc. Much of this required data is private or protected, which makes PCMs challenging to use for economic evaluation.

To address this challenge, we propose a shortcut algorithm to estimate dispatch schedules for individual market participants, requiring only generator characteristics, bid curves, and historical LMPs. Figure 1 shows the three-step process, which includes solving two mixed integer linear programs (MILP). To evaluate the proposed method, we simulate a single generator in the open-source RTS-GMLC data set ("GridMod/RTS-GMLC") over a month-long horizon using a rolling-horizon algorithm. The resulting dispatch is then compared to results from conducting a full market clearing using the open-source Pyomo-based PCM Prescient ("Prescient").

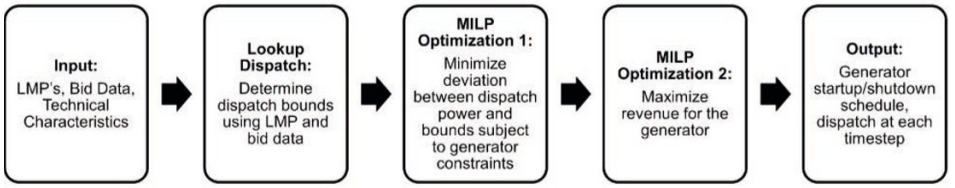


Figure 1: Shortcut Market Simulator Process

2. Methods

Figure 1 summarizes the proposed shortcut market simulator algorithm. The input data are: π_h^{real}, historical LMPs; piecewise "bid curve", a set of power, B_{hl}, and price, π_{hl}, pairs that communicate the total operational costs for the generator; and, technical characteristics including minimum and maximum power output, uptime and downtime constraints, and ramping limits. The latter are used in the thermal generator MILP model adapted from Arroyo and Conejo (2000) and Carrión and Arroyo (2006). The MILP optimization problems shown in Figure 1 are described below. The full simulation can be conducted in one-shot or using a rolling horizon algorithm. The rolling horizon algorithm solved a 24-hour horizon subproblem (from hour 0 to hour 23), saving the results of the first timestep, fixing that timestep, and solving another 24-hour horizon beginning at the next hour (from hour 1 to hour 24 with hour 1 fixed).

2.1. Sets and Variables

All equations in the MILP models are indexed over 2 sets: set $h \in H$ represents the timesteps in the horizon and set $l \in L$ represents the points on the bid curve, or each individual power-price pair. The MILP models include five sets of decision variables. Variable p_h represents the power output of the generator and time h. Variable B_h represents the bid power (bound by the lookup dispatch algorithm) for the generator at time h. Both p_h, and B_h are continuous variables. The remaining three variables are discrete: y_h represents the on/off state of the generator at timestep h (0 is off, 1 is on), y_h^{SU} represents if the generator *is* starting up at timestep h, and y_h^{SD} represents if the generator is shutting down at timestep h.

2.2. Lookup Dispatch Algorithm

The lookup dispatch algorithm compares the LMP, π_h^{real}, to the generator's bid curve prices, π_{hl}, at each timestep of the horizon (the bid curves may be either static, i.e., time-invariant, or indexed by time). The algorithm sets upper and lower bounds, $\underline{B_h}$ and $\overline{B_h}$, on the bid power at that timestep, B_h, according to where on the bid curve the LMP falls. If the LMP is larger than the highest price on the bid curve, the generator has low marginal costs and has cleared the market for that timestep, therefore will be constrained to maximum power output,P^{max}. If the LMP is lower than the lowest point on the bid curve, the generators marginal costs are higher than electricity price at that timestep, so the generator is constrained to either shutdown (zero power output) or operate at minimum power,P^{min}. If the LMP falls between two points on the bid curve, the dispatch of that generator is expected to fall between the associated power values of those points $(B_{hl} \leq B_h \leq B_{h(l+1)})$.

2.3. MILP Optimization Problem

After the bid power bounds are set, a multiobjective optimization problem is solved for the final dispatch of each generator:

$$\min \quad \Delta \tag{1a}$$

$$\max \quad \sum_{h \in H} \underbrace{\pi_h^{real} p_h}_{A} - \underbrace{(\pi_h^0 B_h^0) y_h}_{B} - \underbrace{\sum_{l=1}^{N} \pi_h^l \delta_{hl}}_{C} - \underbrace{c^{SU} y_h^{SU}}_{D} \tag{1b}$$

$$\text{s.t.} \quad \underline{B_h} \leq B_h \leq \overline{B_h} \ \forall h \tag{1c}$$

$$\Delta = \sum_{h \in H} |p_h - B_h| \tag{1d}$$

$$0 \leq \delta_{hl} \leq B_{hl} - B_{h(l-1)} \quad \forall h, l \tag{1e}$$

$$p_h = P^{min} y_h + \sum_{l=1}^{N} \delta_{hl} \quad \forall h \tag{1f}$$

$$\sum_{h \in H} |p_h - B_h| \leq \Delta^* + \varepsilon \tag{1g}$$

The first objective function Eq.(1a) minimizes the sum of deviations for the generator, Δ. The second objective function Eq.(1b) maximizes the revenue of the generator over the entire horizon. Term A represents the profit from the final dispatch, term B represents the minimum operating costs which are represented by the first point on the bid curve, term C is a linear representation of the bid curve of the generator, and term D is the start-up cost (this term is zero if generator is not starting up at timestep h i.e. $y_h^{SU} = 0$) It is constrained by Eq.(1c), bounds on the bid power for each timestep, and Eq.(1d), the definition of deviation between final dispatch, p_h, and bid power, B_h. The continuous auxiliary variable δ_{jhl} is a linear correction for the piecewise bid curve. Eq.(1e) and Eq.(1f) describe the variable's behavior, which allows the selection of the proper segment of the piecewise bid curve when π_{hl} is increasing in l, i.e., the piecewise cost curve is convex. The thermal generator model also adds constraints to the problem and includes all the discrete decisions for the generator (whether it is on/off, starting up, or shutting down at each timestep). To solve the problem, objective functions are solved using lexicographic ordering, placing full priority on Eq.(1a) first, then optimizing with the second objective. To ensure the minimum deviation value is enforced in the second objective, constraint Eq.(1g) is added to constrain the deviation between the optimized first objective, Δ^*, and a small number ε (approximately 10^{-2}).

3. Results and Discussion

To test the shortcut market simulator algorithm, we analyze a single node from the RTS-GMLC data set named "Adams". One month of the node's dispatch was simulated using a rolling horizon algorithm. The historical LMPs came from a full market clearing simulation in Prescient. The dispatch results from the shortcut simulation and Prescient were then compared. Problems M1 and M2 were formulated in Pyomo (Hart et al. 2017) and solved using Gurobi. The optimization problem contained 194 variables (122 continuous, 72 binary) for the 24-hour sub problem solved during the rolling horizon. The total 31-day shortcut market simulator algorithm took ~532 seconds. In small-scale tests, we found the shortcut market simulator algorithm was approximately 100-times faster than conducting a full market clearing in Prescient.

Comparing the results of Prescient with the shortcut simulation revealed the accuracy of our proposed approximation. Figure 2 (left) shows the generator dispatch schedules from the shortcut simulator (solid line) and Prescient (dotted line) for one quarter of the 31-day rolling horizon case study (hours 186-372). Only three time periods in this portion of the simulation, circled in red, show differences in the dispatch profiles. When analyzing the points where the shortcut simulator's dispatch did not match Prescient's dispatch, we observed two main trends. First, the shortcut simulator heavily

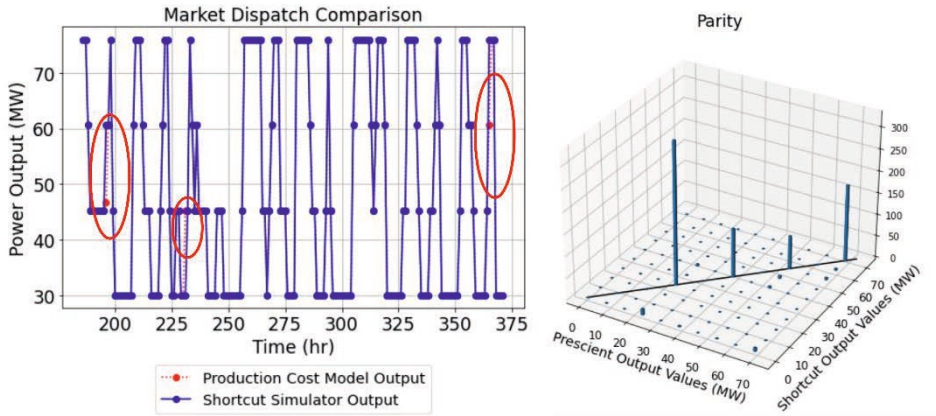


Figure 2: (left) generator dispatch schedule, comparing dispatch results from production cost model (red, dotted line) and shortcut market simulator (blue, solid line) for hours 182-372 of the 31-day simulation. The three red ovals show small discrepancies between output of the shortcut simulator and Prescient PCM. (right) parity between dispatch results from production cost model and shortcut market simulator.

favors the upper bound on bid power, set in the lookup dispatch step. Second, because Prescient makes unit commitment decisions (start-up/shut-down) in the DAM, the shortcut simulator finds different unit commitment while considering RTM prices. Figure 2 (right) shows a 3D parity plot, demonstrating the frequency of timesteps that match exactly. Approximately 67% of the dispatch points match within <1 MW. Overall, the shortcut simulator predictions had approximately 4% error in cumulative power output (summed over the entire horizon) as compared to Prescient.

4. Conclusions and Future Work

The case study provides initial validation of the proposed shortcut simulator to approach dispatch schedules using only historical LMPs, bid curves, and generator characteristics. Coupled with market participation optimization formulations (e.g., Dowling, 2017a), this can enable new approaches to estimate the economic performance of new technologies such as integrated energy systems when participating in markets. Ongoing work includes analysing all nodes of the RTS-GMLC dataset to further benchmark the accuracy of this proposed method including alternate MILP formulations. Moreover, the proposed shortcut simulator can be used to improve the realism of technoeconomic evaluations by considering bidding, the dominant mode to participating in markets, instead of the common self-schedule assumption.

Acknowledgements

This research was conducted as part of the Institute for the Design of Advanced Energy Systems (IDAES) and Design and Optimization Infrastructure for Tightly Coupled Hybrid Systems (DISPATCHES) projects. It was supported by (1) the Simulation-Based Engineering, Crosscutting Research Program within the U.S. Department of Energy's Office of Fossil Energy and Carbon Management and (2) the Grid Modernization Initiative of the U.S. Department of Energy as part of its Grid Modernization Laboratory Consortium, a strategic partnership between DOE and the national laboratories to bring together leading experts, technologies, and resources to

collaborate on the goal of modernizing the nation's grid. This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

References

Arent, Douglas J., Shannon M. Bragg-Sitton, David C. Miller, Thomas J. Tarka, Jill A. Engel-Cox, Richard D. Boardman, Peter C. Balash, Mark F. Ruth, Jordan Cox, and David J. Garfield. 2021. "Multi-Input, Multi-Output Hybrid Energy Systems." *Joule* 5 (1): 47–58.

Arroyo, J. M., and A. J. Conejo. 2000. "Optimal Response of a Thermal Unit to an Electricity Spot Market." *IEEE Transactions on Power Systems* 15 (3): 1098–1104.

Carrión, Miguel, and José Manuel Arroyo. 2006. "A Computationally Efficient Mixed-Integer Linear Formulation for the Thermal Unit Commitment Problem." *IEEE Transactions on Power Systems* 21 (3): 1371–78.

Dowling, Alexander W., Ranjeet Kumar, and Victor M. Zavala. 2017. "A Multi-Scale Optimization Framework for Electricity Market Participation." *Applied Energy* 190 (March): 147–64.

Dowling, Alexander W., Tian Zheng, and Victor M. Zavala. 2017. "Economic Assessment of Concentrated Solar Power Technologies: A Review." *Renewable and Sustainable Energy Reviews* 72 (June 2016): 1019–32.

Ela, E, M Milligan, A Bloom, A Botterud, A Townsend, and T Levin. 2014. "Evolution of Wholesale Electricity Market Design with Increasing Levels of Renewable Generation," no. September: 1–139.

"GridMod/RTS-GMLC: Reliability Test System - Grid Modernization Lab Consortium." n.d. https://github.com/GridMod/RTS-GMLC.

Hart, William E., Carl D. Laird, Jean-Paul Watson, David L. Woodruff, Gabriel A. Hackebeil, Bethany L. Nicholson, and John D. Siirola. 2017. *Pyomo — Optimization Modeling in Python*. Vol. 67. Springer Optimization and Its Applications. Cham: Springer International Publishing.

"Prescient." n.d. https://github.com/grid-parity-exchange/Prescient.

Seel, Joachim, Andrew Mills, Ryan Wiser, Sidart Deb, Aarthi Asokkumar, Mohammad Hassanzadeh, and Amirsaman Aarabali. 2018. "Impacts of High Variable Renewable Energy Futures on Wholesale Electricity Prices, and on Electric-Sector Decision Making." https://emp.lbl.gov/publications/impacts-high-variable-renewable.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50087-7

Scheduling of Material and Information Flows in the Manufacturing of Chemicals for the Order-to-Cash Process of a Digital Supply Chain

Hector D. Perez[a], John M. Wassick[b], Ignacio E. Grossmann[a*]

[a]*Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh 15213, USA*
[b]*The Dow Chemical Company, 2211 H.H. Dow Way, Midland 48674, USA*
grossmann@cmu.edu

Abstract

A scheduling model is proposed to schedule order transactions and manufacturing operations in the order-to-cash process of a digital supply chain. The proposed model is compared to scheduling models that focus on either the order transactions or the manufacturing operations. The advantage of the integrated approach is found in the accuracy of the solutions attained, whereas the purely transactional model is found to be suboptimal, and the production scheduling model is found to be infeasible under certain circumstances. An illustrative example is presented where the integrated model is shown to increase both the system revenue (60% increase) and order-fulfilment (100% increase), compared to the transactional model. The production scheduling model is shown to be infeasible and overestimate the system revenue.

Keywords: Scheduling, Business Processes, Supply Chain.

1. Introduction

With the advent and widespread drive towards digitalization in the fourth industrial revolution, a clear opportunity has emerged for a more holistic approach to supply chain management. This endeavour requires reimagining supply networks as systems that unite physical, information, and financial flows, with multiple interactions across the enterprise where material, data, humans, and intelligent agents interact in a coordinated fashion (Büyüközkan and Göçer, 2018). Within the PSE community, Laínez and Puigjaner (2012) have called for an integrated approach to Supply Chain Management (SCM), with a shift from operations-based decision support systems to decision frameworks that integrate operational, financial, and environmental models.

In the last two decades, research has begun to respond to these trends and address this need for integration. One such study in this space is that of Guillen, et al. (2006), who present a planning/scheduling model for a chemical supply chain that integrates process operations and financial decisions. This work highlights the value obtained when financial and material flows are integrated in the scheduling decisions. However, their approach does not consider information flows in the supply chain, which is an area that has not received much attention. Information flows are captured in business processes, which model the transactions that occur on requests to the system involved. In supply chain, these requests can be external customer orders, such as in the order-to-cash process, or replenishment orders, such as in the procure-to-pay process. Scheduling events in business processes has been studied by the computer science and information systems communities (Xu, et al., 2010). The business process scheduling done in these works

targets purely transactional business processes, such as banking processes that are executed in the cloud (Hoenisch, et al., 2016). However, when physical goods are involved, such as in material procurement or physical goods sales, the associated business processes become tightly coupled with the material flows in the system. Although scheduling business processes in this context has not received much attention, their close integration with physical flows is critical in chemical supply chains, where business processes like the order-to-cash process depend on the availability of inventory and the manufacturing of goods.

The scheduling of business process transactions in supply chain has been the focus of previous work by the authors. In their prior work, scheduling models have been applied to optimize the performance of the order-to-cash business process in a digital supply chain (Perez et al. 2020; 2021). However, the models have been applied primarily to the information flows in the supply chain and represent any physical processes as nodes in the transactional process network with a lumped processing time. The goal in this work is to extend what has been done previously by integrating a batch chemical manufacturing scheduling model with the order fulfilment supply chain model. The aim is to provide a more complete and accurate view of the supply chain by coupling material flows from chemical processing with the information flows from business processes. Thus, this work takes a step forward in the development of holistic decision support systems for digital supply chains.

2. Problem Statement

The order-to-cash business process manages the sequence of transactions that occur when a customer places an order. At each step, one or more agents is capable of performing certain transaction on an order. These agents can be human agents (e.g., planner, freight forwarder, customer service representative) or digital agents (for automated steps). Agents can be dedicated to a specific transaction, or they can be flexible such that they can perform transactions at different steps in the business process.

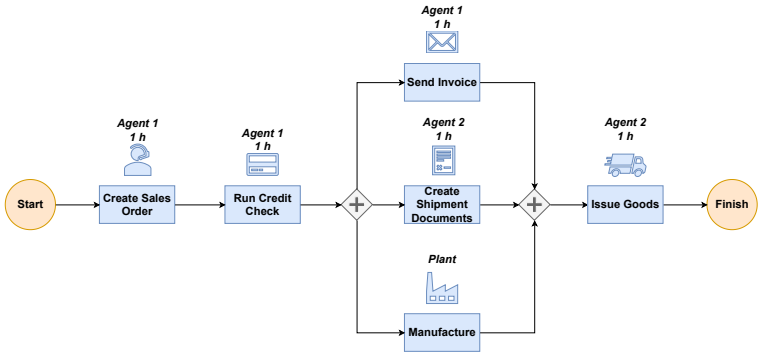


Figure 1. Illustrative order fulfilment process in a chemical supply chain

Consider the illustrative example in Figure 1, which shows a simplified customer order fulfilment process with five business transactions and one manufacturing step. The credit check step is a representation of many things that could hold or delay an order from being released to manufacturing. The invoice creation, shipment document creation, and manufacturing steps can be performed in parallel. Two agents are available to perform the five transactions on the orders as indicated in Figure 1. The manufacturing node can

represent a batch chemical plant as the one in Kondili, et al. (1993), shown in Figure 2. The plant flowsheet involves a heating step, three reaction pathways, and a purification step to produce products P1 and P2 from raw materials A, B, and C. The main equipment in the batch plant includes a heating vessel with 100 kg capacity, two reactors with 50 and 80 kg capacities, and a distillation column (still) with 200 kg capacity. Intermediate storage tanks include a 100 kg tank for hot A, a 150 kg tank for BC, a 200 kg tank for AB, and a 100 kg tank for E. Raw material and final product storage are uncapacitated. Processing times are indicated next to each transaction in Figure 1 and each manufacturing step in Figure 2.

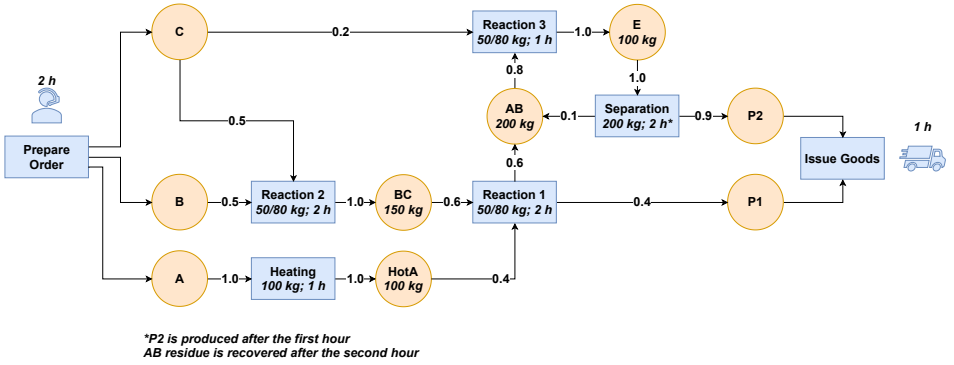


Figure 2. Flowsheet for batch chemical plant

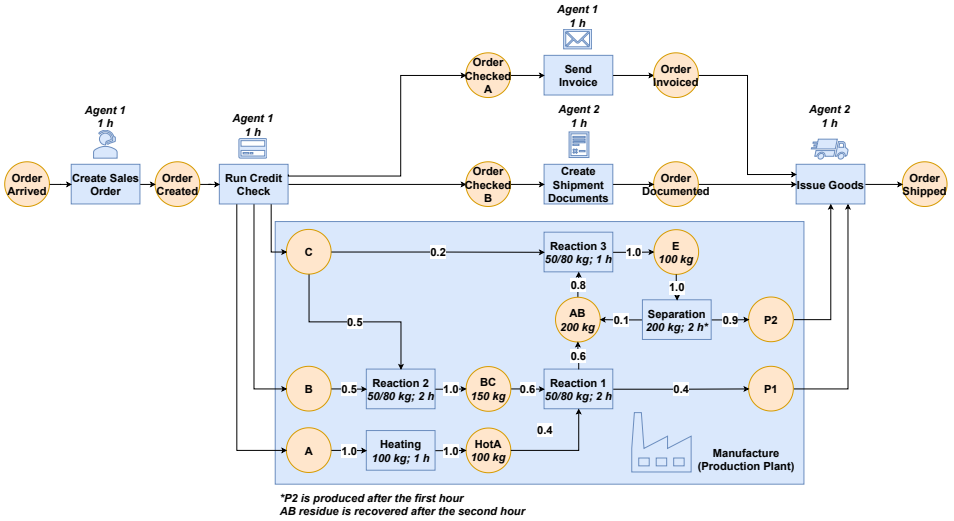


Figure 3. State-Task Network representation for the integrated model

3. Mathematical Model

The process can be modelled as purely transactional, as shown in Figure 1, by viewing the manufacturing node as a transactional step with a fixed processing type for each product type (4 h for P1 and 6 h for P2). Alternatively, from a purely plant operations standpoint, the system can be modelled using the flowsheet in Figure 2, adding a 2 h delay

after an order enters the system and a 1 h delay after a batch of product is produced to account for the first two and last order transactions, respectively. A third option, the proposed approach, is to model the system holistically, accounting for the order transactions and the detailed chemical plant model as shown in Figure 3. In this approach, the stochiometric amount of each raw material for each order quantity is made available when the credit check step is completed. Producing one unit of P1, requires one unit of A, 0.75 units of B, and 0.75 units of C. Produce one unit of P2, requires 0.59 units of A, 0.44 units of B, and 0.66 units of C. For each of the three modelling approaches, a State-Task Network (STN) model (Shah, et al., 1993) is used to schedule the system events.

The objective function of the optimization models is to maximize the revenue as indicated by the first term in Eq. (1). For the purely physical model (plant model) and the integrated model (transactions + plant model), a small ϵ penalty (10^{-4}) is assigned to the binary task triggering variables for the plant tasks (heating, reactions, and separation) to force the optimizer to favour fewer large batches over many small batches. The margin (revenue) for each order is modelled as a monotonically decreasing piecewise linear function. Eq. (2) gives the upper bound on the order margin (z_o), where T_o^{fulfil} is the time that order o is fulfilled, and $m_{i,o}$ and $b_{i,o}$ are the slope and intercept parameters for each linear function i. The discontinuity in the order margin function occurs at the order due date (t_o^{due}), where a penalty is assessed because of backordering ($m_{1,o} \cdot t_o^{due} + b_{1,o} \geq m_{2,o} \cdot t_o^{due} + b_{2,o}$). The fulfilment time, T_o^{fulfil}, is constrained by Eq. 3, where $D_{o,t}$ is a binary variable used to indicate that order o was completed at time t and leaves the State-Task Network (external consumption term in the state balance). Backordering is governed by the binary variable B_o as shown in Eq. 4. Eq. 5 forces unfulfilled orders to have zero revenue. F_o is a binary variable that indicates if an order was fulfilled within the scheduling horizon, as shown in Eq. 6. The disjunctions in Eq. (2) and Eq. (5) are reformulated using Big-M constraints.

$$\max \sum_o z_o - \sum_{k \in K^{plant}} \sum_{r \in R_k} \sum_t \epsilon \cdot W_{k,r,t} \tag{1}$$

$$\begin{bmatrix} \neg B_o \\ z_o \leq m_{1,o} \cdot T_o^{fulfil} + b_{1,o} \end{bmatrix} \vee \begin{bmatrix} B_o \\ z_o \leq m_{2,o} \cdot T_o^{fulfil} + b_{2,o} \end{bmatrix} \quad \forall o \tag{2}$$

$$D_{o,t} \cdot t \leq T_o^{fulfil} \quad \forall o, t \tag{3}$$

$$B_o = 1 - \sum_{t \leq t_o^{due}} D_{o,t} \quad \forall o \tag{4}$$

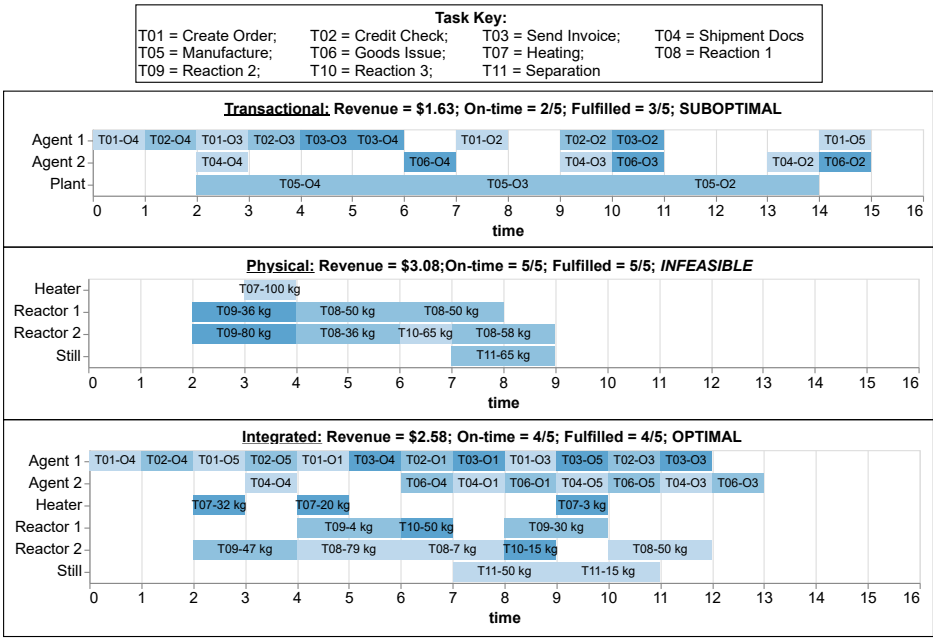
$$\begin{bmatrix} F_o \\ z_o \leq z_o^{UB} \end{bmatrix} \vee \begin{bmatrix} \neg F_o \\ z_o \leq 0 \end{bmatrix} \quad \forall o \tag{5}$$

$$F_o = \sum_t D_{o,t} \quad \forall o \tag{6}$$

4. Illustrative Example

In the illustrative example depicted in Figure 1, five orders are generated with random due dates and order margin parameters. Orders 2, 3, and 4 are for material P1, and orders 1 and 5 are for material P2. The demand of each material is also sampled randomly with a mean of 25 kg. A scheduling horizon of 15 h is used. The three modelling approaches (purely transactional with lumped plant processing times, purely physical plant model with upstream and downstream delays, and the integrated model) are implemented in JuMP 0.21 (Julia 1.6), using CPLEX 20.1 as the optimizer on a PC with an Intel i7, 1.9 GHz processor with 24 GB of RAM. CPLEX is allowed to access all 8 threads. The problem is relatively small (approximately 1,400 binary variables, 290 continuous variables, and 4,200 constraints for the integrated model), and solves within 1 s or less.

Figure 4 shows the results for each of the three scheduling approaches. The limitations of the purely transactional or purely physical models are seen in the results obtained. The purely transactional model ignores the integration of physical flows, making it suboptimal. Because intermediate AB, which is required to produce P2, is a by-product of P1 and P2, the time to produce P2 can be decreased when a batch of P1 has already been produced or is being processed alongside an order for P2. On the other hand, the purely physical model ignores the resource constraints on the transactional side, producing an infeasible schedule. The infeasibility arises from the fact that there are not enough agents to perform the first two steps on each order in the first 2 hours of the schedule. Thus, the production of intermediates for all orders cannot begin at t = 2 h. Furthermore, the model assumes that there are enough agents to issue goods once they are ready, which overestimates the system revenue as not all orders can be fulfilled immediately after the material is produced. In contrast, the integrated model finds the optimal schedule which fulfils 80% of the orders in the 15 h horizon, accounting for both agent availability and process integration at the plant.



5. Conclusions

An STN-based scheduling model is proposed to schedule orders in the order-to-cash process of a chemical supply chain. The information flows in the order-to-cash process are integrated with the physical flows of the manufacturing facility. An illustrative example is given, in which the model that integrates the transactional and the material flows is shown to attain a 60% improvement in terms of revenue over the model that lumps the material flows in a single manufacturing node. The integrated model also doubles the number of orders fulfilled in the scheduling horizon. The manufacturing-only model that lumps the initial order transactions and the goods issue step, is shown to yield infeasible schedules in a make-to-order system when the transactional steps are resource constrained. The infeasibility demonstrated in the illustrative example is indicative of actual circumstances encountered in industrial supply chains. The lack of rigorous coordination between manufacturing scheduling and order processing often leads to telephone calls and email exchanges between the scheduler and customer service representative to ultimately resolve conflicts between their respective domains. The proposed modelling approach is a first attempt to integrate the different flows involved in a digital supply chain. Future work will include adding financial flows (accounting ledger), and extending the material flows to those in a multi-echelon supply network.

References

Büyüközkan, G., Göçer, F., 2018, Digital Supply Chain: Literature Review and a Proposed Framework for Future Research, Computers in Industry, 97, 157–77.

Guillén, G., Badell, M., Espuña, A., Puigjaner, L., 2006, Simultaneous optimization of process operations and financial decisions to enhance the integrated planning/scheduling of chemical supply chains, Computers and Chemical Engineering, 30, 421-436.

Hoenisch, P., Schuller, D., Schulte, S., Hochreiner, C., Dustdar, S., 2016, Optimization of complex elastic processes, IEEE Transactions on Services Computing, 9, 700-713.

Laínez, J.M., Puigjaner, L., 2012, Prospective and perspective review in integrated supply chain modeling for the chemical process industry, Current Opinion in Hcemical Engineering, 1, 430-445.

Kondili, E., Pantelides, C.C., Sargent, R.W.H., 1993, A general algorithm for short-term scheduling of batch operations-I. MILP formulation, Computers and Chemical Engineering, 17, 211–227.

Perez, H.D., Amaran, S., Erisen, E., Wassick, J.M., Grossmann, I.E., 2020, Optimization of Business Transactional Processes in a Digital Supply Chain, Computer Aided Chemical Engineering, 48, 1159–1164.

Perez, H.D., Amaran, S., Erisen, E., Wassick, J.M., Grossmann, I.E., 2021, A Digital Twin Framework for Business Transactional Processes in Supply Chains, Computer Aided Chemical Engineering, 50, 1755–1760.

Shah, N., Pantelides, C.C., Sargent, R.W.H., 1993, A general algorithm for short-term scheduling of batch operations-II. Computational issues, Computers and Chemical Engineering, 17, 229-244.

Xu, J., Liu, C., Zhao, X., Yongchareon, S., Business Process Scheduling with Resource Availability Constraints, 2010, Lecture Notes in Computer Science, 6426, 419-427.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50088-9

Optimization of Maximum Completion Time of Polymerization Section Based on Improved Estimation of Distribution Algorithm

Jian Su[a], Yuhong Wang[a*], Su Zhang[a], Xiaoyong Gao[b]

[a]*College of Control Science and Engineering, China University of Petroleum, Qingdao 266580, Shandong, China;*
[b]*Department of Automation, China University of Petroleum, Beijing 102249, China*
Y.H.Wang@upc.edu.cn

Abstract

The polymerization section is an essential link in the production process of PVC, and the quality of its scheduling tasks directly affects the benefits of the enterprise. Given the scheduling optimization problems of unreasonable allocation of tasks and low efficiency, an improved estimation of distribution algorithm (IEDA) is proposed in this study to minimize the maximum completion time. The elitism strategy and the binary search strategy are employed to improve the global search ability and the solution speed of the estimation of distribution algorithm (EDA). Then the effectiveness of the proposed algorithm is verified in a case. The result of comparison demonstrates that compared with the EDA, the maximum completion time of the 3×5SP problem solved by the IEDA is reduced by 1.31% on average, and the speed of convergence is accelerated, which verified the accuracy and effectiveness of the IEDA in solving the scheduling problem of the PVC polymerization section.

Keywords: Process systems; Estimation of distribution algorithm; Polymerization; PVC; Scheduling optimization

1. Introduction

Polyvinyl chloride (PVC) is made of vinyl chloride monomer polymerization. The polymerization section is an imperative link. It is essential to develop a scientific and reasonable scheduling scheme to make full use of existing resources, and improve the economic benefits of enterprises.

PVC polymerization section is a batch production with large-scale and high complexity processes. How to optimize the solution of its model and improve the efficiency has attracted more attention. Wang et al. (2016) established the whole process model of PVC production, proposed an optimal decomposition algorithm based on hierarchical division, and decomposed the complex MINLP problems into MILP and NLP problems for solving. Gao et al. (2017) introduced a planning optimization model based on piecewise linear approximation to process the nonlinear characteristics in actual production, reducing the complexity of the model and improving the solving efficiency. However, it becomes difficult for the exact solution based on mathematical programming to adapt to the demand for the rapid solution with the increase of solution scale in the actual production environment. The EDA is a commonly used heuristic approach in project scheduling and related areas (Zhou et al., 2021). Research reveals that EDA breaks the pattern of traditional optimization algorithms, avoids premature convergence that frequently occurs

in overcoming complex optimization problems, and has strong chain learning ability and evolutionary orientation. Therefore, it is of practical significance to solve the problem of PVC scheduling using EDA.

In this paper, the scheduling model of the PVC aggregation section is firstly described, and several strategies are proposed to improve EDA. Then, the IEDA is applied to the scheduling optimization problem to minimize the maximum completion time. Finally, a case is provided for analysis.

2. Scheduling model of PVC polymerization section

2.1. Modeling hypotheses

1) The number of polymerization reactors, PVC grades, and orders are known, each polymerization reactor can process any grade of PVC.
2) Each polymerization reactor can only produce PVC of the same grade in the same batch.
3) The same polymerization reactor can produce different grades of PVC within a scheduling period, the kettle and other operations should be cleaned when switching between different grades of PVC, resulting in waiting time.
4) The polymerization reaction time depends on the factors such as PVC grade and catalyst, temperature and pressure in the polymerization reactor.

2.2. Mathematical model

(1) Optimization objective

Assuming that the completion time on the *j*-th polymerization reactor is TE_j, the mathematical description of the optimization objective to minimize the maximum completion time *T* is:

$$T = min(max(TE_j)) \tag{1}$$

(2) Constraints

$$TE_j = \sum TH_{j,i_1,i_2} + \sum TP_{j,i}, i \in MP \tag{2}$$

$$TP_{j,i} = N_{j,i} \times T_{j,i} \tag{3}$$

$$\sum_{j=1}^{n} N_{j,i} = ceil\left(\frac{D_i}{FS \times \alpha_i}\right) \tag{4}$$

$$TS_{j,i} = TH_{j,i_1,i_2} + TC_{j,i_1} \tag{5}$$

$$TC_{j,i} = TS_{j,i} + TP_{j,i} \tag{6}$$

$$N_{j,i} \geq 0, D_i \geq 0, TS_{j,i} \geq TC_{j,i_1}, TC_{j,i} > 0, i \in MP \tag{7}$$

Eq.(1) presents the minimum completion time; Eq.(2) expresses the completion time constraint; Eq.(3) presents the polymerization time; Eq.(4) indicates the feed times constraint; Eq.(5) describes the constraints on the start time and end time of polymerization of the same grade in one polymerization reactor; Eq.(6) indicates the polymerization completion time; Eq.(7) describes the constraints on the number of feeding times, the total quantity of PVC of each grade in the order, and the start and end time of polymerization reactor processing.

3. Improved estimation of distribution algorithm

3.1. Improvement strategy

3.1.1. The elitism strategy

In the EDA, the proportion of selected dominant groups is fixed, which easily leads to local optimality. The number of dominant groups is adjusted based on the elitism strategy to more accurately express the information of solution space. Specifically, the individuals are sorted according to the fitness value; then the minimum value r_{min} and maximum value r_{max} of the ratio of the number of dominant population N to the total number of individuals in the population are set. Besides, M denotes the population size, I refers to the current iteration number, and K is the total iteration number. N is updated as:

$$N = M \times (r_{min} + (r_{max} - r_{min}) \times I/K) \tag{8}$$

3.1.2. Constructing probability model

The probability model was constructed in the following way to better describe the distribution of PVC production of different grades on the polymerization reactor. Suppose there are M individuals in the population, and N dominant individuals are selected to construct the probability model matrix P according to the fitness value function. Then, the feeding times of each grade of PVC in each polymerization reactor in dominant individuals are counted. The probability model matrix P is expressed as:

$$P = \begin{bmatrix} \frac{\sum_{k=1}^{N} b_{1,1,k}}{\sum_{k=1}^{N}\sum_{i=1}^{m} b_{i,1,k}} & \frac{\sum_{k=1}^{N} b_{1,2,k}}{\sum_{k=1}^{N}\sum_{i=1}^{m} b_{i,2,k}} & \cdots & \frac{\sum_{k=1}^{N} b_{1,n,k}}{\sum_{k=1}^{N}\sum_{i=1}^{m} b_{i,n,k}} \\ \frac{\sum_{k=1}^{N} b_{2,1,k}}{\sum_{k=1}^{N}\sum_{i=1}^{m} b_{i,1,k}} & \frac{\sum_{k=1}^{N} b_{2,2,k}}{\sum_{k=1}^{N}\sum_{i=1}^{m} b_{i,2,k}} & \cdots & \frac{\sum_{k=1}^{N} b_{2,n,k}}{\sum_{k=1}^{N}\sum_{i=1}^{m} b_{i,n,k}} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\sum_{k=1}^{N} b_{m,1,k}}{\sum_{k=1}^{N}\sum_{i=1}^{m} b_{i,1,k}} & \frac{\sum_{k=1}^{N} b_{m,2,k}}{\sum_{k=1}^{N}\sum_{i=1}^{m} b_{i,2,k}} & \cdots & \frac{\sum_{k=1}^{N} b_{m,n,k}}{\sum_{k=1}^{N}\sum_{i=1}^{m} b_{i,n,k}} \end{bmatrix} \tag{9}$$

Where, $1 \le i \le m$, $1 \le j \le n$, $1 \le k \le N$; $b_{i,j,k}$ represents the polymerization matrix code of the k-th individual in the dominant population, the feeding times of grade i PVC on the j-th polymerization reactor. The updating mode of P is shown in Eq.(10).

$$p_{i,j}(g+1) = (1-\beta)p_{i,j}(g) + \beta\frac{1}{N}\sum_{k=1}^{N} C_{i,j}^{k} \tag{10}$$

Where, $p_{i,j}(g)$ represents the probability matrix of the g-th generation, binary variable $C_{i,j}^{k}$ takes the value one when grade i PVC in the k-th individual is produced on the j-th polymerization kettle. β represents the learning rate of P, and the value is adaptive, as shown in Eq.(11), in which a_1 and a_2 are set to 0.7 and 0.4, respectively.

$$\beta = a_1 - a_2 \times (g/g_{\max}) \tag{11}$$

3.1.3. Binary search strategy

The binary search strategy is adopted to update individuals, so as to improve the breadth of algorithm search and the speed of convergence. The specific process is detailed as follows.

Firstly, the probability model matrix P and individual coding matrix B are constructed according to Eq.(12).

$$P=\begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,m} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,m} \\ \vdots & \vdots & \cdots & \vdots \\ p_{n,1} & p_{n,2} & \cdots & p_{n,m} \end{bmatrix}, \quad B=\begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,m} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,m} \\ \vdots & \vdots & \cdots & \vdots \\ b_{n,1} & b_{n,2} & \cdots & b_{n,m} \end{bmatrix} \tag{12}$$

Where, $p_{i,j}$ represents the probability of polymerization production of grade i PVC in the j-th polymerization reactor. The higher the probability value is, the more likely it is to choose this polymerization reactor for production, and the more the polymerization quantity is. And $b_{i,j}$ indicates the number of feeding times of grade i PVC in the j-th polymerization reactor.

Secondly, the maximum and minimum values corresponding to the i-th column of matrix B and matrix P and polymerization kettle number are determined. Finally, individual updates according to whether the PVC quantity corresponding to the individual and the distribution mode of the polymerization kettle comply with the combination mode described by the probability model.

3.2. IEDA

The steps for solving the PVC production scheduling problem are described as follows.

1) The initial population was randomly initialized using a coding method based on the sequence of aggregation and feed times.
2) The fitness value of each individual in the population was calculated according to the PVC production scheduling mathematical model.
3) The individuals with higher fitness values were selected to form the dominant group following the elitism strategy.
4) The probability matrix of different grades of PVC produced in different polymerization reactors was constructed depending on the dominant group obtained in step 3.
5) The coding mode of individuals in the population was transformed into aggregation matrix coding, and the binary search strategy was adopted to perform individual updating.
6) If the set number of iterations is reached, end the iteration, output the optimal individual and decode; otherwise, turn to step 2.

4. Case analysis

An actual industrial case study was conducted using MATLAB R2017b programming with the PVC polymerization section process as the research object to further verify the performance of IEDA. Experimental environment: Intel(R) Core(TM) i5-8400 CPU, 2.80GHz processor, 8.00GB memory, operating system Win10.

4.1. Case parameters

Two examples are selected: one is scheduling optimization problem of 3 polymerization reactors with 5 grades of PVC (3×5PS), the order is {700, 950, 1500, 1100, 500}; the other is scheduling optimization problem of 4 polymerization reactors with 3 grades of PVC (4×3PS), the order is {700, 1100, 680}.

4.2. Analysis of simulation results

Simulation experiments were conducted to verify the effectiveness of IEDA in solving the PVC scheduling problem. Genetic algorithm (GA), particle swarm optimization (PSO) and estimation of distribution algorithm (EDA) were used for simulation comparison experiments. Their parameters are set as follows: the maximum number of iterations of the algorithms is 1000, the population size is 300, the mutation probability of GA is 0.09, and the crossover probability is 0.7. The two learning factors of PSO are 1.62, and the inertial weights are 0.5 and 0.8, respectively.

Table 1 presents the maximum (T_{max}), minimum (T_{min}), average (T_{avg}) completion time, relative value (R) of IEDA to EDA and standard deviation (δ) of completion time (T) obtained through 20 experiments conducted by the algorithms under two examples.

Table 1 Optimization results of the algorithms

cases	algorithms	T_{max}	T_{min}	T_{avg}	R	δ
4×3PS	GA	192.3	187.5	189.61	--	1.268
	PSO	190.6	187.1	188.17	--	1.026
	EDA	185.3	180.4	183.76	--	1.137
	IEDA	183.6	179.3	181.45	1.25%↓	1.015
3×5PS	GA	401.4	399.1	399.35	--	0.494
	PSO	402.1	400.7	401.28	--	0.473
	EDA	398.5	397.0	397.88	--	0.445
	IEDA	393.3	391.7	392.66	1.31%↓	0.389

As revealed in Tables 1, the completion time obtained by IEDA are shorter than those obtained by other algorithms in the comparative experiments, besides, as another important evaluation index of algorithm performance, the standard deviation of IEDA is lower than others, which demonstrates that IEDA has better convergence ability and solution stability.

With the purpose of directly reflecting the superiority of IEDA in optimal solution and convergence speed, the average convergence curves of GA, PSO, EDA and IEDA for solving 3×5PS scheduling optimization problems 20 times were drawn, as illustrated in Figure 1.

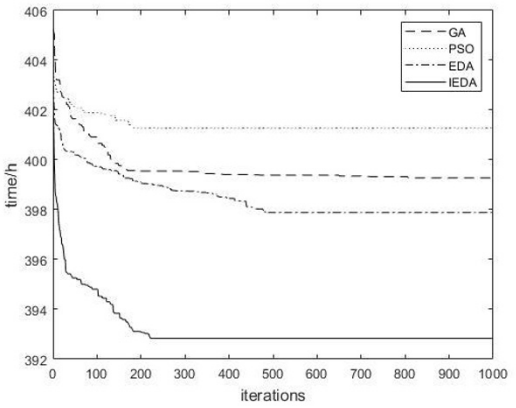


Fig.1 Optimization curves of the algorithms when solving 3×5PS problem

The convergence curves of the algorithms in Figure 1 suggest that after several iterations, the optimal solution solved by IEDA is superior to other algorithms, and the convergence speed and accuracy of IEDA are superior to EDA.

The Gantt chart corresponding to the optimal solution obtained by IEDA in 20 experiments is exhibited in Figure 2. Regarding the convenience of representation, a slash '/' and two groups of numbers before and after are used in the Gantt chart to indicate the PVC grade to be polymerized on each polymerization reactor and the feeding times.

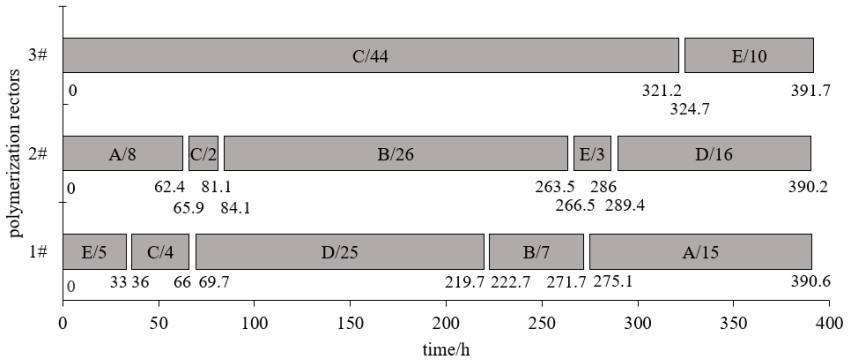


Fig.2 Optimal scheduling result of 3×5PS solved by IEDA algorithm

The analysis of completion time, evolution curve, and Gantt chart demonstrates that IEDA is superior to GA, PSO and EDA in terms of stability, convergence speed and accuracy, optimization performance, and comprehensive performance in solving PVC production scheduling problems. This verifies the effectiveness of IEDA in optimizing PVC production scheduling problems.

5. Conclusions

In this paper, a mathematical model of scheduling optimization problem of the PVC polymerization section is established, and IEDA is proposed to minimize the maximum completion time of polymerization. The results of the comparison of algorithms reflect that IEDA has higher convergence speed and optimization ability. Furthermore, as a general strategy, IEDA is applicable for the solution of PVC whole process production model and other similar process industries scheduling problems.

Reference

Y. H. Wang, X. Lian, X. Y. Gao, Z. H. Feng, D. X. Huang, T. Chan, S. S. Liu, J. X. Bai, 2016, Multiperiod Planning of a PVC Plant for the Optimization of Process Operation and Energy Consumption: An MINLP Approach, Industrial & Engineering Chemistry Research, 55, 48, 12430-12443.

X. Y. Gao, Z. H. Feng, Y. H. Wang, X. Huang, 2018, Decomposition algorithm for PVC plant planning optimization based on piecewise linear approximation, CIESC Journal, 69, 03, 953-961.

Y. F. Zhou, J. D. Miao, B. Yan, Z. S. Zhang, 2021, Stochastic resource-constrained project scheduling problem with time varying weather conditions and an improved estimation of distribution algorithm, Computers & Industrial Engineering, 157.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50089-0

Evolutionary Algorithm-based Optimal Batch Production Scheduling

Christian Klanke[a*], Engelbert Pasieka[a*], Dominik Bleidorn[b], Christian Koslwoski[b], Christian Sonntag[b], Sebastian Engell[a]

[a]*Process Dynamics and Operations Group, TU Dortmund University, Emil-Figge-Straße 70, 44227 Dortmund, Germany*
[b]*INOSIM Software GmbH, Joseph-von-Fraunhofer-Str. 20, 44227 Dortmund, Germany*
{christian.klanke, engelbert.pasieka}@tu-dortmund.de

Abstract

In this work, a simulation-optimization strategy is applied to a benchmark scheduling problem from the pharmaceutical industry, as published by Kopanos, et al. (2010). The optimization is performed by a meta-heuristic using a commercial Discrete Event Simulation software as the schedule builder (simulation-optimization approach). Our work is motivated by commonly encountered real world scenarios where detailed simulation models of the production processes are available and can be used to validate and evaluate the schedules in the presence of many, often non-standard constraints. Moreover, the effort for re-modelling and for the maintenance is reduced by using the available simulator. The meta-heuristic applied here is an Evolutionary Algorithm and we discuss different variants of the encoding of the problem. It is demonstrated that for regular objectives the performance is similar to the tailored MILP-based solution strategy of Kopanos et al. (2010) where a two-stage decomposition strategy is employed.

Keywords: Batch Production Scheduling, Pharmaceutical Industry, Simulation-Optimization, Meta-heuristics, Evolutionary Algorithm, Discrete Event Simulation.

1. Introduction

Simulation-optimization (SO) is a versatile tool for the solution of planning and scheduling problems. In industrial practice, often simulation models of different degrees of accuracy are available from the plant design stage and/or used as a tool in operations for example to validate delivery promises and to determine bottlenecks. Some tools can represent complex constraints of the execution of the orders, include maintenance, the availability of personnel, feedstock and packing materials and the like. Also stochastic effects as e.g. disturbances or varying processing times can often be included. Usually such tools are implemented as Discrete Event Simulators (DES) where rules for the execution of the production can be implemented flexibly. Such simulators enable the end-user to model the production processes in detail and therefore to validate production schedules with respect to constraints which are difficult to formulate otherwise. The models that are built for commercial DES software are typically maintained by the industrial end users, which provides flexibility with respect to changing rules, constraints, recipes or even the set-up of the plant in an intuitive manner. If such a simulation is available for a given plant or process, it is often desired by the user to use it also for planning and scheduling purposes beyond manual generation of plans or schedules and

testing them in simulations. As the models do not conform to a specific mathematical formalism, combining simulation by the DES for the execution and performance evaluation of the schedules with meta-heuristics for schedule generation is an obvious approach. Clearly, this does not provide provably optimal or near-optimal solutions.
In contrast, mathematical programming (MP) provides exact solutions with performance guarantees. It can be applied even to large-scale problems (Harjunkoski, et al., 2014) but often the problem size and complexity leads to unacceptably long computation times or large optimality gaps. This issue is usually dealt with in a semi-heuristic manner, i.e. by employing decomposition approaches that exploit the problem structure in a tailored manner, generating sub-problems with reduced numbers of degrees of freedom. These sub-problems can be solved faster and their solutions are then combined to yield the solution of the full problem (Klanke, et al., 2021, Georgiadis, et al., 2019). MP yields solutions with a measure of optimality and the problems can be solved deterministically to proven optimality. However, as soon as decomposition approaches are used, a measure of optimality usually also is not provided, as only the optimality gaps of the sub-problems are known, but not the optimality gap of the final solution. In addition, the quality of the solutions depends on the heuristics that are employed to perform the decomposition, e.g., the assignments of orders to sub-problems in case of order-based decomposition. A major disadvantage of MP solutions is the need for expert knowledge to formulate the problem at hand and to maintain the models which is far less intuitive than parameterizing a DES.
In related work, a SO approach based on the same commercial simulator (Klanke, et al., 2021) was applied to a complex industrial make-and-pack scheduling problem for which no solutions from exact optimization approaches were available. In this work, we address the well-studied problem from Kopanos, et al., (2010), a large-scale benchmark batch scheduling problem from the pharmaceutical industry, to investigate the quality of the solutions obtained with the SO approach in comparison to those obtained from the tailored MP formulation in Kopanos, et al., (2010). There a two-step MILP decomposition strategy was proposed and the authors stated that "[…] a comparison of the solution method with elaborated metaheuristics would be of great interest." The remainder of this contribution is structured as follows: We start by giving a short overview of the case study and its key features in Section 2. Then, in Section 3, we introduce our methodology, including the representation of solutions and the genetic operators. In Section 4, we present the results of our approach for regular objectives and compare them with those obtained by Kopanos, et al. (2010). Additionally, we present and discuss results with our proposed approach for the non-regular Weighted Lateness objective where timing decisions had to be added to our approach. In the last Section we conclude our findings and present an outlook.

2. Case Study

The case study addressed in this work is taken from Kopanos, et al., (2010). It is a multiproduct batch plant with 17 units (machines) that are organized in 6 stages. The problem, which is a variant of a hybrid flow shop problem, comprises 12 instances that vary in the number of orders, the objectives, and in the storage policy. The features of the problem include limited product-unit flexibility, machine-dependent processing times, sequence-dependent changeover times, and product-specific recipes, meaning that certain jobs are not processed on some of the available stages.

This paper considers three problem instances with 30 orders, unlimited intermediate storage (UIS), and the objectives Makespan (C_{max}), Overall & Changeover Cost (O.&C.C.) and Weighted Lateness (W.L.), which are minimized. The objectives are defined as

$$C_{max} = \max(C_1, .., C_{|I|})\,[h] \tag{1}$$

$$O.\&C.C = \omega C_{max} + \Sigma_{i \in I}\, cc_i\,[h], \text{ with } \omega = 0.9 \cdot 10^3 \tag{2}$$

$$W.L. = \Sigma_{i \in I}\, \alpha E_i + \beta T_i\,[h], \text{with } \alpha = 0.9 \text{ and } \beta = 4.5 \tag{3}$$

C_i denotes the completion time of job i, cc_i denotes the sum of all changeover times multiplied with a sequence dependent impact factor associated with job i, and E_i and T_i denote the earliness and the tardiness of job i.

3. Methodology

In this section, a generic solution approach that works without decomposition of the problem is presented. We first focus on the regular objectives C_{max} and O.&C.C..
Our solution method uses modular representations for the different degrees of freedom that can be adapted according to the problem at hand. The case study has three generic degrees of freedom: the allocation of jobs or rather of their operations to units, the sequences of operations on units, and the timing of the operations. Our approach uses separate strings for all decisions similar to the approach presented in Chen, et al., (1999), to maintain flexibility in the choice of the representation. To reduce the search space, heuristics can replace some of these decisions. We tested different combinations of representations for the sequencing and the allocation. Encoding timing decisions explicitly was not necessary for the regular objectives, C_{max} and O.&C.C., because no improvements of the solution quality can be obtained if delays in the starting times of operations are included.

3.1. Encodings

For the two objectives C_{max} and O.&C.C., an encoding of the sequences and, depending on the applied strategy, an encoding of the allocation to units is used.
To keep the dimension of the search space manageable, the global sequence of orders, i.e. a single sequence $\pi = (\pi_1, \ldots, \pi_{|I|})$ permuting the set of jobs $i \in I$, which is imposed on all stages, is employed. In the simulation, this sequence is decoded by the DES software by processing the jobs in the order of appearance in the global sequence π. Consequently, the sequences of operations on each stage are tightly coupled, i.e., if job i follows i' on unit M01, i' cannot finish before i if they follow the same recipe and are therefore processed on the same units. However, as shown in (Kopanos, et al., 2010), in the schedule for the 30-product case, minimizing C_{max} under UIS-policy, better solutions can be obtained when the sequence of operations on the same units are swapped within two consecutive stages. However, encoding individual sequences for all stages S, would increase the search space significantly and lead to the need of a much larger number of calls of the simulator by the EA.
Two different ways to decide on the allocation of jobs to units were investigated. One option is to determine the allocation dynamically during the simulation using a rule that is implemented in the simulator such that the highest-priority operation is allocated to and executed by the machine that first becomes idle, and started as soon as it becomes idle,

leading to non-delay schedules. So the unit on which an operation is processed in each stage is determined by an EST heuristic.
In the second option, the allocation encoding assigns to every operation $o_i \in O_i$ of a job i a unit $u_s \in U_s$ in the corresponding stage $s \in S$. The allocation encoding $\alpha \subset O \times U$ therefore is a partial relation of the set of all operations $O = \bigcup_{i \in I} O_i$ and the set of all machines in all stages $U \in \bigcup_{s \in S} U_s$.
Employing these encodings two optimization strategies are obtained: *Strat1*, where the global sequence and the allocation are optimized simultaneously and *Strat2*, where the global sequence is optimized and the EST heuristic is used for the determination of the allocation during the simulation.
Weighted lateness is a non-regular objective where timing decisions are important, because the objective value is non-decreasing with the completion time of the scheduled jobs (Baker and Scudder, 1990). For the W.L. minimization, a simple heuristic improvement strategy was used for pre- and post-processing of the EA solutions. Prior to the optimization, the jobs were sorted according to their earliest due dates. The number of possible job permutations is still very high, because many jobs share the same due date. In a simple repair step, each operation on the last stage was delayed if it was early and the delay would not increase the tardiness of a following job.

3.2. Genetic operators

The parent and survivor selection operators are identical to those that were used in Klanke, et al. (2021), i.e. a rank-based parent selection and a rank-based/elitist survivor selection. In the latter operator, a fixed percentage of the best individuals are guaranteed to survive, while all remaining individuals are chosen via rank-based selection. Parameters for which the values are not stated explicitly in this work are also chosen as in Klanke, et al. (2021).
For the allocation chromosome, the Point Mutation operator that randomly picks an operation o and assigns it to a new unit of the same stage, and a Uniform Crossover operator that iterates over all products and stages and assigns a new unit, either from parent P1 or parent P2, with equal probability, is used.
As the mutation operator for the global sequence chromosome, we employ the Permutation Mutation operator from Eiben and Smith (2015) that cuts a sub-sequence of random length and permutes its elements before reinsertion into the chromosome. As the crossover operator, the Cycle Crossover, as reported in Larranga, et al. (1996) is used.

4. Results

In this section the solutions obtained by *Strat1* and *Strat2* are presented. The optimization was run on an i7-7700K Intel CPU under Windows 10 for approx. 2.5 h of computational time. As schedule builder, the commercial software INOSIM 13.0 which, on average, took about 2.5 seconds for a single fitness evaluation, is used. Within a computational time budget of 2.5 h, by parallelization of the fitness evaluation of all individuals of the same generation, in total 3040 evaluations could be performed per problem instance. The results of our approach are presented in Table 1 together with the results from Kopanos, et al. (2010), where the computation time was limited to 1 h. This led in some instances to non-feasible solutions for the monolithic approach (see O.&C.C. in Table 1 in Kopanos, et al. (2010)). The solutions found by the two-step MILP decomposition approach in Kopanos, et al. (2010), the construction (MILP decomp. 1st Step) and the improvement (MILP decomp. 2nd Step), serve as a benchmark for the strategies proposed in this paper. The runs for *Strat2* were repeated several times to evaluate the

reproducibility of the solution. *Strat1* could barely reach the solution quality of the MILP monolithic approach for all three objectives, whereas the second strategy Strat2 outperformed the 1st step solution in Kopanos, et al. (2010), and leads to a solution quality between the 1st and 2nd step solutions of the MILP decomposition approach. The best solution that was observed with *Strat2* for the Makespan objective came very close to the 2nd step MILP solution. For the O.&C.C. objective, the best solution was slightly better than the 1st step solution from the decomposition approach. For the W.L. the best run led to a value of 47.22 h, which was reduced to 37.07 h after the repair step. Clearly here a tailored improvement strategy is needed.

Table 1: Results of the EA and MILP approaches for the 30 batches case from Kopanos, et al. (2010). $\mu^1 = 5$, $\lambda^2 = 40$ and $N_{gen}{}^3 = 600$

Solution Approach	Makespan [h]	O.&C.C. [h]	W.L. [h]
MILP monolithic	34.81	-	428.15
MILP decomp. 1st Step	28.51	66.16	48.16
MILP decomp. 2nd Step	26.56	62.91	19.09
Strat1 (Alloc. + Seq. Enc.)	35.02	77.49	693.71
Strat2 (EST + Seq. Enc.)[4]	27.90 ± 1.17	66.17 ± 0.57	47.78 ± 0.51
Best result from Strat2	26.72	65.59	47.22
After repair step	-	-	37.07

[1]Number of children, [2]Population size, [3]Number of generations, [4]Mean and standard deviation of three runs

5. Summary, Conclusion and Outlook

In this work, we investigated the potential of a simulation-optimization approach, combining an EA and DES, for a benchmark scheduling problem, from the pharmaceutical industry.

Our proposed approach benefits from the use of existing models and only encodes the essential degrees of freedom, while the detailed schedules are built by the simulation system. This has the advantage that all constraints that are implemented in the simulator are respected by the solution so the resulting schedule is executable to the best of the available knowledge of the processes in the plant.

For the case study under consideration, the allocation and the sequence degree of freedom were encoded explicitly (*Strat1*), or only the global sequence of jobs was encoded and the allocation was determined heuristically by the simulator (*Strat2*). The encoding of only the global sequence together with the heuristic allocation provided better results due to the smaller search space of the EA. For the three investigated objective functions, the best results are between the 1st and 2nd step solutions of the benchmark approach. From a practical point of view, the solution quality can be considered as sufficient and the small differences are outweighed by the advantages of the simulation-based approach of intuitive modelling, re-use of models and the ability to implement and modify all kinds of constraints in the execution of the schedule. For the W.L. objective, the optimization of a global sequence of the orders turned out to be insufficient. Here a tailored second-stage solution is needed where the allocation and sequencing decisions on the stages are considered explicitly.

The computation time of the detailed simulation models of a commercial simulation environment is significantly higher than that needed for computing a solution with a simple job-shop model due to the larger overhead that is caused by the possibility to implement more detailed models. It can be reduced significantly by a parallelization of

the fitness evaluation. In our case, in a time span of 2.5 h the EA generated results that are similar to those obtained by the MILP decomposition approach for the Makespan objective.
Overall, the combination of a detailed discrete-event simulation and an evolutionary algorithm is attractive from an industrial point of view because of the flexibility to implement non-standard features in the simulation model, the fact that the modelling is more intuitive and the model can be modified and maintained by plant personnel. For timing-related objectives and large problem sizes, further work on suitable refinement strategies is needed.

Acknowledgements

This work was partially funded by the European Regional Development Fund (ERDF) in the context of the project OptiProd.NRW (https://www.optiprod.nrw/en).

References

Kenneth R. Baker and Gary D. Scudder, 1990, Sequencing with earliness and tardiness Penalties. A review, Operations Research, 38(1), 22–36

A.E. Eiben and J.E. Smith, 2015, Introduction to Evolutionary Computing, Springer-Verlag Berlin Heidelberg

Iiro Harjunkoski, Christos T. Maravelias, Peter Bongers, Pedro M. Castro, Sebastian Engell, Ignacio E. Grossmann, John Hooker, Carlos M´endez, Guido Sand, and John Wassick, 2014, Scope for industrial applications of production scheduling models and solution methods, Computers and Chemical Engineering, 62, 161–193

Christian Klanke, Dominik Bleidorn, Christian Koslowski, Christian Sonntag, and Sebastian Engell, 2021, Simulation-based Scheduling of a Large-scale Industrial Formulation Plant Using a Heuristics-assisted Genetic Algorithm, In GECCO ’21: Proceedings of the Genetic and Evolutionary Computation Conference Companion, Association for Computing Machinery, 1587–1595

Christian Klanke, Vassilios Yfantis, Francesc Corominas, Sebastian Engell, 2021, Short-term scheduling of make-and-pack processes in the consumer goods industry using discrete-time and precedence-based MILP models, Computers and Chemical Engineering, 154, 107453

Georgios P. Georgiadis, Apostolos P. Elekidis, Michael C. Georgiadis, 2019, Optimization-Based scheduling for the Process Industries: From Theory to Real-Life Industrial Applications, 7(7), 438

Georgios M. Kopanos, Carlos A. Méndez, and Luis Puigjaner, 2010, MIP-based decomposition strategies for large-scale scheduling problems in multiproduct multistage batch plants: A benchmark scheduling problem of the pharmaceutical industry, European Journal of Operational Research, 207(2), 644–655

Pedro Larranaga, Cindy M. H. Kuijpers, Roberto H. Murga, and Yosu Yurra-Mendi, 1996, Learning Bayesian Network Structures by Searching for the Best Ordering with Genetic Algorithms, IEEE Transactions on Systems, Man, and, Cybernetics-Pakt A: System and Humans, 26(4), 487–493

Haoxun Chen, Jürgen Ihlow, Carsten Lehmann, 1999, A Genetic Algorithm for Flexible Job-Shop Scheduling, 1999 International Conference on Robotics & Automation, 1120-1125

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50090-7

Cream Cheese Fermentation Scheduling

Misagh Ebrahimpour[a], Wei Yu[a*], Brent Young[a]

[a] *Department of Chemical and Materials Engineering, The University of Auckland, Auckland 1142, AUCKLAND*
w.yu@auckland.ac.nz

Abstract

Maintaining high throughput with consistent quality is challenging in industrial cream cheese plants since batch fermentation time varies. However, determining the batch duration right from the batch start time is challenging. This makes the scheduling of this plant difficult. The characteristics of the plant, the main process challenges, and the resulting framework, which included adaptive modelling and scheduling, are presented.

Keywords: Cream cheese fermentation, Batch Scheduling, Online pH prediction

1. Introduction

Fermentation batches are challenging to schedule due to the high inherent biological variability. Batch fermentations are common in various industries such as food, chemical, and pharmaceutical processing; therefore, much work has been carried out to schedule such systems. Harjunkoski *et al.*, (2006) worked on scheduling a copper plant. Raw materials variation affected the reaction time, which made the plant operation challenging. Reaction modelling with raw material changes was used in a mixed-integer formulation for scheduling of the overall production process. Scheduling of penicillin fermentation was studied in Lau *et al.*, (2003). However, the authors did not consider the batch variation, and a nominal batch processing time was used in their scheduling formulation. Baldo *et al.*, (2014) presented a scheduling solution for a beverage plant in the brewing industry. They use a constant fermentation time which is much longer than the mean values for the scheduling time, however the fermentation time variation was not addressed. Their schedule significantly reduced the process throughput. Additionally, the fermentation liquid product could be stored in tanks for several days.

In cream cheese plants the variation of batch duration affects the downstream continuous production rate and quality. Furthermore, the fermentation curds cannot be stored for a long time since over acidification degrades the quality. To avoid batch interferences during cooking, engineers in industry set up the fermentation scheduling with a long buffer time between two fermentation vats. This assures quality; however, the production rate is reduced significantly. Better scheduling of batches can decrease their variation by reducing possible interferences between batches due to variations of their duration. A new framework is presented in this work that provides a primary schedule with updating each batch durations predicted by a fermentation model at each time step. This schedule was updated in real-time by using an adaptive model that predicted the batch duration along with the fermentation when enough measurements were available. A mixed-integer linear (MILP) programming optimization was formulated for real-time scheduling of the vats filling and draining. The constraints of the plant regarding the filling, draining, and cleaning of the vats were considered. The best configuration for scheduling was

determined to minimize the cost and waste and improve the continuous operation of the plant.

2. Methodology

2.1. Scheduling framework

As Figure 1 (A) shows, the downstream and upstream units connected to vats are ideally in continuous operation. The objective is to schedule vats to maintain continuous operation while considering filling, draining, and cleaning constraints, and varying batch duration. As shown in Figure 1 (B), only one vat can be drained or filled at any time due to the draining and filling line architecture. Both filling and draining take 2 hours. After reaching the desired pH, batches should be cooked immediately to stop the fermentation. If one batch's pH reaches the desired value and the draining line is used by another vat, the batch can be cooled in the buffer tank and drained later. However, this will cause more energy consumption and extra cost for the plant. Therefore, interference between batches, as explained in the above example, should be avoided. These are two significant constraints that are considered in the optimization formulation. After the batch is drained, it should be cleaned for future usage. Dairy plants use the Cleaning in Place (CIP) term for cleaning. The CIP time also varies since it is monitored online and can be stopped based on CIP measurements. For this study we used a constant CIP value that was suggested by the plant.

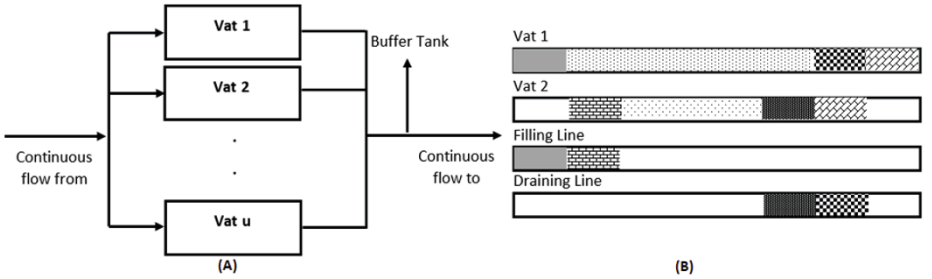


Figure 1. Unit operations: (A) Process flow diagram, (B) Vat operating conditions: filling shown by ■ and ▤, draining shown by ▦ and ▚, and CIP shown by ⚡.

The batch duration (the time required to reach the desired pH from the beginning of the batch) varies due to disturbances such as milk components changing from season to season due to cow nutrition and weather conditions. This makes the scheduling of vats a challenging task. A mixed-integer optimization has been applied for solving this scheduling problem. The scheduling routine is shown in Figure 2. When all vats are available, scheduling is carried out for all of them. The key problem is what batch duration time should we use? As mentioned before, batch duration varies. To deal with this problem, a default value is used first as the initial batch duration. The default value can be defined by engineers based on historical batch duration data. Different default values may impact the scheduling performance. Therefore, three default values, 12, 13 and 14 (hours), are investigated in this paper. After filling up the first batch, measurements from the batch beginning up to a specific time can be used in the pH prediction model to estimate the time for reaching the desired pH. This is important in the plant as the desired pH should be obtained at the end of batch draining, affecting the quality of the end-product. The updated batch duration will be used to reschedule the batches. Rescheduling will be repeated whenever enough data is available for determining the duration for each

batch. The time needed to collect enough data for the pH prediction model and predict the actual batch duration can consequently affect the scheduling performance. The optimal estimate of the initial value of the batch duration is determined by evaluating the scheduling performance discussed in Section 3.

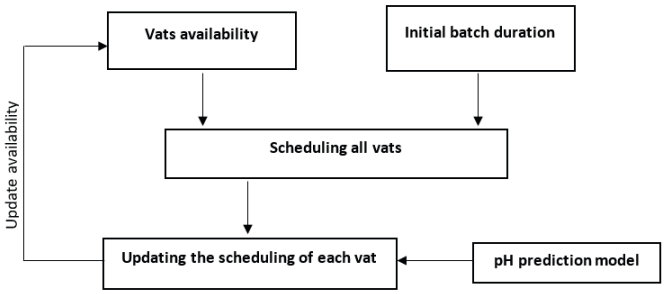


Figure 2. Scheduling framework

2.1.1 pH prediction model

Ebrahimpour *et al.*, (2021) studied the application of white, black, and grey box models for cream cheese pH prediction. A reliable pH prediction model was achieved by applying neural networks to pH dynamics. Additionally, a grey box model developed by Li *et al.* (2021) was discussed. Estimated biomass with measurements along the fermentation can make this model adaptive and improve the prediction. Both models can be used in the scheduling framework for cream cheese pH prediction.

2.1.2 Scheduling formulation

The operating conditions of the batch units were defined by two variables $W_{u,t}$ and $Y_{u,t}$ as shown in Table 1. The processing (fermentation) duration was defined between filling start time and the end of CIP. Variables $B_{u,t}$, $F_{u,t}$, $E_{u,t}$, $D_{u,t}$, and $G_{u,t}$ were used for distinguishing different occasions. The formulation and the details of the variables are given below.

$$C_{u,t} = C_{u,t-1} + W_{u,t} \qquad \forall u \in U, \forall t \in T : t > t_0 \tag{1}$$

$$C_{u,tt} - C_{u,t} \le (T_F - T_0) * (1 - B_{u,t}) \qquad \forall u \in U, \forall t \in T, \forall tt \in (t+1) \ldots \min(t + P_u - 1, T_F) \tag{2}$$

$$C_{u,t} - C_{u,t-P_u} \ge 1 \qquad \forall u \in U, \forall t \in T : \forall t \ge T_0 + P_u \tag{3}$$

$$W_{u,t} + Y_{u,t} \ge 1 \qquad \forall u \in U, \forall t \in T \tag{4}$$

$$Y_{u,t-1} - W_{u,t} \le Y_{u,t} \le Y_{u,t-1} + W_{u,t} \qquad \forall u \in U, \forall t \in T : t > t_0 \tag{5}$$

$$W_{u,t} - Y_{u,t-1} \le B_{u,t} \le \frac{W_{u,t} + Y_{u,t}}{2} \qquad \forall u \in U, \forall t \in T \tag{6}$$

$$F_{u,t+t_{FD}} = B_{u,t} \qquad \forall u \in U, \forall t \in T \tag{7}$$

$$E_{u,t_{BD}} = B_{u,t} \qquad \forall u \in U, \forall t \in T \tag{8}$$

$$D_{u,t_{DD}} = E_{u,t} \qquad \forall u \in U, \forall t \in T \tag{9}$$

$$G_{u,t_{CIPD}} = D_{u,t} \qquad \forall u \in U, \forall t \in T \tag{10}$$

$$\sum_{u=1}^{U} (B_{u,t} + F_{u,t}) = 1 \qquad \forall u \in U, \forall t \in T \tag{11}$$

$$\sum_{u=1}^{U} (E_{u,t} + D_{u,t}) = 1 \qquad \forall u \in U, \forall t \in T \tag{12}$$

T_0 : start of the solution horizon

T_F : end of the solution horizon

P_u : batch duration for batch unit u in U

U : domain of batch units 1…number of batch units

T : total time horizon from T_0 to T_F

$E_{u,t}$ = 1: if batch unit u starts draining a batch at time t, 0 otherwise (Boolean variable)

$C_{u,t}$: the number of batches and unit idle periods from time T_0 to t (integer variable)

u : batch unit number

t : time at any instant

$B_{u,t}$ = 1: if batch unit u starts filling a batch at time t, 0 otherwise (Boolean variable)

$F_{u,t}$ = 1: if batch unit u finishes filling at time t, 0 otherwise (Boolean variable)

$G_{u,t}$ = 1: if batch unit u is doing CIP at time t, 0 otherwise (Boolean variable)

$D_{u,t}$ = 1: if batch unit u finishes draining at time t, 0 otherwise (Boolean variable)

Equations (1)-(12) explanation:

Eq. (1): At any time, t, if a batch starts on a batch unit, u, or the unit is idle, a counter is incremented

Eqs. (2)-(3): Batch cycles times must be longer than the specified value

Eqs. (4)-(6): Boolean relationships for ensuring the feasibility (the Table 1 condition)

Eq. (7): Batch filling time duration specification; t_{FD} is the filling duration

Eq. (8): Batch complete time (from start time to reaching the desired pH)

Eq. (9): Batch draining time duration specification; t_{DD} is the draining duration

Eq. (10): Batch cleaning time duration specification; t_{CIPD} is the cleaning duration

Eq. (11): Batches filling constraint; only one vat can be filled at any time

Eq. (12): Batches draining constraint; only one vat can be drained at any time

The objective function maximizes the started vats which is equivalent to maximizing Y and W for all u vats at any time t.

$$Max \qquad \sum_{u=1}^{U}\sum_{t=T_0}^{T_F} W_{u,t} + Y_{u,t} \tag{13}$$

Table 1. Operating conditions of a batch unit

Variable	Start filling a batch	Processing a batch	Unit is idle	Infeasible
$W_{u,t}$	1	0	1	0
$Y_{u,t}$	1	1	0	0

3. Results

The batch duration varies in the industrial case due to disturbances such as milk composition variation and bacteria activity. Since the batch duration cannot be predicted at the beginning of the batch, a default initial batch duration was assumed to schedule the vats. Three default batch durations (12, 13 and 14 hours) were selected for testing the impact of default batch duration on the scheduling performance. As soon as enough data was measured, the scheduling would be updated by the predicted fermentation time from the pH prediction model discussed in Section 2.1.1. The effect of updating time on the scheduling was tested by considering the pH prediction model output availability 5 and 8 hours after the batch start time.

3.1. Industrial scale scheduling

Batch duration data from a real cream cheese plant was used for testing the scheduling framework performance. The duration times of 20 batches in the sequence were used, which took approximately 70 operating hours in the plant. As shown in Figure 2, when all the vats are available, scheduling was carried out for all the vats.

As mentioned in previous sections, the batch duration cannot be determined before batches start. The initial batch duration in the scheduling algorithm was assumed to be a fixed value at the beginning of all batches run. For obtaining the best initial value, scheduling was applied to the industrial batch duration data. Scheduling was carried out by considering the default batch duration as 12, 13, and 14 hours. The initial batch duration was updated by the predictions from the pH prediction model. The pH model prediction output was assumed to be available 8 hours after the batch start time. The updated batch duration was used to update the scheduling of the vats.

Table 2 summarizes the scheduling results with different initial batch durations for five vats. The performance of the scheduling framework was studied by comparing three indicators - idle time, number of cooled batches, and number of waste batches. Idle time is the summation of hours in which the draining line is not in operation. This time should be minimized in the plant as continuous operation and consequently high throughput is desired. The number of cooled batches represents the draining interference of two batches when one is cooled and drained later. Wasted batches happen when more than two batches draining coincidence happens. One of the batches can be cooled at such a time, but the other one is wasted.

Table 2. Scheduling results with different initial default batch duration

Batch duration	Idle time (h)	Cooled batches	Wasted batches
12	22	3	1
13	20	4	1
14	24	4	2

Table 3. Scheduling results with different update availability

Update availability	Idle time	Cooled batches	Wasted batches
At 5 h	17	2	0
At 8 h	22	3	1

Table 2 shows that the 12 hours initial batch duration led to less cooled and wasted batches. This means that more energy and money are saved in the plant. However, the draining line idle time is more than 13 hours batch duration. The selection between the initial batch duration options should be made based on the plant's production, economic, and quality objectives. Without rescheduling, the idle time, number of cooled and wasted batches were 24 h, 4 and 3 respectively which shows the importance of rescheduling in improving the performance.

Figure 3 shows an example of the scheduling framework performance for the five vats with an initial batch duration of 12 hours. The top part of the Figure indicates the results for an initial batch duration of 12 hours for all vats at the batch start time. The scheduling update was carried out after determining the batch duration by pH prediction model. The bottom part of the Figure shows the actual batch duration. The vat filling time was

updated after time step 13 according to the actual batch duration determined by the pH prediction model.
Scheduling performance can be improved by providing the batch duration prediction earlier. This has been studied by providing the batch predictions 5 and 8 hours after the batches start scheduling with the initial batch duration estimate of 12 hours. Table 3 shows that the earlier update of the scheduling using the pH prediction model outputs can decrease the idle time, and the number of cooled and wasted batches. This will improve the scheduling performance in terms of energy, economy, and quality.

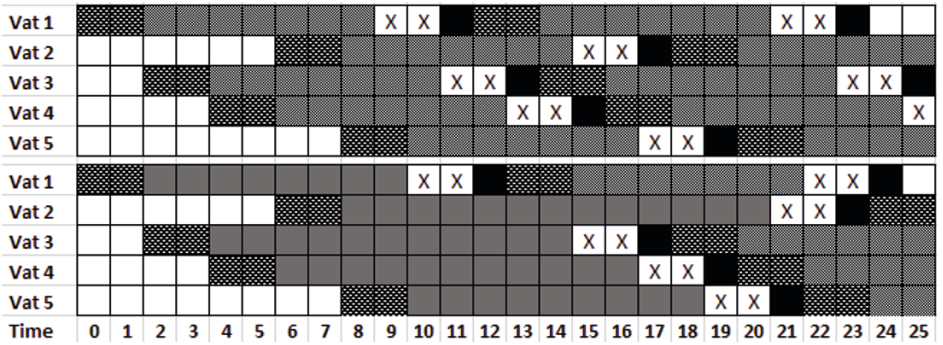


Figure 3. Scheduling of vats before (top) and after (bottom) the update. is the initial batch duration, is the actual batch duration, is the filling, X is draining, is CIP

4. Conclusions

Scheduling cream cheese fermentation is challenging since batch duration varies. This work presented a scheduling framework that included an online pH prediction model along with MILP formulation. Online lactose and lactate measurements improved the pH prediction, which was achieved by reconciling the states. The formulation used the model output to reschedule the primary schedule, which was obtained by assuming a default initial batch duration. The framework performance was tested by scheduling 20 batches in sequence. Results showed that using 12 h as the default initial batch duration with batch prediction updating 5 hours after the fermentation started led to the minimum wasted and cooled batches.

References

M. Ebrahimpour, W. Yu, B. Young, 2021, Artificial neural network modelling for cream cheese fermentation pH prediction at lab and industrial scales, Food and Bioproducts Processing 126 (2021): 81-89.
B. Li, Y. Lin, W. Yu, D. Wilson, B. Young, 2021, Application of mechanistic modelling and machine learning for cream cheese fermentation pH prediction, Journal of Chemical Technology & Biotechnology 96, no. 1 (2021): 125-133.
Lau, S., Willis, M.J., Montague, G.A. and Glassey, J., 2004. Predictive scheduling of a penicillin bioprocess plant. IFAC Proceedings Volumes, 37(1), pp.463-468.
Baldo, T.A., Santos, M.O., Almada-Lobo, B. and Morabito, R., 2014. An optimization approach for the lot sizing and scheduling problem in the brewery industry. Computers & Industrial Engineering, 72, pp.58-71.
Harjunkoski, I., Borchers, H.W. and Fahl, M., 2006. Simultaneous scheduling and optimization of a copper plant. In Computer Aided Chemical Engineering (Vol. 21, pp. 1197-1202). Elsevier.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50091-9

Multi-Objective Optimization of Production Cost and Carbon Loss in the U.S. Petrochemicals Industry

Ioannis Giannikopoulos[a], Alkiviadis Skouteris[a], David T. Allen[a,b], Michael Baldea[a,c] and Mark A. Stadtherr[a*]

[a]*McKetta Department of Chemical Engineering, The University of Texas at Austin, 200 East Dean Keeton Street, Austin, Texas 78712-1589, United States*
[b]*Center for Energy and Environmental Resources, The University of Texas at Austin, 10100 Burnet Road, Bldg. 133, R7100, Austin, Texas 78758, United States*
[c]*Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin, 201 East 24th Street, Austin, Texas 78712, United States*

markst@che.utexas.edu

Abstract

The U.S. petrochemicals industry has been strongly influenced by the shale hydrocarbon boom commenced more than a decade ago. Newly available resources and emerging technologies spurred research focused on new pathways and technologies for producing chemicals. The adoption and deployment of these technologies by the manufacturing industry is driven by the need to minimize costs while meeting demand for chemical products. However, cost minimization is a unilateral objective that may conflict with other societal, environmental or economic goals and constraints. Motivated by this, in this paper we study the optimal configuration of the U.S. petrochemicals industry by considering both economic and environmental objectives. Our work is based on a comprehensive model representing the U.S. industry as a whole, and on multi-objective optimization problem formulations that allow us to elucidate the trade-off between economic costs and net carbon loss. A Pareto-optimal set of solutions is obtained and a comparison between the two extreme cases is performed.

Keywords: Sustainability; Chemical manufacturing; Supply chain; Network modeling

1. Introduction

Advances in hydraulic fracturing and horizontal drilling have led to significant growth in the U.S. petrochemicals industry. Oil and gas production have rapidly increased in the U.S., particularly from shale formations (U.S. Energy Information Administration, 2021a). An important consequence of the shale hydrocarbon boom has been the increased production of natural gas liquids (NGLs), which are often abundant in shale gas. NGLs are composed of ethane, propane, *n*-butane, isobutene and small amounts of less volatile hydrocarbons, molecules that are valuable feedstocks for chemical manufacturing (Siirola, 2014). The increase in the availability of shale gas and associated NGLs has thus provided a unique opportunity to expand the U.S. chemical manufacturing industry. In this paper, we aim to study optimal directions for this historic industry expansion, considering a trade-off between economic and environmental objectives.

We consider this industrial sector at the level of the entire United States, and use a computational model that comprises several hundred of the highest-volume chemicals and a library of hundreds of potential technologies for producing and processing them (DeRosa and Allen, 2016; Skouteris et al., 2021). The model is formulated as a superstructure that allows for determining the optimal configuration of the industrial sector by solving an optimization problem to minimize a given industry-wide objective, subject to material balance constraints as well as supply limitations and demand requirements. The solution of the optimization problem determines the optimal production levels for each process technology, as well as material flows in the network. In this paper, we consider two industry objectives: production cost minimization and minimization of carbon loss (e.g., as CO_2 emissions). In determining carbon loss, both feedstock carbon and the carbon cost of energy used are considered.

2. Background and Problem Definition

Optimization-based network superstructure models of chemical manufacturing originated in the work of Stadtherr and Rudd (1976), and many variations of this approach, with various focuses, have appeared since. Cost minimization is a commonly used industry objective in such studies and can be used to determine an optimal network configuration (i.e., the technologies in the superstructure that have nonzero utilization). In recent work (Skouteris, et al., 2021), we developed a nonlinear, cost-based industry model that propagates cost and price changes within the network as new technology is introduced. However, there are other objectives that may affect industry behavior, such as safety and environmental performance, leading to the use of multi-objective optimization. For example, Fathi-Afshar and Yang (1985) focused on the effect of gross toxicity in process selection and how it conflicts with the minimization of cost for the industry. Similarly, Chang and Allen (1997) studied trade-offs between chlorine usage and industry costs. Here we will use an industry network model with multi-objective optimization to study trade-offs between cost minimization and carbon usage.

In the network model, the industry is represented as a directed graph, where chemical processes are the nodes and the edges correspond to material and utility flows. The data for the model (e.g., process stoichiometries and costs) have been obtained from the IHS 2012 Process Economic Yearbook (IHS Markit, 2012). The model used in this work involves 887 processes, 892 materials and 7 utility types, aiming to represent the bulk petrochemicals industry, including polymer products, in the United States. The methodology can also be used to study the chemical industries in other countries or even different industrial networks within the U.S., provided that the data are adjusted accordingly. The core model consists of balance equations for all materials, and supply and demand constraints. The balance equations for every material i are of the form:

$$F_i + \sum_j a_{i,j} X_j - Q_i = 0 \tag{1}$$

where F_i is the exogenous flow rate of material i as a primary raw material into the network, $a_{i,j}$ is the input-output coefficient for i in process j (positive if i is produced, negative if i is consumed, and unity if i is the main product), X_j is the utilization level of process j (in terms of flow rate of main product), and Q_i is the exogenous flow rate of i as a final product out of the network. The supply and demand constraints are:

$$0 \le F_i \le S_i \tag{2}$$

$$Q_i = D_i \ge 0 \tag{3}$$

where S_i and D_i are specified exogenous raw material supply and final product demand rates for *i*, respectively.

Two objective functions for optimizing the industry behavior are considered. The first objective is minimization of total processing cost (equivalent to maximization of profit since output is fixed by Eq. (3)):

$$\min_{X_j, F_i} C = \sum_j C_j X_j \tag{4}$$

where C_j is the net unit cost of process j (cents/lb of main product), including raw materials and utility costs, fixed capital investment depreciated over a 10-year period, and other fixed operating costs. The second objective is net carbon loss:

$$\min_{X_j, F_i} L = \sum_i F_i w_{C,i} + \sum_k w_{C,k} \sum_j u_{k,j} X_j - \sum_i Q_i w_{C,i} \tag{5}$$

Here, $w_{C,i}$ is the weight fraction of carbon in material *i*, $w_{C,k}$ is the carbon cost per unit of utility *k*, and $u_{k,j}$ is usage of utility *k* per unit of production from process *j*. To determine $w_{C,k}$ we consider the actual carbon content of an energy-equivalent amount of natural gas. This means that, for example, 1 kWh of electricity (equivalent to 3412 BTUs) is assumed to have the same carbon content as 3412 BTUs of natural gas. Information on the energy content of utilities used is taken from the U.S. Energy Information Administration (2021b). The first term in Eq. (5) represents the feedstock carbon input to the industry, the second term the carbon cost of utilities used, and the last term the carbon outputs (constant here due to fixed industry output). Thus, *L* represents the industry-wide loss of carbon (e.g., as CO_2 emissions). As formulated here, the carbon loss function *L* does not consider the possible generation of energy, and thus a carbon credit, by a production process. An alternative formulation for *L* that accounts for energy production can be considered and will be the subject of our future research. Both objective functions are linear so the underlying problem is a linear program (LP). Here we use these two objectives to consider a multi-objective optimization problem. We seek to determine the Pareto optimal set for this problem using the weighted sum and ε-constrained methods.

3. Solution Strategies

The LP problem defined above was first solved separately with each individual objective to obtain the two single-objective optimal configurations, indicating that the two objectives are indeed in conflict. To obtain Pareto optimal solutions using the weighted sum method, the two functions were scalarized into a single objective by assigning each function a defined weight; the weights are then varied to obtain a set of Pareto-optimal solutions:

$$\min_{X_j, F_i} f = w_1 \frac{C}{C^*} + w_2 \frac{L}{L^*}, \quad \text{with } w_1 + w_2 = 1 \tag{6}$$

Here each objective function has also been scaled using the minimum values C^* and L^* obtained by solving the two single-objective LPs.

Additional Pareto-optimal solutions were generated using the ε-constrained method, in which the LP is solved for one objective, while bounding the other objective through an additional constraint:

$$\min_{X_j, F_i} C \tag{7}$$

s.t. $L \leq \varepsilon_1$, where $L^* \leq \varepsilon_1 \leq L^0$

or alternatively:

$$\min_{X_j, F_i} L \tag{8}$$

s.t. $C \leq \varepsilon_2$, where $C^* \leq \varepsilon_2 \leq C^0$

Here C^0 and L^0 correspond to the value each objective function takes when the other objective is minimized. Varying ε_1 and ε_2 within their given ranges will result in several additional Pareto-optimal solutions.

4. Case Study

Proceeding as described above, we found the optimal industry configurations for the two conflicting objectives defined by Eqs. (4) and (5), and determined the Pareto-optimal front, namely the set of non-inferior solutions for the multi-objective problem.

The Pareto front is shown in Figure 1. Here, point A represents the optimal industry configuration when minimizing the total production cost as a single objective, regardless of the carbon loss. This point corresponds to a minimum total production cost of 187 billion dollars per year and is accompanied by a carbon loss of 90.6 billion lbs per year. Similarly, point C represents the optimal configuration when minimizing the net carbon loss as a single objective, without considering the total production cost. At this point, the minimum net carbon loss is 62.8 billion lbs per year, requiring a total production cost of 258 billion dollars per year. All other Pareto-optimal solutions between these two extreme cases represent trade-offs between minimizing production cost or carbon loss, based on the varying importance levels given to each objective. An overall trade-off can be determined by comparing the two extreme solutions. In moving from point A to C to achieve the minimum carbon loss, the total industry production cost increases by approximately 37%, while moving in the opposite direction to achieve the minimum production cost, the net carbon loss increases by approximately 42%.

Point U in Figure 1 refers to the utopia point, which is an ideal solution corresponding to the minimum values found for both objectives, but which never can be reached for these two conflicting objectives. Ideally, though, it is desirable that the Pareto front be as close as possible to the utopia point. In this case, the closeness to U will depend on the "steepness" of the slope of the Pareto front in the vicinity of point C and the "flatness" of the slope in the vicinity of point A. We note that moving leftward from point A, significant reductions of the net carbon loss can be obtained with only small increases in the production cost. Similarly, moving downward from point C, significant reductions of the production cost can be obtained with only small increases in the carbon loss.

The changes that occur in moving along the Pareto front stem from changes in the industry network configuration. Here we highlight some of the most important changes observed between the extreme points, A and C. At point C, there is high production of methanol from natural gas, which is then used for producing olefins, such as ethylene and propylene. In contrast, at point A, propylene is produced from naphtha, and ethylene is produced from propane and ethane by steam cracking. Also, there is significantly higher production of ethylene at point C and its downstream usage differs between the two

extreme points. In both cases, ethylene is used for the production of 1-butylene, polyolefin elastomer, EPDM rubber, ethylene dichloride ethylene/vinyl alcohol barrier resin, 1-hexene, polyethylenes, vinyl acetate-ethylene copolymer and vinylidene chloride. However, at point A, ethylene is also used for the production of ABS resin, methyl methacrylate, polystyrene, ethylene glycol and vinyl chloride, whereas it is used for production of ethanol, ethylene oxide and styrene at point C. These changes take place gradually throughout the network, as we move along the Pareto front. Consider, for example, point B, which is obtained from the weighted sum method when equal importance is given to both objectives. Here, ethylene is produced from a combination of steam cracking and methanol-to-olefins processes, with the production rates of these two types of processes being about the same. A similar behavior can be observed for several other chemicals in the network.

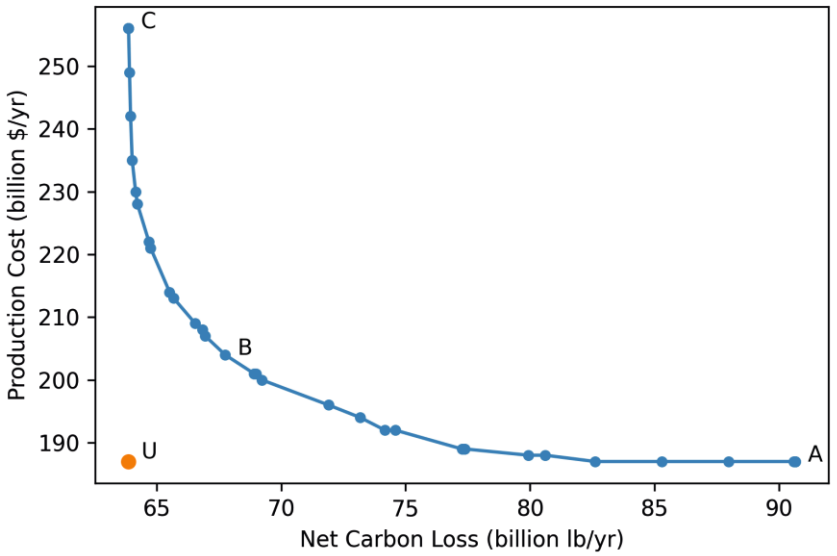


Figure 1: Pareto front and utopia point for the multi-objective optimization problem.

Comparing the amount and type of feedstock carbon that the industry network utilizes in each case is also of interest. Specifically, at point A, when production cost is the sole driving force behind the optimal industry configuration, a higher amount of feedstock carbon is utilized, compared to point C, when lost carbon content is the sole driving force. Moreover, feedstocks with lower carbon weight fraction see increased importance at point C. This leads to a situation at point C in which there is little loss of feedstock carbon, with the carbon loss function L being dominated by the carbon cost of utilities. In contrast, at point A, the contributions of feedstock carbon loss and utility carbon cost to L are roughly the same. This suggests that there may be environmental benefits to prioritizing efficient raw material usage, even if that means more usage of energy utilities. This observation will be reinforced as energy utilities are increasingly decarbonized.

5. Conclusions

In this work, we used a superstructure network model of the U.S. refining and chemical manufacturing industry, and formulated a multi-objective linear program with two conflicting objectives, total industry production cost and net industry carbon loss. The model was first solved for each objective separately, and then the weighted sum and ε-constraint methods were employed to obtain the Pareto-optimal set of solutions. Moving along the Pareto front towards the minimum carbon loss point causes a number of structural changes in the industry network, notably a shift towards natural-gas-derived methanol used for production of olefins. Results also suggest that there may be environmental benefits to prioritizing efficient raw material usage, even if that means more usage of energy utilities, especially as these utilities can be and will likely become increasingly decarbonized.

Acknowledgement

This paper is based on work supported primarily by the National Science Foundation under Cooperative Agreement No. EEC-1647722 (CISTAR: NSF Engineering Research Center for Innovative and Strategic Transformation of Alkane Resources). Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

D. Chang and D. T. Allen, 1997, Minimizing Chlorine Use: Assessing the Trade-offs between Cost and Chlorine Reduction in Chemical Manufacturing, Journal of Industry Ecology, 1, 111–134

S. E. DeRosa; D. T. Allen, 2016, Impact of New Manufacturing Technologies on the Petrochemical Industry in the United States: A Methane-to-Aromatics Case Study, Industrial & Engineering Chemistry Research, 55, 18, 5366-5372

S. Fathi-Afshar and J. C. Yang, 1985, Designing the Optimal Structure of the Petrochemical Industry for Minimum Cost and Least Gross Toxicity of Chemical Production, Chemical Engineering Science, 40, 5, 781–797

IHS Markit, 2012, IHS 2012 Process Economics Program Yearbook, https://ihsmarkit.com/products/chemical-technology-pep-index.html

J.J. Siirola, 2014, The Impact of Shale Gas in the Chemical Industry, AIChE Journal, 60, 3, 810-819

A. Skouteris, I. Giannikopoulos, T. F. Edgar, M. Baldea, D. T. Allen, and M. A. Stadtherr, 2021, Systems Analysis of Natural Gas Liquid Resources for Chemical Manufacturing: Strategic Utilization of Ethane, Industrial & Engineering Chemistry Research, 60, 33, 12377-12389

M. A. Stadtherr and D. F. Rudd, 1976, Systems Study of the Petrochemical Industry, Chemical Engineering Science, 31, 1019-1028

U.S. Energy Information Administration, 2021a, Natural Gas Explained: Where Our Natural Gas Comes From. [Online]. https://www.eia.gov/energyexplained/natural-gas/where-our-natural-gas-comes-from.php

U.S. Energy Information Administration, 2021, Units and Calculators Explained: British Thermal Units (Btu). [Online]. https://www.eia.gov/energyexplained/units-and-calculators/british-thermal-units.php

Proceedings of the14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50092-0

Mapping Anthropogenic Carbon Mobilization through Chemical Process and Manufacturing Industries

Amrita Sen[a], George Stephanopoulos[b,c], Bhavik R Bakshi[a*]

[a]*William G Lowrie Department of Chemical Engineering, The Ohio State University, Columbus Ohio 43210, USA*
[b]*The Global KAITEKI Center, Arizona State University, Tempe, Arizona 85287, USA*
[c]*Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

bakshi.2@osu.edu

Abstract

The long-term impact of global warming and the resulting climate crisis, brought about by human-induced emission of greenhouse gases, is an imminent environmental concern. The Paris Agreement aims to limit global temperature rise to below 2° C over pre-industrial levels, to curb this impact. Meeting this limit necessitates reaching carbon neutrality by 2050, which imply no net transport of carbon dioxide to the atmosphere. The chemical process industry along with associated manufacturing industries such as iron and steel, cement and aluminum contributes significantly towards global carbon dioxide emissions. Mapping the precise routes of Carbon mobilization is the first step towards establishment of a sustainable, circular and Carbon neutral chemical industry. There exist no C flow models for aforementioned energy intensive industries. Current published literature also does not account for C mobilized to meet the energetic needs of global chemical processes. They also do not account for the emissions offset by material exchange between different production processes. In this work, we develop a steady state model of Carbon flow through chemical process and associated industries. Our model traces the flow of carbon from fossil feedstock, to energy carriers and chemical intermediates, and finally valuable products, by-products and emissions. This model makes use of process data, life-cycle inventories models developed by existing studies on the chemical and petrochemical industries, government databases, greenhouse gas emissions data and economy models . Fundamental laws like mass and energy balance are used in conjunction with stoichiometric calculations to estimate missing data and reconcile incorrect data. We represent this model as a Sankey Diagram to better facilitate visualization of the process network and identify scope of process improvement. We elaborate how this model helps the placement of process alternatives such as use of renewables, electrification, green hydrogen and carbon capture and storage in the value chain. These alternatives can be highly energy intensive, requiring a large amount of "net zero" electricity to function. The dependence of renewably sourced electricity on land area availability necessitates its efficient use. Thus, the integration of fossil alternatives in the model paves the path for their targeted and optimal usage towards decarbonization.

Keywords: Decarbonization, Modelling, Supply chain, Sustainability

1. Introduction

As global average temperature continues to rise and predictions for climate change turn more grim with every new assessment, closing the global anthropogenic carbon (referred to as "C" from here on) cycle has become more important than ever. Reducing the production of materials for consumption, infrastructure and healthcare in the face of a rising global population, or bringing about drastic reduction in consumptive behaviour overnight, is unlikely. Therefore, focussing on emissions reduction while still maintaining production volumes (or establishing a circular economy of materials) may arguably have better payoffs in the immediate future.

The efficiency (yield or selectivity) of conventional chemical processes cannot be increased indefinitely. Thus, there is a minimum C emission associated with all products. Any further reduction requires us to choose alternate pathways or retrofit mitigation technologies to conventional pathways. Many innovative technologies have been

developed to leverage both these options. [8, 7, 1, 10] The general idea behind carbon neutral technologies is to limit CO_2 emissions to the atmosphere. With prices of renewable electricity dropping steadily, electrically powered emerging technologies show promise in competing with incumbent routes of production. Examples of such processes include electrochemical means of converting CO_2, that would otherwise be emitted to the atmosphere,to value added chemicals, or electrification of H_2 production. While renewables and emerging technologies may lower emissions to their credit, their usage does not absolve industries of all environmental concerns. C capture itself might increase energy needs of a process enough, to offset its C credits. Renewables also raise issues of waste management of noble materials, land use concerns. Thus, the application of these technologies needs to be weighed in with its tradeoffs to guide policymaking.

Attempts at large scale, sector-wide decarbonization imply the implementation of alternate low carbon pathways wherever possible, and C capture from exhaust gases elsewhere. This naturally requires a thorough knowledge of C flows, sources and sinks. A superficial knowledge of emissions from the chemical industry is not very useful in this regard. The contribution of different processes, pathways and fuels is needed to target processes with the highest decarbonization potential. The US EPA traces a majority of US greenhouse gas (GHG) emissions back to the transportation sector, followed closely by electricity generation and industrial sectors. Many existing works break down these emissions across manufacturing and process industries. These diagrams however, suffer from various inadequacies of their own. The diagram developed by Lawrence Livermore National Lab uses data from the US Energy Information Administration (EIA) and does not distinguish between different processes in the industrial sector. [3] The diagram by Global Climate and Energy Project (GCEP) at Stanford, while much more detailed still lumps all chemicals into one node and does not provide any insight into the individual consumption or emission of processes.[9, 2] The mass flow balance on the process network in the chemical industry developed by Levi et al. while comprehensive, does not have information on the energetic needs of these processes. Finally, there is a distinct need for and lack of distinction between the direct and indirect C requirements of a process. The decarbonization potential of a process changes significantly depending on whether C is required as feedstock for the process, or simply for its energy or heating needs. This insight is valuable and missing from the current literature.

The current major scientific efforts in this field focus on development of decarbonization technologies. However, this reductionist approach may have rebound effects, whereby increase in consumptive behaviour offsets the marginal benefits of emissions reduction, and may hinder the longer scale goals of establishing a circular, sustainable chemical industry which is still carbon neutral. For example, attempts at electrification of ammonia, source hydrogen from electrolytic processes and attempt to electrify the operation of Haber Bosch process, to avoid releasing carbon dioxide as a co-product of fossil sourced hydrogen. A superficial analysis may reveal the abatement of a large fraction of greenhouse gad emissions when such production routes are taken. However, as we see in the results of our work, other processes like methanol, urea and acetic acid production are dependent on this carbon monoxide for their feedstock. Decarbonizing these sectors thus becomes much harder. This insight is easy to miss in traditional reductionist thinking and may lead us to grossly overestimating the emissions reduction. This incites the need of development of more holistic models which will accurately reflect the dependence of different processes in the network and avoid chances of shifting impacts and inadvertently increasing the environmental burden. In this work, we build a model that captures the co-dependence of different processes and outline a protocol for resolving the feedstock and utility C needs of the chemicals and material industries, while distinguishing between different processes and pathways of production.

2. Methodology

This model traces most conventional processes centering the chemical industry. We outline the C flows in the feedstock, product as well as that associated with energy requirement of the process. For all cases, we begin with a process description and a stoichiometric model of the ideal process. Yield and selectivity data collected from published surveys of operational plants lets us calculate realistic values of feedstock. Energy data is estimated

from specific energy consumption (SEC) values or from life-cycle data. C content of these material and energy flows is then determined. Finally, we put these intensive flows into global perspective using production tonnage of each material. [11, 12] These calculations are demonstrated below for methyl alcohol.

Synthesis	$CO + 2H_2 \rightarrow CH_3OH$	(1)
Coal/Oil POX	$CH_n + H_2O \rightarrow CO + \frac{n+2}{2}H_2$	(2)
NG SMR	$2CH_n + O_2 \rightarrow 2CO + nH_2$	(3)
WGS	$CO + H_2O \rightarrow H_2 + CO_2$	(4)
RWGS	$H_2 + CO_2 \rightarrow CO + H_2O$	(5)
SMR to methanol	$2CH_n + \frac{n-2}{3}CO_2 + \frac{8-n}{3}H_2O \rightarrow \frac{4+n}{3}CH_3OH$	(6)
POX to methanol	$2CH_n + O_2 + \frac{4-n}{3}H_2O \rightarrow \frac{2+n}{3}CH_3OH + \frac{4-n}{3}CO_2$	(7)

To estimate feedstock requirement of methanol production, we use chemical synthesis route as shown in Eq.(1). The feedstock for methanol synthesis are sourced from syngas. Syngas can be generated by steam methane reforming of natural gas or coal gasification or partial oxidation of oil, as shown in Eq.(2) and Eq.(3) respectively. The kind of fossil feedstock used, determines the ratio of CO and H_2 in syngas. This ratio can be corrected by water gas shift or reverse water gas shift reactions for direct use in the synthesis process, depending on which gas is in excess. This is illustrated in Eq.(4) and Eq.(5). The value of 'n' can be approximated as 0.456 for coal, 1.873 for oil and 3.951 for natural gas. Thus, while syngas produced from coal and oil have excess CO and need to be subjected to water gas shift, NG sourced syngas is lean in CO and is followed up with reverse water gas shift. The final equations for methanol production, combining syngas generation, WGS/RWGS and methanol synthesis can be represented as Eq.(6) for NG and Eq.(7) for coal/oil.[6]

As can be observed from Eq.(6) and (7), the SMR/POX reactions coupled with WGS/RWGS reactions have two sources of C flows. The first is associated with the fossil feedstock requirement and the second is process emissions of CO_2. The coal/oil POX process emits CO_2 which is released to the atmosphere. On the other hand, the NG SMR process consumes CO_2 which we assume is sourced from ammonia plants nearby. Ammonia plants use only H_2 from syngas mixture. Thus, CO, which is generated as co-product of this H_2, can be separated and oxidized, according to Eq. (4), to provide feedstock to the methanol plants.

We assume an efficiency of 0.99 for the synthesis route (η_s) and 0.861 (NG), 0.808 (oil) and 0.76 (coal) for syngas generation steps (η_c). [6] Thus, the yield data along with stoichiometric information lets us calculate feedstock demand and C input thereof for methanol generation. [6] These calculations for process C of methanol are elucidated in Table 2 .

For energy requirement, we consider specific energy consumption (SEC) data for different feedstocks. Methanol generated from NG SMR has an SEC of 24 GJ/ton whereas partial oxidation of coal or oil lead to an energy consumption of 13.9 GJ/ton methanol. [5] To estimate the emissions associated with energy use, emisson coefficients associated with each fuel are used. The final calculations are shown in Table 3.

	Feedstock $\frac{CH_n}{CH_3OH}$	CO_2 input $\frac{CO_2}{CH_3OH}$	η_c	η_s	n	Actual feed $\frac{kg}{kgCH4O}$	Actual CO2 $\frac{kg}{kgCH4O}$
Coal	$\frac{6}{2+n}$	$-\frac{4-n}{2+n}$	0.76	0.99	0.456	1.26	-2.004
Oil	$\frac{6}{2+n}$	$-\frac{4-n}{2+n}$	0.808	0.99	1.873	0.84	-0.763
Natural Gas	$\frac{6}{4+n}$	$\frac{n-2}{4+n}$	0.861	0.99	3.951	0.44	0.396

	SEC for feedstock use $\frac{GJ}{ton\ CH_3OH}$	Feedstock fraction %	Emissions for energy use $\frac{kg\ CO2}{MJ}$	Energy fraction %
Coal	24	20.5	0.091	50
NG	13.9	71	0.013	50

Finally, these intensive product flow calculations are scaled up for global production tonnages. Currently, the annual production capacity of methanol stands at 102 million metric tonnes. [4] Of this, we estimate 71% is sourced from natural gas derived, 8.5% from oil derived and the remaining from coal derived syngas. [6] The flows are scaled accordingly and thus we arrive at the total C flows associated with fulfilling the global demand of methanol. The treatment of process C follows a protocol illustrated by Levi et al. [6]

3. Results and Discussion

Fig 1 shows the mobilization of C in feedstocks for manufacture of major chemicals. All flows shown here correspond to mass units of C. Their relative values are scaled by their annual global production capacities i.e. the flows are extensive in nature. Therefore, in this case, the production tonnage and specific C consumption both dictate magnitude of the flows. Process flows resulting in output of C flows are shown to be released to the environment. Unreacted reactants or leakages are grouped together in a separate category as "loss".

The diagram starts with fossil resources like crude oil, natural gas and coal. Refining processes yield primary hydrocarbons like olefins, aromatics, cyclic compounds and alkanes. Transformative reactions alter the relative production of these compounds. For example, toluene hydrodealkylation (THD) and toluene disproportionation (TDP) convert toluene to xylenes and benzene, ethylene and butene get converted to propylene via metathesis, and propane on dehydrogenation produces propylene. Alongside, we also show the synthesis of urea and methanol. The next stages show the production of platform chemicals like cumene, acrylonitrile, caprolactam, phenol, ethylene glycol, terephthalic acid, vinyl chloride monomer, acetic acid, formaldehyde etc from hydrocarbons. This is followed by production of polymers like polyethylene, polypropylene, polyvinyl chloride, polyethylene terephthalate, polystyrene etc. Finally, we track the end uses of some of these polymers to their major products. The diagram shows a large amount of C embodied in products used by the packaging industry, consumer goods and so on.

Large volume polymers like polyethylene, polypropylene, PET sequester C through their long half lives whereas C embodied in solvents such as acetone, ethylene glycol, toluene are either incinerated or disposed off. While attempts to circularize use of such stable plastics is underway, efforts can be made to source their feedstock C from non-fossil sources. While captured C or biogenic C can replace their fossil counterpart for use as feedstock, substituting the hydrogen or energy requirement of such processes presents a formidable task. Since captured C is only available as CO_2, it does not have any value as fuel. While the diagram itself only represents C flows, co-product flows are captured in process models. This information is indispensable in searching for lower C pathways of production.

We also observe a significant amount of process loss, throughout the industry. This can be attributed to unconverted reactants or inefficient separation of products. When the consumption of process and energy C are compared, we see emissions distributed similarly across both categories. This points to the vast decarbonization potential of the chemicals industry whereby energy can be alternately sourced from non emitting resoources.

4. Conclusions

Visualization of the C Flow model points us to the industries with greatest C footprint and the best ways to retrofit decarbonization technologies to these specific cases. This model is a stepping stone towards possibly answering bigger questions about the deep decarbonization of the chemicals industry. The data used in this model, can be used in conjunction with data on emerging technologies to optimize the economics and electricity demand of a decarbonized chemical industry. Additionally, the need for innovation and the direction in which it is needed may also be investigated. Eventually, the need for a policy change to support and affect the shift to decarbonized technologies can be supported with models such as this one. Thus, this model is a foundation in the vision of a sustainable, circular and C neutral chemical industry.

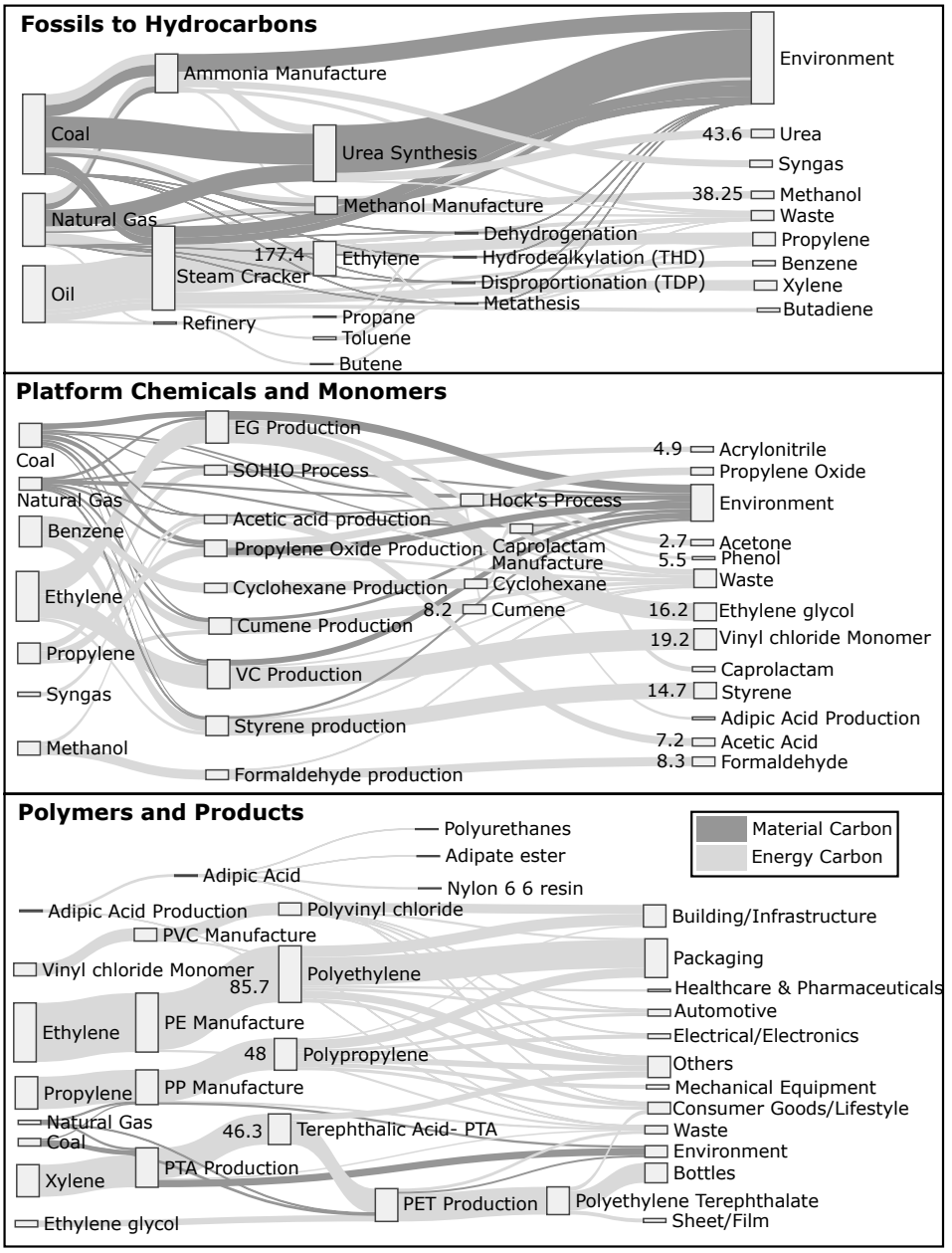


Figure 1: Flow of Fossil C through the Chemical Industry for in 2018 (in million metric tonnes of C)

5. Acknowledgments

Partial financial support was provided by "The Global KAITEKI Center" at Arizona State University (ASU), a university-industry partnership between ASU and The KAITEKI Institute of Mitsubishi Chemical Holdings Corporation.

References

[1] Mai Bui, Claire S. Adjiman, André Bardow, Edward J. Anthony, Andy Boston, Solomon Brown, Paul S. Fennell, Sabine Fuss, Amparo Galindo, Leigh A. Hackett, Jason P. Hallett, Howard J. Herzog, George Jackson, Jasmin Kemper, Samuel Krevor, Geoffrey C. Maitland, Michael Matuszewski, Ian S. Metcalfe, Camille Petit, Graeme Puxty, Jeffrey Reimer, David M. Reiner, Edward S. Rubin, Stuart A. Scott, Nilay Shah, Berend Smit, J. P. Martin Trusler, Paul Webley, Jennifer Wilcox, and Niall Mac Dowell. Carbon capture and storage (CCS): the way forward. *Energy & Environmental Science*, 11(5):1062–1176, 2018.

[2] W HERMANN. Quantifying global exergy resources. *Energy*, 31(12):1685–1702, sep 2006.

[3] https://www.llnl.gov/. Lawrence Livermore National Laboratory: C Emissions and Energy Flow Charts.

[4] https://www.methanol.org. Methanol Institute.

[5] IEA. Energy and GHG Reductions in the Chemical Industry via Catalytic Processes. 2013.

[6] Peter G. Levi and Jonathan M. Cullen. Mapping Global Flows of Chemicals: From Fossil Fuel Feedstocks to Chemical Products. *Environmental Science and Technology*, 52(4):1725–1734, feb 2018.

[7] Hélène Pilorgé, Noah McQueen, Daniel Maynard, Peter Psarras, Jiajun He, Tecle Rufael, and Jennifer Wilcox. Cost Analysis of Carbon Capture and Sequestration of Process Emissions from the U.S. Industrial Sector. *Environmental Science & Technology*, 54(12):7524–7532, jun 2020.

[8] Giulia Realmonte, Laurent Drouet, Ajay Gambhir, James Glynn, Adam Hawkes, Alexandre C. Köberle, and Massimo Tavoni. An inter-model assessment of the role of direct air capture in deep mitigation pathways. *Nature Communications*, 10(1):3277, dec 2019.

[9] Richard E Sassoon, Weston A Hermann, I-Chun Hsiao, Ljuba Milkovic, Aaron J Simon, and Sally M Benson. Quantifying the Flow of Exergy and Carbon through the Natural and Human Systems. *MRS Proceedings*, 1170:1170–R01–03, jan 2009.

[10] Nils Thonemann and Massimo Pizzol. Consequential life cycle assessment of carbon capture and utilization technologies within the chemical industry. *Energy & Environmental Science*, 12(7):2253–2263, 2019.

[11] www.ihsmarkit.com. IHS Markit Production Market Analysis.

[12] www.statista.com. Statista: Global Production Capacity Forecasts.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50093-2

Optimal Designing and Planning of Ethiopia's Biomass-to-Biofuel Supply Chain Considering Economic and Environmental Dimensions under Strategic and Tactical Levels

Brook Tesfamichael[a,b,c], Ludovic Montastruc[a,b,*], Stéphane Negny[a,b], Abubeker Yimam[c]

[a]*Université de Toulouse, INP-ENSIACET, 4, allée Emile Monso, F-31432 Toulouse Cedex 04, France*
[b]*CNRS, LGC (Laboratoire de Génie Chimique), F-31432 Toulouse Cedex 04, France*
[c]*School of Chemical and Bio Engineering, Addis Ababa Institut of Technology, Addis Ababa University, Addis Ababa, Ethiopia*
**ludovic.montastruc@ensiacet.fr*

Abstract

This contribution provides an optimization-based decision support model of biomass to biofuels supply chain (BBSC) in order to offer optimal strategic decisions and tactical plans in the entire supply chain. This model is designed for long-term planning studies, in that it is utilized to optimize the Ethiopian BBSC over a 20-year horizon. Moreover, a comprehensive LCA is conducted on the BBSC by broadening the assessed impact categories adverse to most of the previous economic-environment models, which considers only one type of impact. To this end, through the LCA method and ecocost approach, an effort is made here to evaluate the environmental impacts associated with different biomass preprocessing technologies (drying and size reduction, mechanical and solvent extractions), biorefinery technologies (biochemical and thermochemical conversions, and homogeneous and heterogeneous transesterification) and material transportations along the BBSC. Since this planning problem relies on two objective functions, namely, profit and ecocost, a set of optimal solutions are generated to come up with the best compromise solution between the two antagonistic objectives.

Keywords: Biomass-to-biofuel supply chain (BBSC), economic-environment optimization, life cycle assessment, long-term planning.

1. Introduction

Over the last decade, biofuels are of rapidly growing interest in Ethiopia for reasons of saving foreign currency, export earning, job creation and greenhouse gas mitigation. However, the implementation of biomass-to-biofuel projects is in its infancy despite there is abundant biomass availability in the country [1]. Moreover, the existing infant biomass-to-biofuel supply chain (BBSC) is not carried out in a sustainable manner. This is majorly associated with the numerous economic and environmental challenges along the supply chain. Therefore, policymakers and other actors in the biofuel sector require a framework, which supports them to make scientifically valid and sustainable strategic decisions and tactical planning.

These days, numerous process systems engineering tools have been developed to assist decision making in the design and plan of various supply chains, of which optimization-based mathematical models take the largest share [2]. Based on the model outputs, recommendations had been forwarded for political decision-makers as well as for potential investors. Several strategic-tactical level models are formulated with objectives of maximizing economic and environmental benefits of supply chain in general and BBSC in particular. However, the models developed so far are not comprehensive and realistic enough to be applied at national level for long term planning of the biofuel sector. Some of the drawbacks include addressing the BBSC in a partial way (focusing on either upstream or midstream or downstream activities), dealing only on one type of biofuel product in a single supply chain and focusing on a one-year planning period. Moreover, the environmental concern in most of the previous studies emphasis on global warming, greenhouse gas (GHG) emissions and fossil energy consumption [3]. Nevertheless, these are not the only environmental impacts generated from the lifecycle of biofuel supply chain, which results in the importance of broadening the impact categories considered while dealing BBSC.

Henceforth, to address the limitations stated above, this work provides a comprehensive optimization-based decision support model in order to design and plan both bioethanol and biodiesel supply chain over a long-term. The objective of the model relies on both economic and environmental aspects, in that it aims to maximize the profit and minimize the ecocost of the BBSC. The model is applied to the real case of Ethiopia to offer optimal strategic and tactical decisions along the BBSC over 20-years horizon. Moreover, this study tries to make a comprehensive LCA along the supply chain of corn stover-, molasses-, and bagasse-based bioethanol and jatropha-, and castor-based biodiesel by broadening the assessed impacts. The impacts considered in this study are broadly classified into four; namely, carbon footprint (global warming potential), ecosystem (acidification, eutrophication and fresh water aquatic ecotoxicity potentials), human health (fine dust, human toxicity and photochemical oxidants potentials) and resource scarcity (metal and water scarcity, fossil fuel depletion and waste generation potentials).

2. Methodologies

2.1. Life cycle environmental assessment

The LCA method, which consists of goal and scope definition, life cycle inventory, and impact assessment, is considered in this work to analyze the BBSC impact on the environment. The LCA is applied herein is from biomass supplier (farm) gate to biofuel market. The BBSC stages considered in this work include biomass feedstocks transportation and preprocessing, preprocessed biomasses transportation, biofuels production and transportation. The technologies considered to preprocess bioethanol feedstocks is drying and size reduction, whereas mechanical and solvent extraction technologies are considered to preprocess the biodiesel feedstocks. Furthermore, the two biorefinery technologies considered for bioethanol production are biochemical and thermochemical conversions. Besides, homogeneous and heterogeneous base-catalyzed transesterification technologies are considered for biodiesel production. One ton of raw or prepocessed biomass is taken as the functional unit for this study. The life cycle inventory, which quantifies the amount of raw materials and product as well as energy requirements, of each biomass preprocessing and biofuel production (biorefinery) technology are estimated based on previous works. The GHG and other criteria pollutant emissions for the different technologies by taking into account each type of biomass and

preprocessed biomass are obtained from the GREET model. On the other hand, previous researches are used to estimate the GHG and other criteria pollutants emissions from vehicles used in transporting materials along the supply chain, i.e., Euro II trailer truck and Euro III tanker truck for transportation of solid and liquid materials respectively. Then, based on the emission, inventory and utility consumption data, the impacts of all activities of the BBSC on the environment is analyzed through the ecocost method. This method is developed based on the marginal prevention costs needed to control the negative impact of toxic emissions. The total ecocost of a product or activity is the sum of ecocosts natural resources depletion, ecosystem, human health and carbon footprint during its life cycle.

2.2. Model formulation

The intention of this study is to provide a comprehensive mathematical model of BBSC, which includes all the principal supply-chain components upstream and downstream of the biorefineries. This model is designed for long-term planning studies, in that it is utilized to optimize the Ethiopian BBSC over a 20-year horizon. To account the whole BBSC behavior, the problem is designed as a spatially explicit, multi-product, multi-feedstock, multi-period, and multi-echelon MILP modeling framework. The model considers yearly and monthly time periods. The design mechanism is perceived as a multi-objective optimization problem that intends to maximize the profit and minimize the ecocost of the BBSC. The profit as illustrated in equation (1) is calculated by taking into account the cash inflow (total annual revenue) and cash outflows (total annual cost and investment cost) in a specific year. Moreover, the total ecocost is given as the sum of the ecocosts of biomass preprocessing, biofuel production and material transportation, as indicated in equation (2).

$$Profit = \sum_{n=1}^{Y} \left(\left[\frac{1}{(1 + ir)^{y}} \left(Total\ Revenue_{y} - Total\ Cost_{y} \right) \right] \right) \tag{1}$$

Where ir is the discount rate

$$Total\ EcoCost = \sum_{y} \left[Ecocost\ Preprocessing_{y} + Ecocost\ Production_{y} + Ecocost\ Transportation_{y} \right] \tag{2}$$

Moreover, constraints and mass balances that needs to be fulfilled at each stage of the BBSC are formulated, including: supply and demand satisfaction, inventory balance, production amount, binary and non-negativity decision variables, and storage and weight capacity constraints. The expected output of the model refers to strategic decisions in the BBSC, which includes the network configuration, capacity, technology, and location of the biorefineries and preprocessing units as well as the capacity and location of distribution centers. Moreover, optimum decisions related to annual production, inventory, and transportation of materials along the BBSC are expected results of the model. The model is described on [4].

2.3. Case Study

The developed model in this study is applied to the real case of Ethiopia. All the assumptions taken and detail technical and economic data of the case study are available on [4]. These data include biomass feedstocks availability and purchasing price, biofuels

and coproduct demand and selling price, investment and production costs as well as storage and processing capacities of each BBSC components, conversion factors of each candidate technology, transportation distance between different zones of the country and transportation cost of materials in the supply chain.

2.4. Solution method for multi-objective optimization

The MILP model was solved using the ILOG CPLEX solver. The antagonistic nature between the economic and environmental objectives was solved using the following procedure. First, the profit was maximized to obtain the resulting optimal profit (upper bound of profit) and the ecocost value (upper bound of ecocost). Second, the ecocost was minimized to obtain the resulting profit value (lower bound of profit) and the optimal ecocost (lower bound of ecocost). Then, the profit was maximized 10 times by constraining the ecocost in different ranges between the lower and upper bound. Similarly, the ecocost was minimized 10 times by constraining the profit in different ranges between the lower and upper bound. To this end, a set of efficient solutions ware generated.

3. Result and Discussion

3.1. Environmental impact

Based on the GREET outputs, utility consumption data and inventory analysis, the environmental impacts (ecocost) associated with biomass preprocessing, biofuel production and material transportation are calculated. The ecocosts associated with biomass preprocessing via drying and size reduction, mechanical and solvent extraction are depicted in Figure 1. Similarly, Figure 2 presents the ecocosts of bioethanol and biodiesel production via biochemical or thermochemical conversion, and homogenous or heterogeneous transesterification.

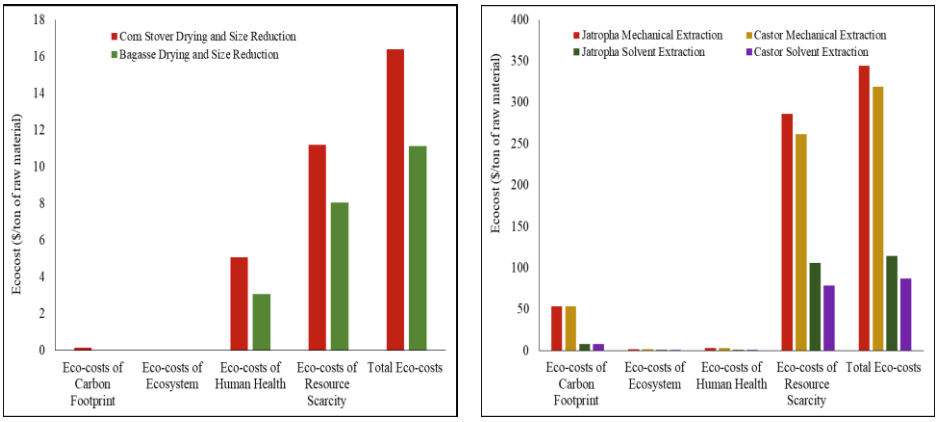


Figure 1: Ecocosts of carbon footprint, ecosystem, human health and resource scarcity for preprocessing corn stover and bagasse using drying and size reduction (left) and for preprocessing jatropha and castor using mechanical and solvent extraction (right)

In the drying and size reduction unit, ecocosts of human health and resource scarcity are significant among others as there is considerable fine dust emission during size reduction of the feedstocks and the requirement of energy by the hammer mill. Due to the higher energy demand in mechanical and solvent extraction, the resource scarcity ecocost is significant in these two preprocessing technologies compared with other ecocost types.

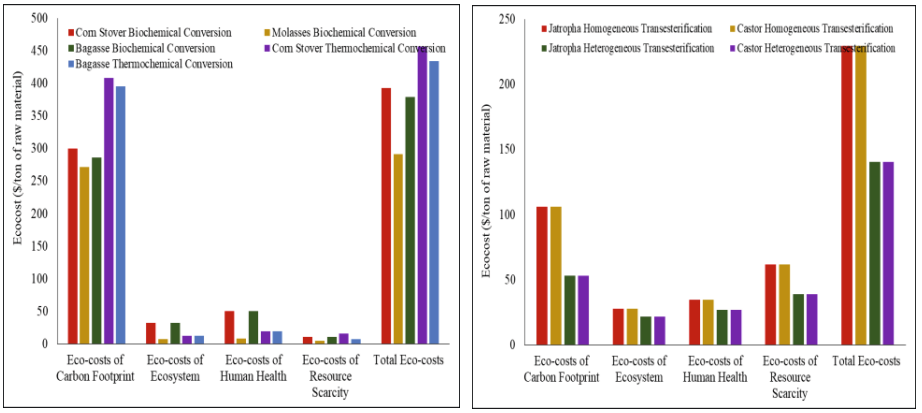


Figure 2: Ecocosts of carbon footprint, ecosystem, human health and resource scarcity for corn stover-, molasses- and bagasse-based bioethanol production using biochemical and thermochemical conversion (left) and jatropha oil-, and castor oil-based biodiesel production using homogeneous and heterogeneous base-catalyzed transesterification (right)

Biochemical and thermochemical conversions are the highest ecocost contributor compared with the other activities due to the highest GHG, specifically CO_2, emission generated from the two processes, in which the latter technology contributes more. Similar to the bioethanol technologies, the carbon footprint ecocost of biodiesel producing technologies (homogeneous and heterogeneous transesterification) are the most significant ones because of their higher energy requirement that is fulfilled by fossil fuel combustion, which is responsible for emission of CO_2 and other criteria pollutants.

The significant amount of CO_2, NOx, CO and VOC emissions from the vehicles, especially in the Euro II trailer truck, has resulted in a considerable contribution of ecocosts of human health, ecosystem and carbon footprint on the total ecocost.

3.2. Multi-objective optimization results

Each run of the MILP model had 1,348,768 constraints and 2,404,604 continuous decision variables of which 174 were binary variables. Optimal solutions were found between 58.42-102.42 minutes on an Intel 2.60-GHz processor. The set of solutions generated to show the tradeoff between the two objectives of the BBSC are depicted in Figure 3.

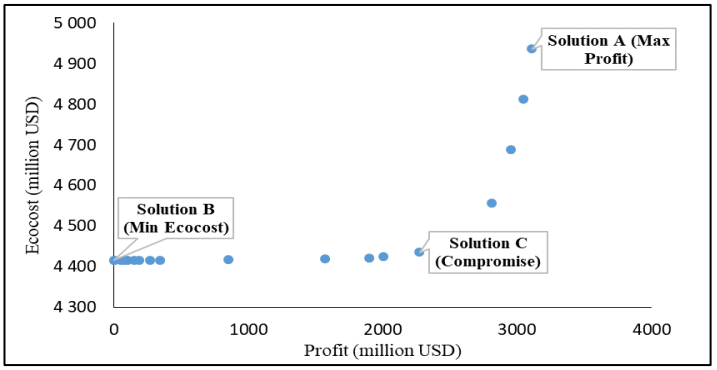


Figure 3: Set of solutions for profit maximization vs ecocost minimization

3.3. Optimum strategic decisions

The optimum solution (Solution C of Figure 3) promotes molasses and jatropha as the only biomass feedstock to address the bioethanol and biodiesel demand of Ethiopia in the next 10 and 5 years respectively, which is then joined by the other biomass feedstocks. Moreover, the optimal solution does not promote any drying and size reduction unit for the first 10 years and solvent extraction technology is mainly preferred to preprocess the biodiesel feedstocks in the entire planning period. Most of the preprocessing units are located proximate to the feedstock supply centers although not necessarily in the same zones. Regarding the biorefinery, the installation of biochemical refineries is proposed to attain the required amount of bioethanol production in the next 10 years, which is joined by thermochemical refineries after a decade. The locations proposed to build the biochemical refineries are in zones where sugar industries are available and the capacities are driven by the molasses availability. Besides, the model prescribes heterogeneous-transesterification refineries installation throughout the planning period for producing biodiesel.

4. Conclusions

To provide optimal strategic decisions and tactical plans of biomass to biofuels supply chain (BBSC), this study aims to develop an optimization-based decision support model. It is a long-term planning model with objectives of maximizing both economic and environmental benefits. The model is utilized to optimize the Ethiopian BBSC over a 20-year horizon. Unlike to previous researches, a number of environmental impact categories are considered to assess the impact of the BBSC on the environment. The addressed impact categories include carbon footprint, human health, ecosystem and resource scarcity potential of the BBSC. To this end, all these impacts are evaluated on each activities of the supply chain (biomass preprocessing, biofuel production and material transportation) by taking the type of biomasses and technologies into consideration. The results have indicated that biofuel production via biochemical or thermochemical conversions is the highest environmental impact contributor in case of bioethanol supply chain. Contrary, in supply chain of biodiesel, the highest impact is generated from biomass preprocessing using mechanical or solvent extraction. Since this planning problem relies on two objective functions, namely, profit and ecocost, a set of optimal solutions are generated to come up with the best compromise solution between the two antagonistic objectives.

References

[1] M. Berhanu, S.A. Jabasingh, Z. Kifile, Expanding sustenance in Ethiopia based on renewable energy resources – A comprehensive review, Renew. Sustain. Energy Rev. 75 (2017) 1035–1045. https://doi.org/10.1016/j.rser.2016.11.082.

[2] A.P. Barbosa-Póvoa, C. da Silva, A. Carvalho, Opportunities and challenges in sustainable supply chain: An operations research perspective, Eur. J. Oper. Res. 268 (2018) 399–431. https://doi.org/10.1016/j.ejor.2017.10.036.

[3] S. Soam, R. Kumar, R.P. Gupta, P.K. Sharma, D.K. Tuli, B. Das, Life cycle assessment of fuel ethanol from sugarcane molasses innorthern and western India and its impact on Indian biofuel programme, Energy. 83 (2015) 307–315. https://doi.org/10.1016/j.energy.2015.02.025.

[4] B. Tesfamichael, L. Montastruc, S. Negny, A. Yimam, Designing and planning of Ethiopia ' s biomass-to-biofuel supply chain through integrated strategic-tactical optimization model considering economic dimension, Comput. Chem. Eng. 153 (2021) 107425. https://doi.org/10.1016/j.compchemeng.2021.107425.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50094-4

A Novel Integrated Optimal Scheduling Framework for Holistic Refinery Supply Chain Management

Li Yu, Qiang Xu*

Dan F. Smith Department of Chemical & Biomolecular Engineering, Lamar University, Beaumont, Texas 77710, USA
Qiang.xu@lamar.edu

Abstract

The refinery supply chain management is critically important. It covers three highly correlated sub problems: the front-end crude-oil management sub problem (CM), the refinery manufacturing sub problem (RM), and the multi oil-product pipeline distribution sub problem (MOPD). By coordinating the management and operations of the three sub problems, it can greatly minimize the operating cost of the entire supply chain. In this paper, a continuous-time and continuous-volume based general integrated optimal MINLP scheduling framework for holistic refinery supply chain covering the crude-oil management, the refinery manufacturing, and the multi oil-product pipeline distribution has been developed (CM&RM&MOPD). The objective is to minimize the total operating cost subject to various constraints such as operating rules, product specifications, inventory limits, delivery constraints, and oil-product demands at each oil depot. The efficacy of the developed CM&RM&MOPD model has been demonstrated by a large-scale case study.

Keywords: Integrated scheduling, MINLP, Holistic refinery supply chain, Crude-oil unloading and transferring, Refinery manufacturing, Pipeline distribution.

1. Introduction

The refinery supply chain management is critically important to the oil industry. It contains three subsystems as shown in Figure 1, starting from the crude-oil vessels unloading at ports to the oil-product exported to local consumer markets. Specifically, the first subsystem includes crude-oil unloading from vessels to storage tanks, transferring crude-oil to charging tanks, blending, and feeding to crude distillation units (CDUs); the second subsystem involves the refinery manufacturing process, including major refinery processing units such as crude distillation, catalytic reforming, fluid catalytic cracking, hydrocracking, delayed coking, hydrotreating, gas fraction, alkylation, hydrogen pooling, blending, as well as sulfur recovery facilities; the third subsystem is about refinery oil-product distribution, which consists of inventory management at the refinery product tank farm, multi oil-product pipeline transportation, as well as oil-product receiving, exporting, and inventory management at different depots. The three subsystems are highly correlated. According to Chima (2007), each subsystem should respond quickly to the exact demand of its downstream customers, protecting itself from problems with suppliers, and buffering its operations from both demand and supply uncertainty. Thus, by coordinating the management and operations of the three subsystems can greatly maximize the potential benefit margin of the entire refinery supply-chain. There exists a lot of works explicitly and deeply exploited and studied these subsystems separately, e.g., some for the front-end crude-oil management (Zhang and Xu, 2015; Qu et al., 2019),

some for oil-product pipeline distribution (Cafaro and Cerdá, 2010; Yu et al., 2020), and some works have simultaneously considered the first two subsystems (Yang et al., 2020; Xu et al., 2017). However, few studies addressed these three sub systems simultaneously. Guyonnet et al. (2009) studied the simplified crude oil unloading, production planning, and distribution sub models on a tactical decision level by solving each part in a sequential push or pull manner, where the planning horizon is discretized into a day or a week. Generally, systematic studies for the integrated scheduling the holistic refinery supply chain are still lacking.

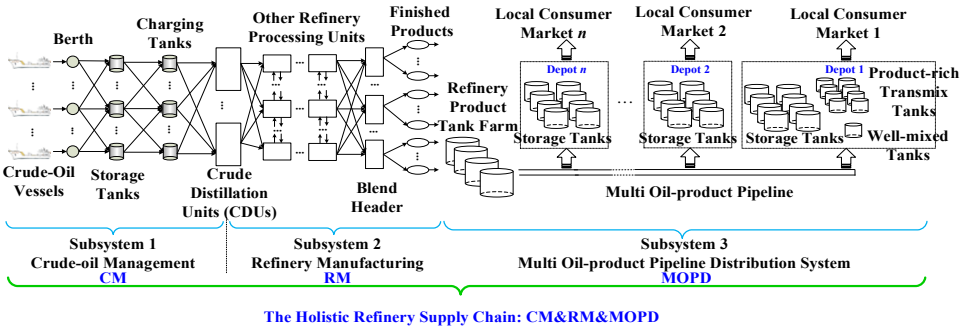


Figure 1. The studied scope of this study.

In this paper, a general integrated optimal scheduling model for the crude-oil management, the refinery manufacturing, and the multi oil-product pipeline distribution has been developed (CM&RM&MOPD). It consists of five sub-models: (i) a crude-oil management (CM) sub-model including the crude-oil unlading, transferring, and charging CDUs; (ii) a refinery manufacturing (RM) sub-model covering all major refinery processing units; (iii) a joint sub-model coupling the CM sub-model and its downstream RM sub-model; (iv) a new multi oil-product pipeline distribution (MOPD) sub-model that considers comprehensive handling measures for oil transmix including downgrading, blending, and distillation operations; and (v) a joint sub-model coupling the upstream RM and the downstream MOPD. The developed CM&RM&MOPD model is a large-scale continuous-time and continuous-volume based MINLP model, where the objective is to minimize the total operating cost subject to various constraints such as operating rules, product specifications, inventory limits, delivery constraints, and oil-product demands at each oil depot.

2. CM&RM&MOPD model

The detailed CM sub-model is remodeled from a previous study (Zhang and Xu, 2014); the RM sub-model is based on the study of Xu et al. (2017); while the MOPD sub-model involves the comprehensive TM handling measures and tanks inventory management is based on the MOPD sub-model development (Yu et al., 2020) with modest modifications. Due to the limited space, more detailed model equations and assumptions could be referred to these three corresponding articles.

2.1. CM&RM joint sub-model

Equation (1) is employed to mathematically connect the CDU charging amount at the front-end CM outlet and the RM inlet. For example, the CDU charging amount at the CM

outlet during time event one (n=1) will be equal to the CDU charging amount at the RM inlet during the injection of the first new slug (i=i_0+1) into the downstream pipeline.

$$Fed_{i,c,unt} = F^c(unt,c,n),\ \ \forall i \in I^{new}, 1 \le n < |N|, \\ n = the\ order\ of\ new\ slug\ i, c \in C, unt \in DU \tag{1}$$

2.2. RM&MOPD joint sub-model

Sub-models RM and MOPD are materially linked through the mass balance of refinery storage tanks as shown in Eq. (2). Once the new slug *ii* is fed into the pipeline, Eq. (2) will calculate the leftover inventory of the refinery storage tank containing the oil product *p* ($Invp_{ii,p,rst}$), which is equal to the previous inventory after feeding slug *ii*-1 ($Invp_{ii-1,p,rst}$) plus the amount of oil product produced by refinery during the time period of feeding new slug *ii* ($Bldp_{ii,p}$), and minus the amount of oil product injected into the pipeline during the same time period ($Fedp_{ii,p}$). Eq. (3) constraints both the lower and upper bound of the refinery storage tanks.

$$Invp_{ii,p,rst} = Invp_{ii-1,p,rst} + Bldp_{ii,p} - Fedp_{ii,p},\ \ \forall ii \in I^{new}, p \in P, rst \in RST \tag{2}$$

$$Invp^{lo}_{ii,p,rst} \le Invp_{ii,p,rst} \le Invp^{up}_{ii,p,rst},\ \ \forall ii \in I^{new}, p \in P, rst \in RST \tag{3}$$

2.3. Objective function

The objective function of the CM&RM&MOPD model is to minimize the total process cost, which is defined in Eq. (4). It contains three main items representing the cost for the three sub problems, specifically, the first item *Cost_CM* represents the total CM sub-problem cost, the second item *Cost_RM* represents the total RM sub-problem cost, while the third item *Cost_MOPD* represents the total MOPD sub-problem cost.

$$\min Cost\_Holistic = \left(Cost\_CM\right) + \left(Cost\_RM\right) + \left(Cost\_MOPD\right) \tag{4}$$

3. Case study

The scope of the scheduling problem consists of (i) three single-parcel vessels carrying their respective crudes, one single docking berth, four storage tanks, four charging tanks; (ii) a refinery plant processing two types of crudes and producing four blending oil products and four corresponding refinery storage tanks; (iii) one long-distance pipeline connected to four oil depots; (iv) each depot has two storage tanks for each oil product; and (v) the farthest depot along the pipeline has nine additional storage tanks, including two product-rich transmix tanks for storing rich-product of each type of oil product (i.e., eight tanks) and one well-mixed tank storing well-mixed *ic*-TMs.

3.1. Computational performance and economic analysis

Based on our study, the developed CM&RM&MOPD model has been programmed with GAMS v25.1.2 and implemented on Intel 3.4 GHz Windows PC with 16.0 GB memory. The optimization solver DICOPT (based on the extensions of the outer-approximation algorithm) is adopted to solve the MINLP problems (Duran and Grossmann, 1986), where CPLEX and CONOPT4 are employed as the sub solvers for MIP and NLP sub problems, respectively. The problem size and results of economic and computational performance are summarized in Table 1. Note that the objective only considers the utility cost and

various operating cost, while other costs like crude-oil purchasing cost, labor, maintenance, and royalties are not considered.

Table 1. Economic and computational performance results for the CM&RM&MOPD case

Problem Size and Solution Efforts			
No. of constraints	25,552	No. of binary var.	3,698
No. of continuous var.	16,132	Non-zeros	107,878
Optimality Gap	0.001%	CPU time (s)	2,038
Economic Results (k$)			
Total cost (objective)	281,471.68	MOPD sub-problem cost	226,911.55
CM sub-problem cost	430.54	Blending credit	-771.25
RM sub-problem cost	54,129.59		

3.2. Results of crude-oil management scheduling

The CM scheduling results is shown in Figure 3, where the numbers above bars represent transferred crude-oil volumes (Mbbl). Various filling patterns indicate the source units/facilities of crude-oils; while different colors denote specific time events when operations occur. Small solid black schedule bars represent the RPST time (Xu et al. 2017). Overall, four time events are employed for the CM scheduling. The first blue bar in the figure, for example, it means at time event 2, a total volume of 1,000 Mbbl crude-oil is transferred from parcel 1 to storage tank 2 from day 0 to day1.1. As shown, two refinery CDUs have received different types of crude blends from charging tanks at different time events. Thus, a RPST time is located between any pair of time events.

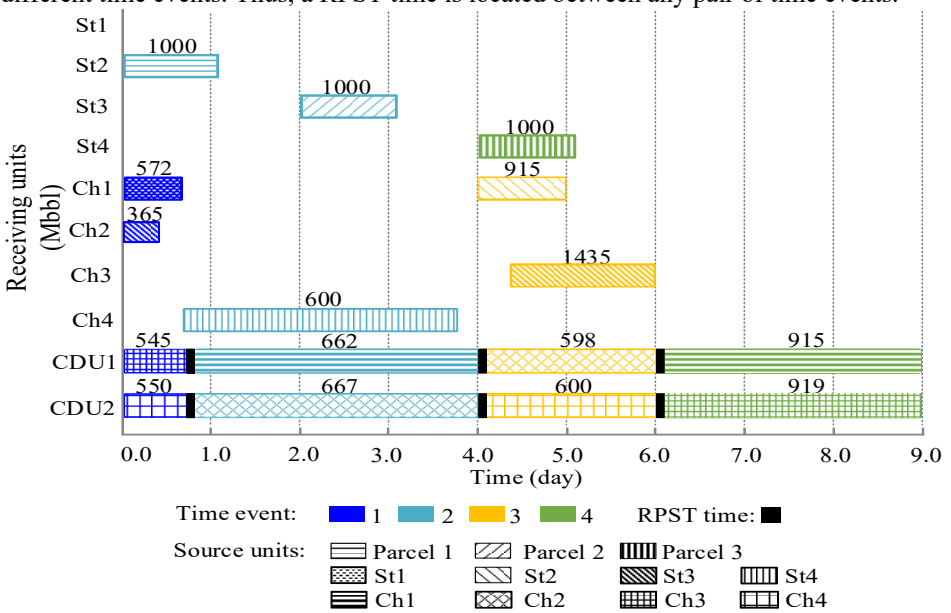


Figure 3. Scheduling result by considering the RPST time.

3.3. Results of refinery manufacturing scheduling

The scheduling results of the RM sub-problem are presented in Figures 4. It displays the production profiles of the blended oil products, where the colour of the 3-D column represents the scheduling time duration. The overall inventory based on the holistic scheduling are constrained within their capacity constraints.

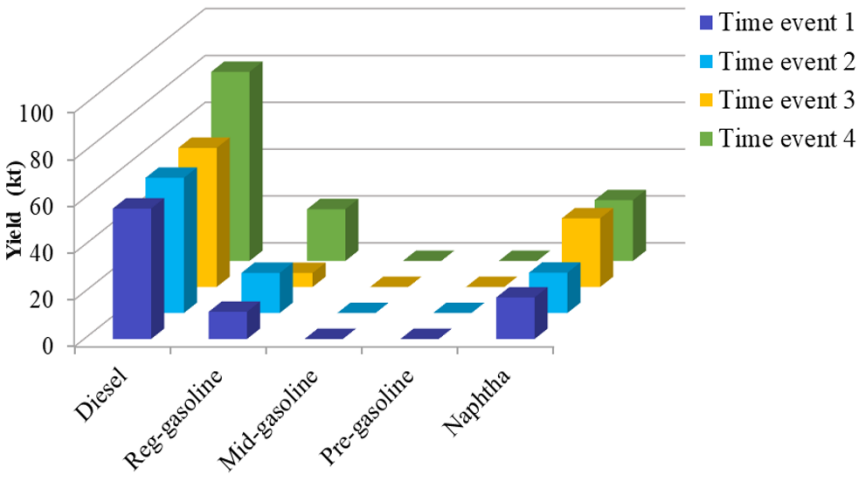


Figure 4. Production profiles of refinery blending product.

3.4. Results of multi oil-product pipeline distribution scheduling

The Gantt charts for the scheduling results of MOPD sub-problem is shown in Figure 5. The filling color of bars and arrows represents the type of oil product. The solution values of injected oil-product type, volume, and time duration are marked near the patterned horizontal arrows at the pipeline inlet. A recycling operation and the recycled volume of an *ic*-TM at the farthest depot are specifically denoted by a bent arrow with a volume value. The *ic*-TMs inside the pipeline during scheduling are represented by short and slash-patterned bars. The delivery operations and delivered volumes of oil products to corresponding depots are denoted by vertical arrows and values above.

4. Conclusions

In this paper, a continuous-time and continuous-volume based general integrated optimal scheduling framework for holistic refinery supply chain covering the crude-oil management, the refinery manufacturing, and the multi oil-product pipeline distribution has been developed. The objective is to minimize the total operating cost subject to various operating constraints. It can simultaneously provide optimal solutions for crude-oil vessels unloading, transfer connections, timings, and volumes; refinery unit operations, production profiles; oil-product inventory management at refinery and depots, oil slug movement profiles inside pipeline, as well as oil-product discharging strategies from pipeline to multiple tanks at different depots.

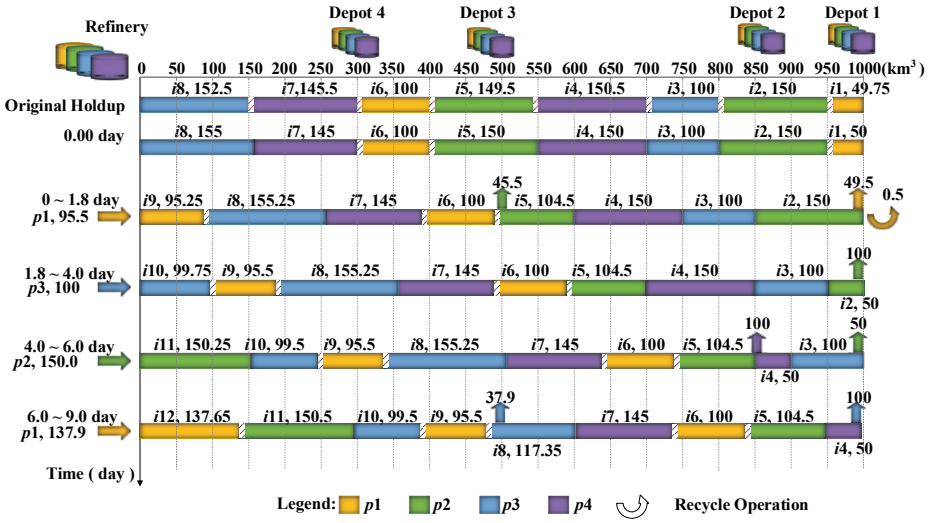


Figure 5. MOPD schedule and pipeline profiles.

Acknowledgements: This work was supported in parts by The Center for Midstream Management and Science (CMMS) and The Center for Advances in Port Management (CAPM) at Lamar University in Beaumont, Texas, USA.

References

Cafaro, D. C., & Cerdá, J. (2010). Operational scheduling of refined products pipeline networks with simultaneous batch injections. Computers & Chemical Engineering, 34(10), 1687-1704.

Chima, C. M. (2007). Supply-chain management issues in the oil and gas industry. Journal of Business & Economics Research (JBER), 5(6).

Duran, M. A., & I. E. Grossmann (1986). An outer-approximation algorithm for a class of mixed-integer nonlinear programs. Mathematical Programming 36(3): 307-339.

Guyonnet, P., Grant, F. H., & Bagajewicz, M. J. (2009). Integrated model for refinery planning, oil procuring, and product distribution. Industrial & Engineering Chemistry Research, 48(1), 463-482.

Qu, H., Wang, S., & Xu, Q. (2019). Integrated proactive and reactive scheduling for refinery front-end crude movement with consideration of unit maintenance. Industrial & Engineering Chemistry Research, 58(27), 12192-12206.

Xu, J., Zhang, S., Zhang, J., Wang, S., & Xu, Q. (2017). Simultaneous scheduling of front-end crude transfer and refinery processing. Computers & Chemical Engineering, 96, 212-236.

Yang, H., Bernal, D. E., Franzoi, R. E., Engineer, F. G., Kwon, K., Lee, S., & Grossmann, I. E. (2020). Integration of crude-oil scheduling and refinery planning by Lagrangean Decomposition. Computers & Chemical Engineering, 138, 106812.

Yu, L., Chen, M., & Xu, Q. (2020). Simultaneous scheduling of multi-product pipeline distribution and depot inventory management for petroleum refineries. Chemical Engineering Science, 220, 115618.

Zhang, S., & Xu, Q. (2015). Refinery continuous-time crude scheduling with consideration of long-distance pipeline transportation. Computers & Chemical Engineering, 75, 74-94.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50095-6

Roadmap to digital supply chain resilience

Adnan Al-Banna[a,b], Robert E. Franzoi[a], Brenno C. Menezes[a*], Ahad Al-Enazi[a,c], Simon Rogers[d], Jeffrey D. Kelly[e]

[a]*Division of Engineering Management and Decision Sciences, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar*
[b]*Department of Logistics and Supply Chain, Milaha, Doha, Qatar*
[c]*Commercial Planning and Optimization, QatarGas, Doha, Qatar*
[d]*Digital Solutions, Yokogawa, Tokyo, Japan*
[e]*Industrial Algorithms Ltd., 15 St. Andrews Road, Toronto M1P 4C3, Canada*
bmenezes@hbku.edu.qa

Abstract

As supply chains evolve from local trading entities to global physical and virtual markets, today's organizations are privileged with enhanced access to unprecedented opportunities in volume, variety, deliver time, transportation mode, of resources and goods. However, given today's level of interdependence of these entities that permits reduced mismatches in their processes, such organizations are more susceptive to disruptions in their networks. Over the past decades, global economies have experienced several crises induced by irrepressible circumstances that are either human-influenced, such as geopolitical conflicts and cyber-attacks, or provoked by nature, such as natural disasters and pandemic outbreaks. Hence, organizations have been placing substantial emphasis on supply chain resilience (SCR), with the objective of mitigating the impact of unforeseen risks on their supply chains, logistics, and their subsequent consequences on cost control and revenue maximization. Conventional SCR relied on increased safety stock levels, partial order allocations, supplier's diversifications, among others. On contrary, next generation of resilient supply chain operations can be reached by the digital transformation of its elements into the SCR modeling and control. The objective of this paper is twofold. First, it introduces the role of the digital transformation, advanced analytics, automation, and augmentation of the SCR with the support of cyber-physical systems and security (CPSS) solutions from the industry 4.0 age. Second, it provides a structured view on artificial intelligence (AI) and internet of things (IoT) technologies, aiming to establish a robust, timely, and successful digital supply chain resilience (DSCR). We believe the discussion provided and provoked herein will aid organizations towards addressing proper digital capabilities for achieving higher levels of visibility and control, enhanced revenues, reduced costs, and improved supply chain resilience augmented by the power of digitalization, automation, and artificial intelligence.

Keywords: Supply chain management, digital transformation, digital supply chain resilience, artificial intelligence, internet of things.

1. Introduction

The prerequisites of efficient business management are not limited to human and financial resources and their supporting ecosystem. Instead, organizational survival and success

are underpinned by supply chain resilience (SCR), which encompasses the abilities to predict, avoid, contain, manage, recover from, and eventually alleviate adverse impacts of continuous disruptions and uncertainties (Melnyk et al., 2014). Previous literature has predominantly addressed the conventional supply chain efficiency (SCE) strategies involving multiple sourcing, partial order allocation, and extra inventory stocks (Tang, 2006; Vanany et al., 2009). Although the aforementioned strategies could have been sufficient to manage organizational risks in the previous decades, limiting the current organizational resilience toolbox to SCE strategies may lead to incurring losses due to a lack of no costly responsiveness to sudden events. Schreckling et al. (2017) highlight that for most industries, regardless of the field, location, and application, it is fundamental to advance towards enhanced digital transformation capabilities to interconnect entities and levels of the supply chain (ELSC). Agrawal (2018) emphasizes that digital transformation is not a choice in the current world of globalization, but it is imperative for all industries to find synergies of a collection of companies' segments and avoid the pitfalls of segregated ELSC. Most scholars addressed SCE and SCR in isolation, while few works have acknowledged their interconnected nature. In contrast, the distinctive approach of Dolgui et al. (2020) provides a concise differentiation of SCE and SCR and elaborates on their interconnections with digital supply chain (DSC) and sustainable supply chain (SSC), as illustrated in Figure 1, which shows the main strategies of the DSC, SCR, SSC, and SCE interplays. Although Dolgui et al. (2020) offer an efficient representation of the interconnectedness of these supply chain (SC) frameworks, it is noteworthy to mention that most previous works have not addressed their correlation.

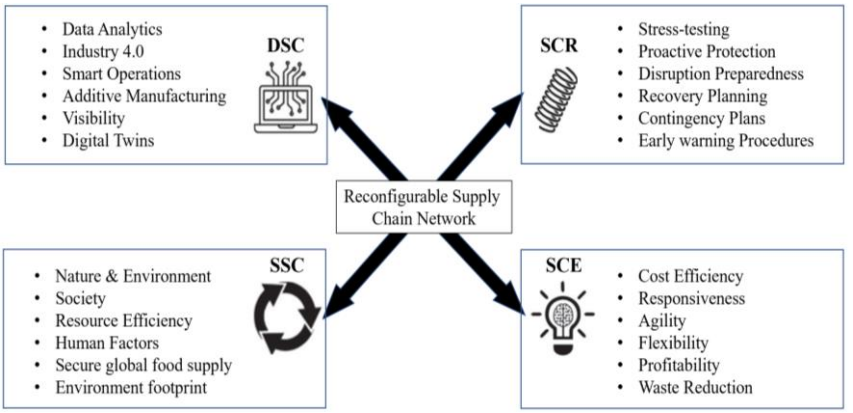


Figure 1. The reconfigurable supply chain network (adapted from Dolgui et al., 2020).

The main contributions of this work rely on addressing a multi-domain supply chain, which involves the interconnectedness of SCE, SCR, DSCE, and DSCR. We emphasize that DSCR elements must be embedded in the organization processes throughout its suppliers-input-process-outcome-customer (SIPOC) workflows. This ensures optimum preparedness to build enhanced: i) resistance against disruptions by employing avoidance and containment strategies; and ii) recovery capabilities by stabilizing and returning to the pre-disruption and pre-disturbance performance levels.

Early publications on SC have predominantly focused on SCE, whereas concepts of SCR have only recently been addressed and studied. Researchers have primarily addressed the impact of digitalization in the supply chain as SCE and later focused on its influence as SCR; therefore, most of the research is found in the SCE domain. Literature indicates a

significant evolution in the supply chain research over the past decades, primarily because of the fast expansion of global markets and their associated opportunities and threats. However, DSCE and DSCR remain to be premature research areas where significant opportunities exist. The digital transformation of companies must be augmented with the relevant resilience and associated optimality that involve automation and digitalization of business processes to capture the status of the ecosystems and provide the necessary scalability and evolution into optimized and automated operations and controls (Menezes et al., 2019a). Gartner (2018) predicts that in 2022, 85% of all artificial intelligence projects are expected to fail due to data inconsistencies, inappropriate algorithms, and inefficient human capital. Hence, there is a fundamental need for a robust roadmap of digital transformation towards a supply chain resilient state.

2. Stages towards the digital transformation

2.1. First stage: SIPOC (suppliers-input-process-outcome-customer) integration

The first stage of any digital transformation project demands a detailed gap analysis that considers the overall SIPOC processes and interdependencies, whereby the current digital standpoint, desired ultimate organization status, and project milestones are identified. Hence, the strategic macro-level visions and operational micro-level requirements should be carefully envisioned, discussed, and outlined to allow future scalability and straightforward implementation. This should be ideally performed in a cross-departmental fashion and under strong sponsorship from top leadership to achieve a company-wide project environment. Similarly, the success of digital transformation projects requires critical factors, including competent human capital, adequate technology selection, efficient implementation, top management sponsorship, sufficient training and incentives for technology adoption, enterprise resources planning (ERP) integration capabilities, optimizing on-premise versus cloud storage, adequate data migration and protection, robust cybersecurity policies and infrastructure, clear KPI's with margins-of-error and escalation mechanisms and strategies for minimizing staff resistance to changes. The aforementioned factors are summarized by Bascur (2020) in three distinct domains, namely: a) people, b) business processes, and c) adopted technologies, whereby the importance of creating a digital transformation environment is highlighted as well. These three domains are considered the corner stones for the digital transformation success triangle and are required to provide optimal results.

2.2. Second stage: SIPOC autonomy

Once the digital transformation roadmap is established, agile organizations can use the aforementioned approach to combine their knowledge of economy, finance, technology, market dynamics, and business resilience. This provides efficient capabilities for handling continuous risks and uncertainties and progressively enhancing business resilience measures with the support of cyber-physical systems and security (CPSS) solutions. Throughout the transformation process, organizations need to acknowledge the interconnection of DSCR with the capabilities provided by artificial intelligence (AI) and the Industrial Internet of Things (IIoT). However, DSCR is expected to be as reliable as its weakest Industry 4.0 components. Hence, the concept of IIoT ecosystem trustworthiness is fundamental at all levels within digital ecosystems. Sharma et al. (2020) categorize an IIoT ecosystem into four interconnected layers, as illustrated in Figure 2, encompassing the internet, devices, support, and applications. The trustworthiness of any IIoT ecosystem depends on the proper and continuous functioning of all layers to enable

the sensors to acquire information correctly. Information is transferred through the internet in a timely and holistically fashion. The support layer efficiently receives and decodes it to allow the application layer to display, process, and present it on the associated user interface for an appropriate decision-making process. We propose the addition of a fifth layer, which is integrated to the other layers towards decoding and converting data and information into a decision-making process.

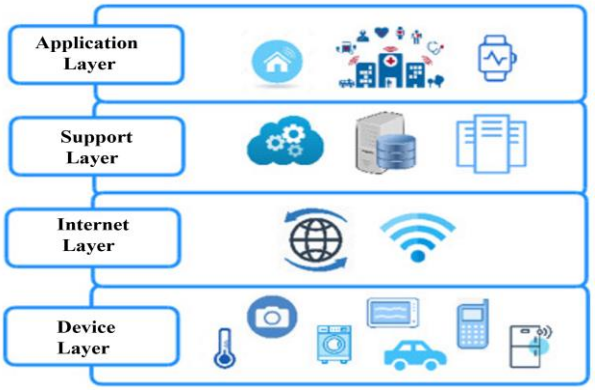


Figure 2. Four-layer architecture of IIoT ecosystem (adapted from Sharma et al., 2020).

The growth of connected IIoT devices is expected to generate increasing amounts of data, emphasizing the high need for a better-managed cybersecurity infrastructure that shields data integrity (IDC, 2019). This highlights the importance of adopting a regular reviews and PDCA (Plan-Do-Check-Act) cycles to continuously ensure data reliability and efficient decision-making. Rocher et al. (2019) adopts this concept and argues that cyber-physical systems should be continuously evaluated to ensure their reliability. The digital ecosystem vulnerability is further amplified whenever the instantaneous SCA (Sense-Calculate-Actuate) cycle is used (Menezes et al., 2019b). This becomes especially critical as the IIoT transforms into the state of AoT (Automation-of-Things). The ecosystem automatically manages the operation by sensing, calculating, and actuating more intelligently and autonomously, without human intervention, within seconds or minutes, depending on the business and industry applications. Such technologies provide capabilities to identify opportunities of (re-) designing and (re-) implementing supply chains to improve the production state within an online closed-loop fashion (Franzoi et al., 2021). The massive potential explains the increasing global spending on IIoT, estimated to surpass 1 trillion US dollars in 2023 (IDC, 2020). Hence, digital transformation in supply chains is expected to provide further insights beyond intrinsic knowledge and historical experience, which can be enhanced by ubiquitous visibility and connectivity among relevant systems.

3. Continuous evolution towards the digital supply chain resilience

Organizations are not more myopic to the importance of continuous evolution for the complete digital transformation on their supply chains. They rely on continuously exploring new advanced solutions and alternatives that could deliver higher reliability, reduced costs, higher efficiency, enhanced resilience, etc. For example, customers have been satisfied with monitoring of the geographical locations of their shipments a decade ago, but now they are able to monitor the shipments location, temperature, humidity, and

visually. The IoT devices of a decade ago had to undergo many challenges of limited battery life, expensive data transfer charges, harsh weather, among other. While today's equipment and advances in technologies have eased these challenges. The fast-paced environment of our Society 5.0 and Industry 4.0 leads to sudden innovation changes, in which today's state-of-the-art technology shall soon become obsolete. Hence, organizations are encouraged to continuously transform and evolve their digital capabilities in their supply chains to avoid being among the 40% expected to fade away in their processes (Schreckling et al., 2017).

4. Conclusions

The development and continuous evolution of digital supply chain resilience is fundamental for today's organizations survival and success. DSCR and its prerequisites of digital transformation and automation success relies on three cornerstones related to human capital, business processes, and adopted technologies. In the human capital realm, the relevant stakeholders responsible for designing, implementing, and managing the digital transformation should be up to date with state-of-the-art technologies and continuously explore new inventions and trends. It is imperative to introduce training and incentive strategies for employees with the objective of enhancing employees' technological literacy and adoption rates within an organization. In the business processes realm, it is a must to conduct a detailed gap analysis of the organizational business requirements that serves the organization strategic vision, in an agnostic perspective that's receptive to innovative solutions and new methodologies with the objective to create an efficient SIPOC that meets the customer requirements, reduces costs, and maximizes revenues. The drawback of ignoring the business processes of the digital transformation triangle yields a mere digitalization project without tangible digital transformation. In the adopted technologies realm, it is critical to explore the organizational digital and technological requirements that meet the previously established business processes substantiated by SIPOC reviews and structured analysis of the right technologies, ERP, licenses, implementation, data storage, right cybersecurity measures, among others. This is crucial for building future technology scalability, enhancing efficiency, thereby decreasing costs, and increasing revenues. The above-mentioned success triangle supports digital supply chain resilience as it results in enhanced preparedness for achieving resistance against disruptions and disturbances (by employing avoidance and containment strategies) and recovery capabilities (by stabilizing and returning to the pre-disruption/disturbance performance). Nevertheless, AI and IoT ecosystems trustworthiness levels remain an imperative accompanying factor to be carefully considered. A key objective of augmenting organizational SCR with the SCA and AoT capabilities is to enhance real-time visibility, responsiveness, and control. However, there are multiple risks of malfunctions, cyber-attacks, and power failures, severely hindering reliability, robustness, and reliability. Hence, rigorous and continuous reviews and upgrades are required to ensure optimum results of DSCR technological elements. A multi-domain supply chain involving the SCE, SCR, DSCE, DSCR, and their interconnectedness are addressed herein, where DSCR elements are embedded in the organizational processes throughout SIPOC workflows. The proposed discussion guides the implementation of enhanced digital transformation capabilities towards better organizational visibility and control, and more efficient operations.

References

P. Agrawal, R. Narain, 2018, Digital supply chain management: an overview, IOP Conf. Series: Materials Science and Engineering, 455(1), 012074.

O. Bascur, 2020, Digital transformation for the process industries: a roadmap, CRC Press.

A. Dolgui, D. Ivanov, B. Sokolov, 2020, Reconfigurable supply chain: the X-network, International Journal of Production Research, 58(13), 4138–4163.

R. E. Franzoi, B. C. Menezes, J. D. Kelly, J. A.W. Gut, 2021, A moving horizon rescheduling framework for continuous nonlinear processes with disturbances. Chemical Engineering Research and Design, 174, 276-293.

Gartner, 2018, Gartner Says Nearly Half of CIOs Are Planning to Deploy Artificial Intelligence. Available at: https://www.gartner.com/en/newsroom/press-releases/2018-02-13-gartner-says-nearly-half-of-cios-are-planning-to-deploy-artificial-intelligence, accessed on 15 September, 2021.

International Data Corporation, 2019, Steady Commercial and Consumer Adoption Will Drive Worldwide Spending on the Internet of Things to $1.1 Trillion in 2023. Available at: https://www.idc.com/getdoc.jsp?containerId=prUS45197719, accessed on 15 September 2021.

International Data Corporation, 2020, Worldwide Spending on the Internet of Things Will Slow in 2020 Then Return to Double-Digit Growth. Available at: https://www.idc.com/getdoc.jsp?containerId=prUS46609320#:~:text=A%20new%20update%20to%20the,in%20the%20November%202019%20release.&text=Consumer%20spending%20on%20IoT%20solutions,year%20over%20year%20in%202020.https://www.idc.com/getdoc.jsp?containerId=prUS45197719, accessed on 15 September 2021.

L. Koh, G. Orzes, F. Jia, 2019, The fourth industrial revolution (Industry 4.0): technologies disruption on operations and supply chain management, International Journal of Operations and Production Management, 39, 817–828.

S. Melnyk, D. Closs, S. Griffis, C. Zobel, J. Macdonald, 2014, Understanding supply chain resilience, Supply Chain Management Review, 18(1), 34-41.

B. C. Menezes, J. D. Kelly, A. G. Leal, G. C. Le Roux, 2019a, Predictive, prescriptive, and detective analytics for smart manufacturing in the information age. IFAC Papers Online, 52(1), 568-573.

B. C. Menezes, J. D. Kelly, A. G. Leal, 2019b, Identification and design of Industry 4.0 opportunities in manufacturing: examples from mature industries to laboratory level systems. IFAC-PapersOnLine, 52(13), 2494-2500.

G. Rocher, J. Y. Tigli, S. Lavirotte, N. Le Thanh, 2018, A possibilistic i/o hidden semi-markov model for assessing cyber-physical systems effectiveness, IEEE International Conference on Fuzzy Systems, 1-9.

C. K. Sahu, C. Young, R. Rai, 2020, Artificial intelligence (AI) in augmented reality (AR)-assisted manufacturing applications: a review, International Journal of Production Research, 59(16), 4903-4959.

E. Schreckling, C. Steiger, 2017, Digitalize or drown. In Shaping the digital enterprise: trends and use cases in digital innovation and transformation, Springer Cham, 3-27.

A. Sharma, E. S. Pilli, A. P. Mazumdar, P. Gera, 2020, Towards trustworthy Internet of Things: A survey on Trust Management applications and schemes, Computer Communication, 160, 475–493.

C. S. Tang, 2006, Perspectives in supply chain risk management, International Journal of Production Economics, 103(2), 451-488.

I. Vanany, S. Zailani, N. Pujawan, 2009, Supply chain risk management: literature review and future research, International Journal of Information Systems and Supply Chain Management, 2(1), 16-33.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50096-8

Development of Flexible Framework for Biomass Supply Chain Optimisation

Ken-Ichiro Sotowa

Dept. of Chemical Engineering, Kyoto University, Nishikyo-ku, Kyoto 615-8510 Japan
sotowa@cheme.kyoto-u.ac.jp

Abstract

Many studies are being conducted to develop methods for the conversion of biomass into useful chemicals. To exploit the results from these studies and realise a biomass-based chemical industry, an appropriate supply chain system must be designed. It is therefore necessary to select the best factory locations and reaction pathways as well as derive the best operational strategy. When biomass is used as feed stock, seasonal variations in the amount of available biomass induces variations in the amount of materials flowing in the system. Thus, several thousands of variables and equations are involved in such design problems. In this study, different elements in supply chain networks were classified into four types according to the feature of their mathematical representations: storage, conversion, transport, and utility models. This abstraction enabled us to model a wide class of supply chain networks in a unified framework. An optimal design system for a supply chain system based on this concept was implemented as a web system.

Keywords: biomass, supply chain, seasonal variation, optimisation.

1. Introduction

Biomass has been attracting a significant amount of attention as a raw material for the chemical industry. There are several substances that can be chemically extracted or synthesised from biomass, and synthetic methods for various substances starting from biomass are currently being developed (Serrano-Ruiz *et al*., 2010, Gérardy *et al*., 2020). If a system capable of producing high-value chemical substances can be implemented, the dependence of the chemical industry on petroleum resources can be reduced, thereby contributing to the building of a sustainable society highly based on renewable resources.

When implementing a technology for the conversion of biomass to chemical products, it is necessary to construct an optimal supply chain network starting from obtaining raw materials to the production and delivery of the products. The network should be designed by considering the characteristics of the area to which the network is to be implemented. This is due to the characteristics of chemical production systems that utilise biomass as a raw material. Because biomass is widely dispersed in an area, it is necessary to optimally select the location of the processing factories. The amount of available biomass varies seasonally. For a biomass-based energy system, there is only one product, that is, energy. However, for a chemical production system, multiple products with different prices and demands must be manufactured. It is therefore necessary to select the optimal reaction pathways from many possible alternatives. The reuse of waste energy and heat should also be considered.

Several researchers working in the field of biomass supply chain formulate the design problem as an MILP and solve it using optimisation software. By using a superstructure of the supply chain network covering all possible factory locations, reaction pathways, and transport of materials, the design problem can be described as an MILP. However, the number of variables and formulas that appear in MILP often exceeds thousands because there are several types of substances and seasonal fluctuations that must be considered. Even a small modification of the problem, such as the incorporation of new reaction techniques or changing the candidate location of the factory, requires a significant amount of effort. However, to implement a biomass-based manufacturing process for chemicals, various case studies must be conducted to evaluate the impact of any possible variations in the conditions. However, because the formulation and modification of the optimisation problem is extremely time-consuming, it is only possible to conduct several case studies.

In this study, we aimed to develop a technology that will enable us to formulate the optimisation problem quickly and subsequently derive the optimal design of a biomass supply chain network. Any change in the superstructure, number of reactions, and substances can be easily reflected in the system. In this presentation, we report on the key to the modelling method and the development of a web system based on this modelling approach.

2. Assumptions during model development

The following assumptions were made in the creation of the supply chain model.

- Each chemical substance whose state changes due to drying, chipping, and packaging is treated as a substance different from the one that has not undergone processing.
- To consider seasonal fluctuations, the target period is divided into N_t terms, and the change in the stored amount of a substance is calculated for each term.
- Only one substance can be stored in one storage.
- Only one type of substance can be transported via one transportation means at a time.
- Each utility is treated as a substance that cannot be stored.

3. Generalised representation of the supply chain elements

Several different types of operations are involved in the synthesis of chemical products from biomass in a supply chain. These operations are typically biomass harvesting,

Table 1 Four generalised elements in the supply chain network

Elements	Function	Usage
Conversion	Convert substance or utility to different ones	Chemical conversion, drying, chipping, separation and purification, packaging
Storage	Store substance according to the difference in the incoming and outgoing flow rates.	Storage of substance
Transport	Transfer substance from a storage to another	Transport, harvesting, product despatch, waste disposal
Utility	Supply or recover utilities	Supply, reuse, purchase and sale of utility

chipping, drying, storage, various reactions, transportation, sale, waste disposal, and purchase of utility. First, we considered the mathematical characteristics of these elements and classified them into four types, as presented in Table 1.

The elements included in a single category can be represented using a common mathematical model. The conversion model changes one substance or utility into another substance or utility and can be expressed by the following mathematical formula:

$$\sum_i r_{ri} R_{ri}(t) = \sum_i p_{ri} P_{ri}(t) \tag{1}$$

$R_{ri}(t)$ and $P_{ri}(t)$ are, respectively, the consumption and production rates of the raw material or utility i in the converter r in the period t, and r_{ri} and p_{ri} are the stoichiometric coefficients. Chipping or drying does not change the chemicals, but because it is assumed that these operations are treated as producing different substances, they can be expressed in this model.

The storage model is expressed using the following equation.

$$S_s(t) = S_s(t-1) - \sum_{r|s} R_{ri(s)}(t) + \sum_{r|s} P_{ri(s)}(t) + \sum_{y|s} \delta_{y|s} Y_{y(i(s))}(t) \tag{1}$$

$S_s(t)$ is the amount of substance in storage s at the end of period t. The substance stored in storage s is represented by $i(s)$. The second and third terms on the right side are the rate of increase or decrease of the substance $i(s)$ by the converter r connected to the storage s. The fourth term is the speed at which the substance enters and exits by means of transportation y, and $\delta y|s$ is +1 when y is brought into storage s and -1 when y is removed.

The transport model solely expresses the amount of transportation $Y_y(t)$ that occurs in y. For transport, it is usually necessary to define the start and end points; however, there are certain exceptions. Transportation with no defined starting point represents the external inputs to the supply chain network, such as biomass harvesting and electricity purchases. On the contrary, transportation without an endpoint represents the emission of a substance outside of the system, either as a product to be sold or as a waste that is discharged. For product dispatch, the cost associated with it is expressed as a negative value. Another rule for a transportation model is that it must always be connected to a storage model.

As stated in Section 2, a utility is treated as a substance that cannot be stored. A utility model is defined as a combination of storage and transport models to model the handling of a utility in a biomass supply chain system. The storage model is expressed by Eq. (2); however, $S_s(t)$ is always zero. Transportation models are used to express the purchase and sale of utilities.

Furthermore, in the conversion, storage, and transport models, the cost of operation is expressed as a function of the conversion rate, amount of storage, and transfer rate. For conversion and storage, the capital cost of the equipment is modelled as a function of the maximum capability.

4. Concept of site and path

By combining the abovementioned four elements, the design problem of a general biomass supply chain network can be easily modeled. In large optimisation problems, however, the superstructure network tends to be very complex owing to the large number of nodes and edges. Notably, in such problems, several candidate sites exist for factory locations, all of which have identical reaction networks. Therefore, in this study, the concept of site was introduced. In a single site, the reaction pathways can be defined using conversion, storage, and utility models; however, no transport models can be in it.

Several different types of substances are involved in a supply chain network; therefore, given a pair of sites, it must be possible to transport several substances between them. Those transport models have common origin and destination sites. A path was defined as a collection of such transport models. This path can be considered as a road connection between the two sites.

5. Formulation of the optimisation problem

The objective function of the supply chain optimisation problem is defined as the total annual cost of operation. The cost includes the depreciation cost of the equipment, in addition to the cost required for biomass collection, transportation, conversion, storage, and utility. The sale of products is subtracted from this cost. The optimisation variables are the transport rate, conversion rate, storage amount, and so on. Because the problem was formulated based on the superstructure, the selection of factory location and reaction pathways is possible with this framework.

Because the storage model connects the conversion and transport models, the mathematical formulation of the entire problem can be performed automatically when all the necessary information is collected. The conversion model consists of only stoichiometric relations and cost formulas. The transport model only expresses the cost mathematically. The storage model can be described using a mathematical formula to calculate the balance between the reaction and transport rate associated with it. Thus, using the concept of the four generalised elements, we can easily generate the mathematical equations for the optimal design of the biomass supply chain network. Because the equations can be obtained automatically from the collected data, this approach is expected to be a powerful tool in the development of optimal designs of systems involving several substances and reaction pathways and showing complex seasonal variations.

6. System development

To design a supply chain network capable of manufacturing chemical products from biomass, it is necessary to collect data from different industries such as agriculture and forestry, chemical industry, transportation industry, trading companies, and local governments. Because the amount of data to be collected is enormous, it is inefficient if a single person performs this task of collecting data and then providing it into the system. Therefore, a supply chain optimisation system was built as a system on a web, to ensure that anyone who owns the data for the supply chain design can enter them directly into the system. Figures 1-3 show an example of the browser screen of this system.

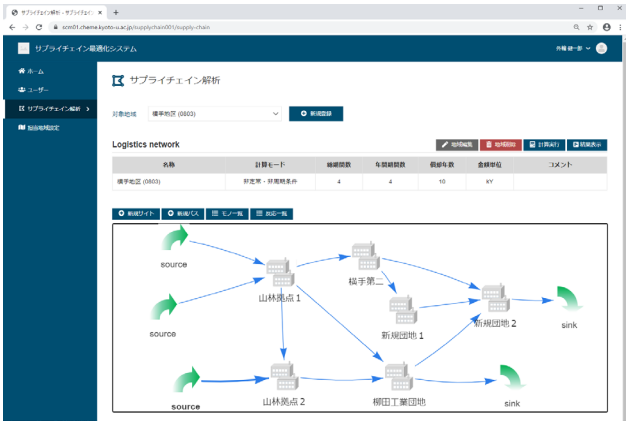


Figure 1 Example of the network comprising consisting of sites and paths (texts are in Japanese)

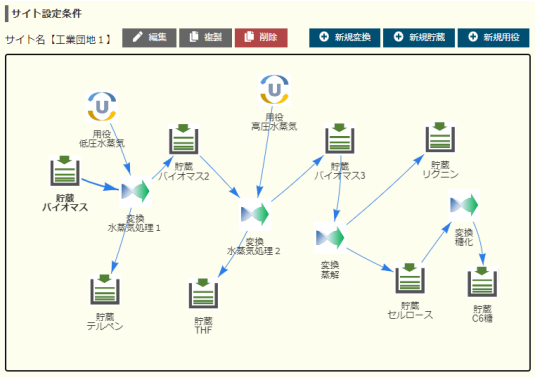


Figure 2 Example of the reaction path network in a site drawn using conversion, storage, and utility models. (texts are in Japanese)

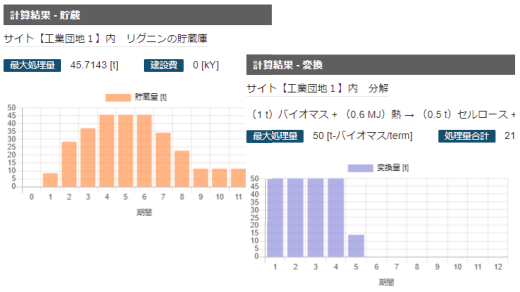


Figure 3 Example of graphical representation of the optimisation results. (texts are in Japanese)

The configuration of the superstructure can be input by a GUI on a web browser. The first step is to define the substances and the available conversion technologies. Then, a network of sites and paths is drawn (Figure 1). For each site, the reaction network is entered using the conversion, storage, and utility models (Figure 2). Transport models

should then be defined for each path. The results can also be viewed using the GUI (Figure 3). In this system, the mathematical formulas representing each model are automatically generated based on the input data, and the optimisation problem is defined as MILP. The problem is solved using the lp_solve or IBM CPLEX software. The data and results are secured in the mysql database. Further, multiple problems can be handled by this system.

A case study was conducted to address the hypothetical supply chain design problem. The network establishment and data inputs were completed in two hours. The problem involved the use of 4626 variables and 4680 constraints. The optimisation calculation was completed within 10 s.

7. Conclusions

Four generalised elements for modelling the superstructure of a supply chain network were proposed. A web-based optimisation system was constructed based on this concept. The mathematical expression of the optimisation problem was automatically generated from the input data of the design problem. The impact of the developed system was demonstrated via a case study.

We plan to utilise this proposed system to design a biomass supply chain system for Yokote City in Japan, where abundant forest biomass exists. Since the proposed elements can also be used to model a wide range of supply chain networks, we plan to apply it to other problems, such as the evaluation of a resource recycling system and in the design of carbon-negative societies.

Acknowledgements

This work was supported by the Cabinet Office, Government of Japan, Cross-ministerial Strategic Innovation Promotion Program (SIP), “Technologies for creating next-generation agriculture, forestry and fisheries” (funding agency: Bio-oriented Technology Research Advancement Institution, NARO).

References

R. Gérardy, D. P. Debecker, J. Estager, P. Luis and J.-C. M. Monbaliu, 2020, “Continuous Flow Upgrading of Selected C_2−C_6 Platform Chemicals Derived From Biomass,” Chem. Rev., **120**, 7219-7347.

J. C. Serrano-Ruiz, R. M. West and J. A. Dumesic, 2010, “Catalytic Conversion of Renewable Biomass Resources to Fuels and Chemicals,” Annu. Rev. Chem. Biomol. Eng., **1,** 79-100.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50097-X

Lagrangean Decomposition for Integrated Refinery-Petrochemical Short-term Planning

Ariel Uribe-Rodriguez[a,b], Pedro M. Castro[c], Gonzalo Guillén-Gosálbez[d] and Benoît Chachuat[b*]

[a] *Center for Innovation and Technology Colombian Petroleum Institute, Ecopetrol, Colombia*
[b] *Sargent Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, United Kingdom*
[c] *Centro de Matemática Aplicações Fundamentais e Investigação Operacional, Faculdade de Ciências, Universidade de Lisboa, Portugal*
[d] *Institute for Chemical and Bioengineering, Department of Chemistry and Applied Biosciences, ETH Zürich, Switzerland*
b.chachuat@imperial.ac.uk

Abstract

We present a methodology for the optimal integration of crude management (CM) and refinery-petrochemical (RP) planning operations. The physical coupling between both CM and RP optimization subproblems is via the flow rate, physical-chemical properties, and composition of the crude blends. For a given economic cost of the crude blends, which either provides a selling price for CM or a purchase price for RP, both subproblems can maximize their profits independently. But failure to integrate these two subproblems can create an imbalance between crude supply and demand. Optimizing CM and RP operations simultaneously entails the solution of large-scale, nonconvex quadratically-constrained quadratic programs (MIQCQPs). We apply a spatial Lagrangean decomposition algorithm to tackle these MIQCQPs and demonstrate it on a full-scale industrial facility. The results show that Lagrangean decomposition can outperform commercial global solvers BARON and ANTIGONE when applied to the monolithic MIQCQP. The Lagrangean decomposition can also reduce the optimality gap faster than with a clustering decomposition algorithm, leading to optimality gaps below 5% within 1 hour of CPU time.

Keywords: Lagrangean decomposition; nonconvex; planning; logistic.

1. Introduction

Integrated operations of petrochemical plants and crude oil refineries are more resilient to volatility of the hydrocarbons market than independent businesses for petrochemical commodities and fuel production. Such integration can be achieved by the exchange of by-products or intermediate streams from the refinery that are transformed into added-value products at the petrochemical units. Some by-products from petrochemical processes can also improve fuel quality at the refinery side. The refinery can provide part of the natural gas required by steam crackers and the petrochemical side can supply part of the hydrogen required by hydrotreating processes (Ketabchi, et al., 2019). Recently, deterministic global optimization and Lagrangean decomposition have been applied to short-term planning of integrated refining and petrochemical operations (Li, et. al, 2016;

Zhao, et. al, 2017; Uribe-Rodríguez, et. al, 2020), formulated as large-scale, nonconvex quadratically-constrained quadratic programs (MIQCQPs).

Herein, we investigate a spatial Lagrangean decomposition-based algorithm to solve such MIQCQPs. This problem is challenging for the following reasons: i) Compared to previous studies, a wider range of crudes are considered, which differ in terms of volume, quality, and cost; these crudes are transported by pipelines or river fleet, depending on their geographic location and can be blended to fulfill the volume and quality needs of the crude distillation units (CDUs). ii) Product demands are set for a large variety of fuels and petrochemical commodities. iii) Process units can be operated in exclusive or non-exclusive campaigns. iv) Higher connectivity between units and intermediate streams is considered in the process network. All these features lead to MIQCQPs with thousands of bilinear terms. Recently, Uribe-Rodríguez et al., (2020) tackled this problem with a deterministic global optimization approach based on process clustering decomposition (CL). Results for several scenarios have produced better incumbent solutions and smaller optimality gaps than BARON and ANTIGONE, but the optimality gap remains high for certain scenarios (11% on average). Therefore, a spatial Lagrangean decomposition-based algorithm is developed to further enhance solution quality and reduce the optimality gap.

2. Methodology

The monolithic short-term planning problem for the integrated refinery-petrochemical facility can be cast as the following MIQCQP:

$$\begin{aligned} z^*: = \max\ & f_0(x,y) \\ \text{s.t. } & f_m(x,y) \le 0\ \forall m \in \{1,\dots,M\} \\ & x \in [x^L, x^U] \subseteq \mathbb{R}^p_+, y \in \{0,1\}^q \end{aligned} \qquad \textbf{(P)}$$

where x are the non-negative continuous decision variables and y the binary decision variables used to select process operating conditions. The objective function and the constraints are furthermore quadratic in x and linear in y: $f_m(x,y) := \sum_{(i,j)\in BL_m} a_{ijm}x_ix_j + B_mx + C_my + d_m\ \forall m \in \{0,\dots,M\}$. BL_m is an (i,j)-index set defining the bilinear terms x_ix_j, while a_{ijm}, d_m, B_m and C_m are parameters. Problem **P** can also describe optimization problems appearing in business units of the facility, subproblems that can be solved independently for a given economic incentive.

In the context of refinery operations, problem **P** exhibits a block structure, which makes it amenable to Lagrangean decomposition (Pinto, 2000). Figure 1 shows the material and economic flows between crude management (CM) and the refinery-petrochemical (RP) plant, which includes refinery (REF), petrochemical (PTQ) and fuel blending (FB) operations. CM includes the operations involved in the selection, transportation, blending and allocation of the crudes. CM buys crude oil from different sources (domestic or import) and sells crude blends to the refinery. CM maximizes profit by buying cheap crude oil on the market, minimizing the transportation cost, and producing crude blends to be sold at a price λ_u that is a function of their quality. The transformation of crude oil in RP involves operations such as crude oil fractionation at the CDUs, naphtha, jet, diesel and gas oil hydrotreating, gas oil catalytic cracking, etc. RP maximizes its profit by buying enough quantity of good-quality crude blends from CM at a cheap price (λ_u), without being concerned about delivering costs.

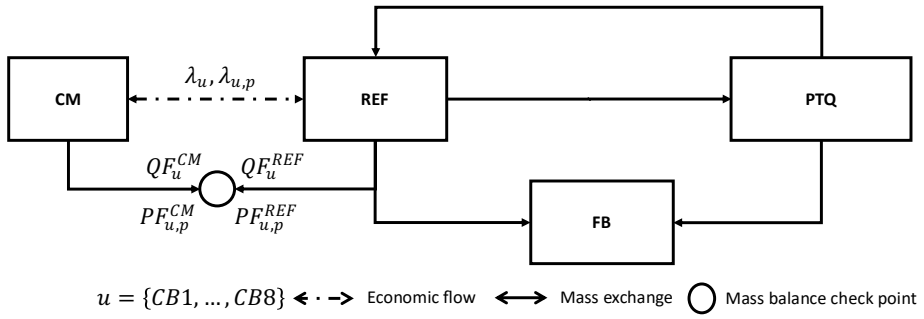


Figure 1. Subproblems derived from **P**.

*2.1. Reformulation of problem **P***

Let the index set X denote the complicating variables (appearing in the mass balance check point in Figure 1) that are shared between subproblems 1 (CM) and 2 (RP). Problem **P'** is made equivalent to **P** (Guignard and Kim, 1987; Grossmann, 2021) after duplicating such variables and adding the constraints $x_e^1 = x_e^2 \; \forall e \in X$. Notice that the cost and constraints are also partitioned over the two subproblems. The reformulated problem **P'** is now amenable to Lagrangean decomposition.

$$\begin{aligned} z^*: = \max & \left\{\sum_{i=1}^{2} f_0^i(x^i, y^i)\right\} \\ \text{s.t. } & f_{m_i}^i(x^i, y^i) \le 0 \; \forall i \in \{1,2\}, m_i \in \{1, \dots, M\} \\ & x_e^1 - x_e^2 = 0 \; \forall e \in X \end{aligned} \tag{P'}$$

*2.2. Lagrangean relaxation of problem **P'***

A Lagrangean relaxation (Guignard, 2003; Guignard and Siwhan, 1987) $\mathbf{LR_\lambda}$ of problem **P'** is created by transferring each constraint $x_e^1 = x_e^2$ into the objective function, multiplied by a Lagrangean multiplier λ_e (unrestricted in sign).

$$\begin{aligned} z_\lambda^{LR*}: = \max & \left\{\sum_{i=1}^{2} f_0^i(x^i, y^i) + \sum_{e \in X} \lambda_e (x_e^1 - x_e^2)\right\} \\ \text{s.t. } & f_{m_i}^i(x^i, y^i) \le 0 \; \forall i \in \{1,2\}, m_i \in \{1, \dots, M\} \end{aligned} \tag{$\mathbf{LR_\lambda}$}$$

2.3. Decomposition of $\boldsymbol{LR_\lambda}$

For fixed values of λ_e, problem $\mathbf{LR_\lambda}$ can be decomposed into 2 parametric subproblems of type $\mathbf{LD_\lambda^i}$, which can be solved independently. The optimal value z_λ^{LD*} is equal to $z_\lambda^{1,LD*} + z_\lambda^{2,LD*}$ and provides an upper bound UB on the optimal value z^* of problem **P**. To obtain the tightest relaxation possible, λ_e is updated by means of an iterative procedure. Herein, we adopt the hybrid method by Grossmann and co-workers (Mouret et al., 2011; F Oliveira et al., 2013; Yang et al., 2020) for updating the Lagrange multipliers, which is based on a subgradient method, cutting plane approach, trust-region method and volume algorithm. Note that the solutions from these subproblems provide good quality initial points to solve the monolithic problem **P** as well.

$$z_{\lambda}^{i,LD*} := \max \left\{ f_0^i(x^i, y^i) + \sum_{j=i+1}^{2} \sum_{e \in X} \lambda_e x_e^i - \sum_{j=1}^{i-1} \sum_{e \in X_{ji}} \lambda_e x_e^i \right\} \quad (\mathbf{LD}_{\boldsymbol{\lambda}}^{\boldsymbol{i}})$$
$$\text{s.t. } f_{m_i}^i(x^i, y^i) \leq 0 \; \forall m_i \in \{1, \ldots, M\}$$

3. Case study

The refinery-petrochemical facility produces several grades of gasoline, diesel and fuel oil, and a set of petrochemical processes for providing BTX, polyethylene, propylene, waxes, and specialty solvents. These commodities mostly supply the Colombian market, with only a small part being exported. A domestic petroleum production equal to 297 kbbl/day is assumed, involving 17 types of crude oil distributed over 8 geographical regions. The refinery can also import 7 types of crude, with up to 15 kbbl/day per crude. The total refining capacity is 248 kbbl/day and the logistic system for crude and commodities comprises 4 river fleet routes and a system of 9 pipelines. The refinery-petrochemical facility is composed of 60 industrial plants, represented by about 125 models. Crude mixing and fuel blending is done in a tank farm, modelled as 30 additional units. The complete model of the system leads to a MIQCQP model with 6975 equations, 35104 nonlinear terms, 9592 continuous and 279 discrete variables. The linking variables between both subproblems in the Lagrangean decomposition algorithm correspond to the flowrate and qualities of the $u \in U^{CRB} = \{CB1, \ldots, CB8\}$ crude streams fed to the CDUs. Flowrates QF_u^{CM} and QF_u^{REF} are traded between CM and RP at the market price λ_u. Multipliers $\lambda_{u,p}$ are penalty costs associated with the $p = 1, \ldots, 3$ crude blend qualities $PF_{u,p}^{CM}$ and $PF_{u,p}^{REF}$. In total, 57 Lagrange multipliers distributed into flowrates (8), bulk properties (24), and crude blend composition (25) are considered.

We define a minimum throughput to the RP of 100 kbbl/day, and set default values for the Lagrange multipliers ($\lambda_u = 0, \lambda_{u,p} = 0$) at the start of the algorithm. At zero crude blends cost, the RP profit is about 11.4 MUSD/day, whilst CM loses 3.5 MUSD/day (the income from selling crude blends is zero), leading to an upper bound (UB) of 7.9 MUSD/day. CM chooses to make 100 kbbl/day of CB7, with 20 API, 1.13 %wt sulfur content and 2.70 mg KOH/g crude oil. These features indicate that CB7 is a heavy, sour, and acidic crude oil with poor qualities for processing at the RP. In contrast, and because it is not paying for the crude blends, RP chooses to include all eight crudes in the basket, leading to a total refinery capacity of 203 kbbl/day. The crude oil throughput to the RP has 32 API, 0.66 %wt sulfur content and 1.25 mg KOH/g crude oil, much more suitable to process at the CDUs than the crude blend produced by CM.

The values of the Lagrange multipliers are updated at each iteration (Figure 2) aiming at tightening the relaxation bound on **P**. The objective function value for the CM subproblem increases (not necessarily monotonically) until reaching its maximum at iteration 17. At that point, CM supplies 240 kbbl/day of a crude basket composed by crude blends CB3 (11%), CB4 (2%), CB5 (10%), CB6 (19%), and CB7 (58%). It can be viewed as a medium crude blend (26 API), with 0.96 %wt sulfur content and 1.92 mg KOH/g crude oil. Likewise, the objective function for the RP subproblem decreases because of the increase in the cost of the crude blends. At iteration 17, the RP throughput is about 137 kbbl/day, with CB1 (23%), CB2 (18%), CB3 (20%), CB5 (18%), CB6 (3%), CB7 (17%) and CB8 (2%). Compared to iteration 1, the RP requires a crude blend with a lower mg KOH/g crude oil (0.53) and keeps the same figures for API and sulfur content. From iterations 18 to 37, the changes in the objective function of CM and RP are steadier.

Overall, finding the optimal values for λ_u^* and $\lambda_{u,p}^*$ determines a transfer price between the crude management and the integrated refinery-petrochemical complex, leading to a trade-off between the CM incomes and REF outcomes.

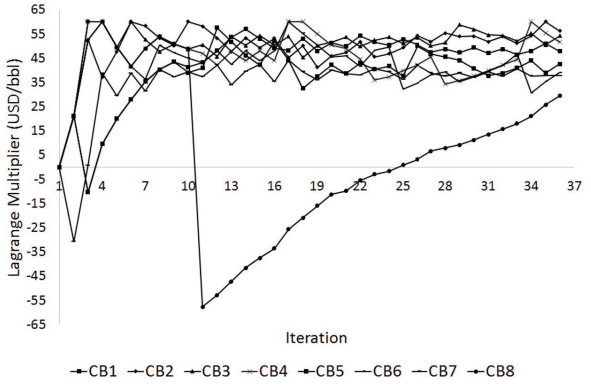


Figure 2. Lagrange multipliers associated with crude blend flowrates.

The performance of the spatial Lagrangean decomposition-based algorithm is summarized in Table 1. All the models used to address the short-term planning problem of the integrated refinery–petrochemical facility were implemented in GAMS 28.2. The CM subproblem was solved using ANTIGONE 1.1. The same is not true for the large-scale MIQCQP resulting from the RP subproblem. To address this challenge, the latter was solved by our process clustering decomposition approach (Uribe-Rodriguez et al., 2020), considering six clusters (crude distillation operations, vacuum and debutanizer columns, refinery units, petrochemical plants, fuels blending). Our Lagrangean decomposition-based approach reaches an optimal solution within 1 hour, with an optimality gap below 5%. Moreover, considering the lower bound reported by the clustering approach, the global optimal solution for this problem is enclosed between 2,964 kUSD/day and 3,063 kUSD/day.

Table 1. Performance comparison between global solvers

Approach	LB [kUSD/day]	UB [kUSD/day]	Gap [%]	Runtime [h]
Lagrangean Decomposition	2,932	3,063	4	0.8
Clustering Decomposition	2,964	3,205	8	5.7
ANTIGONE	2,634	3,898	48	10.0
BARON	2,687	4,250	58	10.0

4. Conclusions

This paper has presented a Lagrangean decomposition-based algorithm for large-scale MIQCQPs derived from the short-term planning problem of an existing integrated refinery-petrochemical facility. The high dimensionality makes it challenging to find high-quality solutions within a reasonable optimality gap. While all algorithms were able to identify feasible solutions, our approach outperforms the process clustering

decomposition, BARON and ANTIGONE, in terms of profit, optimality gap and computational runtime.

Acknowledgments

Financial support from the Colombian Science Council (COLCIENCIAS), the Center for Innovation and Technology Colombian Petroleum Institute (Ecopetrol S.A.) and Fundação para a Ciência e Tecnologia through grants CEECIND/00730/2017 and UIBD/04561/2020.

References

Yang, H., Bernal E. D., Franzoi, R. E., Engineer., F. G., Kwon, K., Lee, S., and Grossmann, I, E., 2020, Integration of crude-oil scheduling and refinery planning by Lagrangean decomposition, Computers and Chemical Engineering, 106812. 138,

Uribe-Rodriguez, A., Castro M. P., Guillén-Gosálbez, G., and Chachuat, B., , 2020, Global Optimization of Large-scale MIQCQPs via Cluster Decomposition: Application to Short–Term Planning of an Integrated Refinery–Petrochemical Complex, Computers and Chemical Engineering 140, 106883.

Sahinidis, N. V., 2004, BARON: A general purpose global optimization software package. J. Glob. Optim. 8, 201–205.

Misener, R., Floudas, C.A., 2014, ANTIGONE: Algorithms for coNTinuous / Integer Global Optimization of Nonlinear Equations. J. Glob. Optim. 59, 503–526.

Ketabchi, E., Mechleri, E., Arellano-Garcia, H., 2019. Increasing operational efficiency through the integration of an oil refinery and an ethylene production plant. Chem. Eng. Res. Des. 152, 85–94.

Li, J., Xiao, X., Boukouvala, F., Floudas, C.A., Zhao, B., Du, G., Su, X., Liu, H., 2016. Data-driven mathematical modeling and global optimization framework for entire petrochemical planning operations. AIChE J. 62, 3020–3040.

Zhao, H., Ierapetritou, M.G., Shah, N.K., Rong, G., 2017. Integrated model of refining and petrochemical plant for enterprise-wide optimization. Comput. Chem. Eng. 97, 194–207.

Pinto, J.M., Joly, M., Moro, L.F.L., 2000. Planning and scheduling models for refinery operations. Comput. Chem. Eng. 24, 2259–2276.

Guignard, M., Siwhan Kim, 1987. Lagrangean decomposition: A model yielding stronger Lagrangean bounds. Math. Program. 39, 215–228.

Grossmann, I. (2021). Advanced Optimization for Process Systems Engineering (Cambridge Series in Chemical Engineering). Cambridge: Cambridge University Press.

Mouret, S., Grossmann, I.E., Pestiaux, P., 2011. A new Lagrangian decomposition approach applied to the integration of refinery planning and crude-oil scheduling. Comput. Chem. Eng. 35, 2750–2766.

Oliveira, F, Gupta, V., Hamacher, S., Grossmann, I.E., 2013. A Lagrangean decomposition approach for oil supply chain investment planning under uncertainty with risk considerations. Comput. Chem. Eng. 50, 184–195.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50098-1

Green Ammonia Supply Chain Design for Maritime Transportation

Hanchu Wang[a], Prodromos Daoutidis[a*], Qi Zhang[a*]

[a] *Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN 55455, USA*

daout001@umn.edu, qizh@umn.edu

Abstract

Recently, there has been increased discussion about the potential of green ammonia as a carbon-free fuel for maritime transportation. If deployed at scale, the demand for ammonia from the shipping industry would be immense such that significant new investments had to be made in green ammonia infrastructure, including entire supply chains of new production sites and ammonia refueling ports. In this work, we develop an optimization model for the design of such a green ammonia supply chain. The proposed model integrates a large set of decisions, including the location of production plants and refueling ports, operational decisions related to green ammonia production using renewable energy, and ship routing decisions. This results in a complex mixed-integer linear programming formulation, which we apply to an illustrative case study to demonstrate its potential to address the given supply chain design problem.

Keywords: sustainability, green ammonia, maritime transportation, offshore wind

1. Introduction

Ammonia is one of the most produced commodity chemicals, and as the basis for most nitrogen fertilizers, it has been a key enabler for the sustained global population growth since the invention of the Haber-Bosch process. In recent years, ammonia has come under increased scrutiny due to the high carbon intensity of the conventional production process, and there have been various efforts in making its production more sustainable. One way to achieve this is to use renewable electricity to produce hydrogen via electrolysis and nitrogen via air separation and then react both chemicals to form ammonia. Ammonia that is produced in this manner is considered *green*. Green ammonia does not only have the potential to decarbonize fertilizer manufacturing, but can also be used as a carbon-free energy carrier [MacFarlane et al., 2020, Palys et al., 2021]. As such, it holds the promise to improve sustainability in multiple sectors. One exciting prospect is the use of ammonia as a marine fuel [de Vries, 2019]. The International Maritime Organization (IMO) has declared the goal of reducing the international shipping sector's annual greenhouse gas emissions by at least 50% compared to 2008 by 2050, which requires the use of alternative, less carbon-intensive fuels. Green ammonia is an ideal candidate as it does not cause any CO_2 emissions, neither in its production nor when it is combusted. In addition, it is sulfur-free, which ensures compliance with new IMO regulations.

In this work, we consider the design of green ammonia supply chains specifically for maritime transportation. For this purpose, we develop an optimization model that incorporates the production and distribution of green ammonia and its use as a marine fuel. Notably, in addition to ammonia produced on land, we also consider green ammonia that is produced on the open ocean using offshore wind. In our recent study [Wang et al., 2021], we have shown the techno-economic feasibility of such green offshore ammonia plants, which provide a means of harnessing wind energy far from the mainland. Here, such plants can further serve as offshore refueling stations, which would allow ships to carry less fuel when embarking on their trips across the ocean. This supply chain analysis aims to determine the optimal locations of ammonia production plants and refueling stations while considering the routing and scheduling of ship fleets as well as the production and storage of green ammonia. We propose a mixed-integer linear programming (MILP) formulation and apply it to an illustrative case study that demonstrates the main features of the model.

2. Mathematical Formulation

We consider a network consisting of a set of ports (including locations on the ocean) $\mathcal{M}$ and a set of ship fleets $\mathcal{V}$ where each fleet is defined by its origin and destination ports and the amount of cargo to be shipped. Each fleet can be split into multiple subfleets (or splits) that can travel on different routes. We consider a set of time periods $\mathcal{T}$ and adapt the arc-load continuous-time maritime inventory routing formulation proposd by Agra et al. [2017] to model the decisions associated with the routing of the subfleets. In the following, we briefly describe the main constraints of the proposed model.

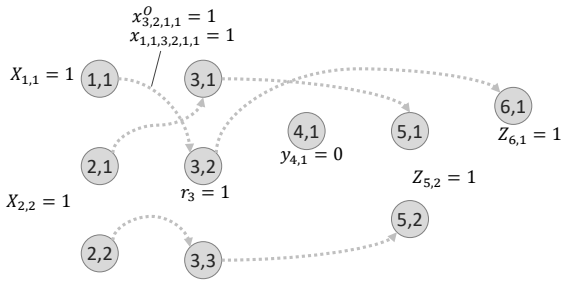


Figure 1: Illustration of the arc-load continuous-time formulation.

Figure 1 provides an illustration of the arc-load formulation. Here, each node is defined by a pair (i, m) representing port i visited the mth time (by any subfleet). We define a binary variable x_{imjnvk} that equals 1 if split k of ship fleet v travels from node (i, m) to node (j, n). Eq. (1) states that the sum of the split fractions, u_{vk}, for each fleet has to be 1. The binary variable h_{vk} equals 1 if fleet v is split. The binary variable w_{imvk} equals 1 if split k of fleet v visits port i at order m. Eqs. (3) and (4) state that each port visit and path traveled can be taken by one split only, and Eqs. (5)-(7) ensure that a port visit of a split only exits if the split exists, where z_{imvk} equals 1 if (i, m) is the terminal node visited by split k of fleet v. Eq. (8) ensures that all fleets travel from their origin ports to

their destination ports, where set $\mathcal{O}$ contains all origin and destination ports of fleet v.

$$\sum_{k\in\mathcal{K}} u_{vk} = 1 \qquad \forall\, v \in \mathcal{V} \quad (1)$$

$$h_{vk}U_v^{\min} \le u_{vk}U_v^{\text{tot}} \le h_{vk}U_v^{\max} \qquad \forall\, v \in \mathcal{V},\ k \in \mathcal{K} \quad (2)$$

$$\sum_{v\in\mathcal{V}}\sum_{k\in\mathcal{K}} w_{imvk} \le 1 \qquad \forall\, (i,m) \in S_v^A \quad (3)$$

$$\sum_{v\in\mathcal{V}}\sum_{k\in\mathcal{K}} x_{imjnvk} \le 1 \qquad \forall\, (i,m,v,k) \in S_v^X \quad (4)$$

$$\sum_{(i,m)\in S_v^A} w_{imvk} \ge h_{vk} \qquad \forall\, v \in \mathcal{V},\ k \in \mathcal{K} \quad (5)$$

$$\sum_{(j,m)\in S_v^A} x_{jmvk}^{\text{O}} = h_{vk} \qquad \forall\, i \in \mathcal{I},\ v \in \mathcal{V} : X_{iv} = 1,\ k \in \mathcal{K} \quad (6)$$

$$\sum_{m\in\mathcal{M}} z_{imvk} = h_{vk} \qquad \forall\, i \in \mathcal{I},\ v \in \mathcal{V} : Z_{iv} = 1,\ k \in \mathcal{K} \quad (7)$$

$$\sum_{m\in\mathcal{M}} w_{imvk} = h_{vk} \qquad \forall\, v \in \mathcal{V},\ (i,m) \in \mathcal{S}_v^A : (i,v) \in \mathcal{O},\ k \in \mathcal{K} \quad (8)$$

Eqs. (9)-(14) are additional routing constraints. The binary variable x_{jmvk}^{O} equals 1 if (j,m) is the first node visited by split k of fleet v, as stated in Eq. (9). Eqs. (10) and (11) are the flow conservation constraints. Eq. (12) ensures that a port visit is recorded by w_{imvk}, and Eq. (13) guarantees that at most one fleet visits node (i,m), where y_{im} equals 1 if port visit (i,m) is taken by a fleet, and the visiting order is constrained by Eq. (14).

$$\sum_{n\in\mathcal{M}} x_{injmvk} = x_{jmvk}^{\text{O}} \qquad \forall\, i \in \mathcal{I},\ v \in \mathcal{V} : X_{iv} = 1,\ k \in \mathcal{K} \quad (9)$$

$$w_{imvk} - \sum_{(j,n)\in\mathcal{S}_v^A} x_{jnimvk} = 0 \qquad \forall\, v \in \mathcal{V},\ (i,m) \in \mathcal{S}_v^A : X_{iv} = 0,\ k \in \mathcal{K} \quad (10)$$

$$w_{imvk} - \sum_{(j,n)\in\mathcal{S}_v^A} x_{imjnvk} = 0 \qquad \forall\, v \in \mathcal{V},\ (i,m) \in \mathcal{S}_v^A : Z_{iv} = 0,\ k \in \mathcal{K} \quad (11)$$

$$w_{imvk} \ge x_{imvk}^{O},\ w_{imvk} \ge z_{imvk} \qquad \forall\, v \in \mathcal{V},\ (i,m) \in \mathcal{S}_v^A :,\ k \in \mathcal{K} \quad (12)$$

$$\sum_{v\in\mathcal{V}}\sum_{k\in\mathcal{K}} w_{imvk} = y_{im} \qquad \forall\, (i,m) \in \mathcal{S}_v^A \quad (13)$$

$$y_{i,m-1} - y_{im} \ge 0 \qquad \forall\, (i,m) \in \mathcal{S}_v^A : m > 1 \quad (14)$$

Eqs. (15)-(19) are the ship fuel inventory constraints. Eqs. (15) and (16) compute the ship fuel level, l_{imvk}, for each subfleet when it departs from port visit (i,m). Here, $\gamma_v T_{ij}$ is the fuel consumption rate for fleet v that travels from port i to j, and $\bar{f}_{jnvk}$ is the total amount of fuel that split (v,k) receives from port visit (j,n). The binary variable g_{imvkt} equals 1 if split (v,k) visits (i,m) at time t; only then refueling can take place.

$$-\bar{C}_{vk}(1 - x_{imjnvk}) \le l_{imvk} + \bar{f}_{jnvk} - \gamma_v T_{ij} x_{imjnvk} u_{vk} - l_{jnvk} \le \bar{C}_{vk}(1 - x_{imjnvk}) \quad \forall\, (i,m,j,n) \in \mathcal{S}_v^X,\ k \in \mathcal{K} \quad (15)$$

$$-\bar{C}_{vk}(1 - x_{jnvk}^{O}) \le l_{vk}^{O} + \bar{f}_{jnvk} - \gamma_v T_{jv}^{O} x_{jnvk}^{O} u_{vk} - l_{jnvk} \le \bar{C}_{vk}(1 - x_{jnvk}^{O}) \quad \forall\, (j,n) \in \mathcal{S}_v^A,\ k \in \mathcal{K} \quad (16)$$

$$\bar{f}_{imvk} = \sum_{t\in\mathcal{T}} f_{imvkt} \qquad \forall\, v \in \mathcal{V},\ (i,m) \in \mathcal{S}_v^A,\ k \in \mathcal{K} \quad (17)$$

$$f_{imvkt} \le M^{big} g_{imvkt} \qquad \forall\, v \in \mathcal{V},\ (i,m) \in \mathcal{S}_v^A,\ t \in \mathcal{T},\ k \in \mathcal{K} \quad (18)$$

$$g_{imvkt} \le w_{imvk} \qquad \forall\, v \in \mathcal{V},\ (i,m) \in \mathcal{S}_v^A,\ t \in \mathcal{T},\ k \in \mathcal{K} \quad (19)$$

Eqs. (20)-(25) are the scheduling constraints where t^E_{im} and t^E_{imvk} denote the arrival and departure time of visit (i,m) for split (v,k), respectively. The parameters T^Q_i, T^S_i, T^O_{iv}, and $\bar{T}$ are the time required to load one unit of fuel at port i, the time to prepare to refuel, the time required to travel from the origin to port i, and the length of the time horizon, respectively. Eq. (20) states that the waiting time at port (i,m) can only be nonzero if the port is visited. Eq. (21) ensures that a fleet can only arrive at the next port after its previous port visit is completed. Eqs. (22) and (23) constrain the time for arrival and departure to and from port (i,m). The disjunction in Eq. (24) states that the arrival and departure times of two consecutive port visits are determined according to the visiting order.

$$t^E_{im} y_{im} \geq t^A_{im} + \sum_{v\in\mathcal{V}} \sum_{k\in\mathcal{K}} T^Q_i \bar{f}_{imvk} + \sum_{v\in\mathcal{V}} \sum_{k\in\mathcal{K}} T^S_i w_{imvk}(1-Z_{iv})(1-X_{iv}) \quad \forall\ (i,m) \in \mathcal{S}^A_v \quad (20)$$

$$t^E_{im} + T_{ij} - t^A_{jn} \leq 2\bar{T}(1 - x_{imjnvk}) \quad \forall\ v \in \mathcal{V},\ (i,m,j,n) \in \mathcal{S}^X_v,\ k \in \mathcal{K} \quad (21)$$

$$\sum_{v\in\mathcal{V}} \sum_{k\in\mathcal{K}} T^O_{iv} x^O_{imvk} \leq t^A_{im} \leq |\mathcal{H}| y_{im} \quad \forall\ (i,m) \in \mathcal{S}^A_v \quad (22)$$

$$t^E_{im} \leq |\mathcal{H}| y_{im} \quad \forall (i,m) \in S^A_v \quad (23)$$

$$\left|y_{im} + y_{i,m-1} \leq 1\right| \vee \left| \begin{array}{c} y_{im} = y_{i,m-1} = 1 \\ t^A_{i,m-1} \leq t^A_{im} \\ t^E_{i,m-1} \leq t^E_{im} \\ t^E_{i,m-1} + TB_i y_{im} \leq t^A_{im} \end{array} \right| \quad \forall\ (i,m) \in \mathcal{S}^A_v \quad (24)$$

$$\sum_{(i,m)\in S^A_v} g_{imvkt} \leq 1 \quad \forall\ v \in \mathcal{V}, k \in \mathcal{K},\ t \in \mathcal{T} \quad (25)$$

Eqs. (26)-(28) are the port inventory constraints. The ammonia inventory level at port i at time t is denoted by s_{it}, P_{it} is the production rate at port i in time period t, and q_{ijt} is the amount of ammonia transported from port i to j in time period t.

$$s_{it} = s_{i,t-1} + P_{it} - \sum_{v\in\mathcal{V}} \sum_{k\in\mathcal{K}} \sum_{m\in M} f_{imvkt} - \sum_{j\in\mathcal{J}} (q_{ijt} - q_{jit}) \quad \forall\ i \in \mathcal{I},\ t \in \mathcal{T} \quad (26)$$

$$s_{i,|\mathcal{T}|} \geq s_{i0} \quad \forall\ i \in \mathcal{I} \quad (27)$$

$$s_{i,t} \leq Q_i \leq Q^{\max}_i \quad \forall\ i \in \mathcal{I},\ t \in \mathcal{T} \quad (28)$$

Eqs. (29)-(32) are the ammonia production constraints. The binary variable o_{it} equals 1 if ammonia is produced at port i in time period t, p_i and r_i equal 1 if port i is an ammonia production site and refueling station, respectively. The production rate P_{it} is bounded by the ammonia plant's capacity C_i which is bounded by the maximum plant capacity. The production rate is a function of wind speed, ω_{it}, and the plant capacity.

$$o_{it} \leq p_i \quad \forall i \in \mathcal{I}, t \in \mathcal{T} \quad (29)$$

$$0 \leq P_{it} \leq C_i \leq C^{\max}_i o_{it} \quad \forall i \in \mathcal{I}, t \in \mathcal{T} \quad (30)$$

$$P_{it} = f(\omega_{it}, C_i, o_{it}) \quad \forall i \in \mathcal{I}, t \in \mathcal{T} \quad (31)$$

$$q_{ijt} \leq p_i,\ q_{ijt} \leq r_j \quad \forall\ i \in \mathcal{I},\ j \in \mathcal{I},\ t \in \mathcal{T} \quad (32)$$

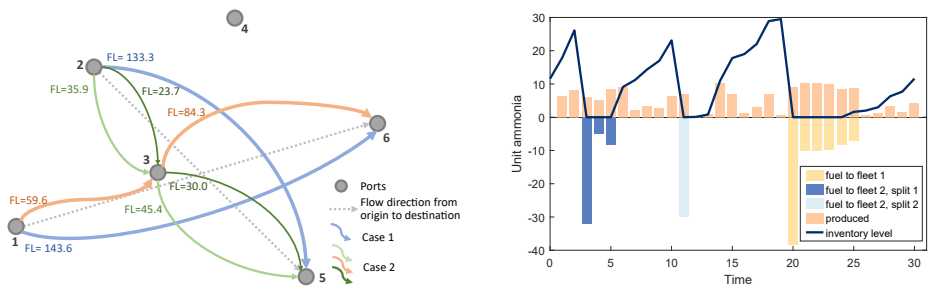
Finally, the disjunction in Eq. (33) links the routing constraints with the production and inventory constraints. The overall problem can be formulated as an MILP.

$$\begin{bmatrix} t^A_{im} \le t \le t^E_{im} \\ w_{imvk} = 1 \\ g_{imvk} = 1 \end{bmatrix} \vee \begin{bmatrix} \neg \begin{bmatrix} t^A_{im} \le t \le t^E_{im} \\ w_{imvk} = 1 \\ g_{imvk} = 1 \end{bmatrix} \end{bmatrix} \qquad \forall (i,m) \in S^A_v,\ k \in \mathcal{K} \quad (33)$$

3. Computational Case Study

The objective is to minimize the overall cost, which includes both the capital and operating costs. The capital cost includes the cost of constructing the green ammonia production plants and refueling ports while the operating cost consists of the costs for producing, storing, and transporting ammonia as well as the cost of operating the ship fleets. In our case study, we consider two proxies for the shipping cost: travel time and fuel cost.

As shown in Figure 2a, we consider a network of six ports and two ship fleets, where one fleet has to transport cargo from Port 1 to Port 6 while the other one has to travel from Port 2 to Port 5. Ammonia can be produced at every port; however, Port 3 is assumed to be a location on the open ocean where wind speeds are significantly higher and steadier [Possner and Caldeira, 2017]. As a result, the cost of producing ammonia at Port 3 is lower than at the other locations. With a time horizon of 30 days (each day being one time period), we optimize the system for two cases. In Case 1, we apply a penalty on the total travel time, assuming that the operating cost of shipping and the opportunity cost of delivering more cargo mainly depend on the travel time. In Case 2, we only consider fuel cost as a proxy for shipping cost.



(a) Optimal routes for Cases 1 and 2

(b) Optimal ammonia schedule for Case 2

Figure 2: (a) Optimal routing decisions in Cases 1 and 2. Ports are depicted as grey nodes, grey dashed lines indicate the origin and destination ports. The blue arcs depict the optimal routes in Case 1 while the green and orange arcs represent the optimal routes in Case 2. (b) Ammonia production and refueling schedule and the resulting inventory profile at Port 3 in Case 2. Positive values indicate production while negative values indicate consumption.

Figure 2a shows the optimal routing decisions in both cases. In Case 1, the optimal route for each fleet is to simply go directly from its origin to its destination (blue arcs), which indicates that the shipping cost as a function of the travel time outweighs the benefit of

producing ammonia at a lower cost at Port 3. In Case 2, the optimal route for each fleet involves stopping at Port 3 to refuel since part of the objective is to reduce fuel cost. Also, in this case, we see that all cargo is transported from Port 1 to Port 6 using one trip (orange arc), while two trips are required to transport all the cargo from Port 2 to Port 5 (green arcs). The reason is that, to reduce capital cost, the optimal solution suggests building an ammonia plant at Port 3 with a capacity that is not large enough to fuel the entire fleet that travels from Port 2 to Port 5. Hence, the fleet has to be split into two subfleets that arrive at Port 3 for refueling at different times. Figure 2b shows the optimal ammonia production and refueling schedule at Port 3. One can see that the three subfleets arrive at the port for refueling at different times such that there is enough time between two visits to replenish the ammonia inventory.

4. Conclusions

In this work, we considered the design of supply chains for the production of green ammonia and its use as a marine fuel in the shipping industry. This required the integration of maritime inventory routing and production scheduling decisions into a facility location formulation, which gives rise to a complex MILP model. We applied the proposed model to an illustrative case study, which demonstrates the model's ability to achieve solutions that optimally balance the costs of investment, shipping, and ammonia production. Future work will focus on the development of efficient solution algorithms that allow us to solve large-scale instances and draw conclusions about the potential benefit of green ammonia as a sustainable marine fuel at the industrial scale.

References

A. Agra, M. Christiansen, and A. Delgado, 2017. Discrete time and continuous time formulations for a short sea inventory routing problem. *Optimization and Engineering*, 18(1):269–297.

N. de Vries. *Safe and effective application of ammonia as a marine fuel*. Master thesis, TU Delft, 2019.

D. R. MacFarlane, P. V. Cherepanov, J. Choi, B. H. Suryanto, R. Y. Hodgetts, J. M. Bakker, F. M. Ferrero Vallana, and A. N. Simonov, 2020. A roadmap to the ammonia economy. *Joule*, 4(6):1186–1205.

M. J. Palys, H. Wang, Q. Zhang, and P. Daoutidis, 2021. Renewable ammonia for sustainable energy and agriculture: vision and systems engineering opportunities. *Current Opinion in Chemical Engineering*, 31:100667.

A. Possner and K. Caldeira, 2017. Geophysical potential for wind energy over the open oceans. *Proceedings of the National Academy of Sciences of the United States of America*, 114(43):11338–11343.

H. Wang, P. Daoutidis, and Q. Zhang, 2021. Harnessing the wind power of the ocean with green offshore ammonia. *ACS Sustainable Chemistry & Engineering*, 9(43):14605–14617.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50099-3

Optimal agriculture residues revalorization as a biofuel alternative in electric power grids

Sergio Iván Martínez-Guido [a], Juan Fernando García-Trejo [a], Claudia Gutiérrez-Antonio [a*], and Fernando Israel Gómez-Castro [b]

[a]*Engineering Faculty, Universidad Autónoma de Querétaro, Campus Amazcala, El Marqués, Querétaro, 76265, Mexico*
[b]*Chemical Engineering Department, Universidad de Guanajuato, Guanajuato, Guanajuato, 36050, Mexico*
claugtez@gmail.com

Abstract

In the last years, the increase in energy demand along with depletion in the production of oil wells has driven the search of alternative energy sources. In addition, these sources must be renewable, in order to contribute to the solution of the climate change problem. Therefore, different alternatives have been proposed and evaluated to satisfy global energy demand. Among them, biomass has received higher attention due its availability, which places it as the third energy source, after crude and oil fuels. Particularly, the agricultural residues represent an option to obtain different biofuels, with the possibility to reduce between 78-94% greenhouse emissions by its use. Moreover, the integration of this kind of residues in new supply chains economies contributes to its valorisation, adding an economic benefit to its origin crop. Thus, it is necessary to evaluate the feasibility of using these biofuels, along with their environmental impact. Hence, this work was optimized the supply chain for the production of solid fuels from different agriculture residues along with its integration into the electric power grid; as a case of study, it was considered the agriculture residues generated in Mexico and its electric power grid. Results show that is possible to generate 140,673,599 GJ, and at the same time achieve a reduction of 20% of CO_2 emissions.

Keywords: Solid biofuels, optimization, supply chain, environmental impact.

1. Introduction

Due to the current increase in energy demand around the world, different bio-energy sources have received higher attention, being biomass waste one of the most promising sources (Schawarzböck et al., 2016). In this way, agro-industrial sector contributes different kinds of residual biomass, which are derived from all the operations or activities involved in the supply chains involved. In this context, agriculture industrial waste can be classified into four main groups (Kumar-Chhetri et al., 2020):

- Crop cultivation waste (stalk, silk, stem, seed, pod, root, weeds, etc.)
- Food loss/waste (edible and non-edible or unprocessed food, spoiled food, etc.)
- Industrial food processing waste (husk, seed, bagasse, fruit, peel, rind, etc.)
- Livestock rising aqua/microbial culture waste (excreta and carcasses dead cell, waste-water, etc.)

Particularly, crop cultivation waste impacts the environment due to these residues are mostly burned in the fields or put into landfills, being used only a lower percent as fodder.

Therefore, some works have focused on finding the optimal alternative to add value to these kinds of residues; highlighting that these kinds of residues are a promising raw material in biofuels processes production (Martínez-Guido et al., 2019). According to Debnath & Babu (2019) biofuels can be produced from different types and forms of biomass and residual biomass. Particularly, agricultural residues have an attractive potential to be used raw material in pellets (solid biofuel) production supply chains (Tauro et al., 2018), incorporating the advantages as overcome the issue with a lower energy per volume unit, easy management, transportation, handling cost, storage, and a better economic efficiency relation in comparison with liquid biofuels (Theppitak et al., 2020). Therefore, the pellets are part of the current renewable alternative sources for electric generation (Sandberg et al., 2019); however, nowadays only a few studies have been focused on the use of pellets obtained from agricultural and into electric power grids, incorporating this process as an agriculture residues revalorization strategy (Martínez-Guido et al., 2019). Hence, in this work is proposed the use of different agricultural residues (obtained from diverse Mexican crops), which nowadays doesn't have any use of the market, as raw material for pellets production. Moreover, the possible supply chain configuration to use these solid biofuels into the electric power grid of México is also analysed.

2. Problem statement

According to Kaza et al. (2018), around the world are generated almost 2.01 billion tonnes/y of solid wastes, and it is expected that this volume increase by 40% by 2050. Particularly, Mexico contributes with approximately 72,172 tonnes of agricultural residues per year (ANBIO, 2020); this residual biomass is constituted by chemical compounds of high value as cellulose, hemicellulose and lignin (Santiago del Rosa et al., 2018). However, as Pradhan et al. (2019) indicated these kinds of residues are left in the fields or even burned, mainly in developing countries. Therefore, the use of these residues as raw material for the pellets production could be a strategy with diverse benefits, such as the dependence on oils reduction, fuels with lower emissions, avoid the scarcity of conventional fuels, and, at the same time, to promote the circular economy. The goal in agriculture residues pelletization is to produce a cheaper and sustainable biofuel with the capacity to satisfy the Mexican electric energy demand. However, is not possible to take into consideration the pellets integrations in all the national electric power plants, due to the available technology in each electric plant; thus, for the analysis only 62% of national demand was considered. Hence, the design, planning and scheduling of a new supply chain for the integration of these kinds of residues into the electric power grid involving the economic, social, and environmental aspects is required (See Figure 1). Therefore, the optimal design could offer a win-win situation where both agriculture residues management and energy security issues are achieved, which is demonstrated through a case study of the electricity supply chain in Mexico.

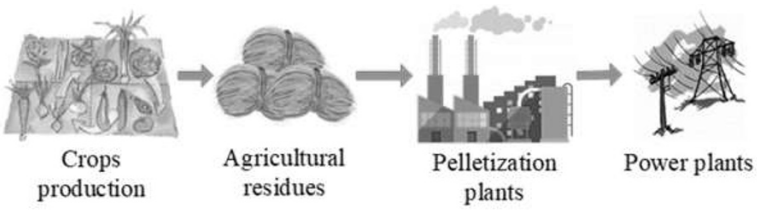


Figure 1. Agro-residues revalorization.

3. Methodology

As it was mentioned before, the proposed approach was performed using the current situation of the Mexican electrical energy demand as a case study. In this way, agricultural residues from the 18 crops (i), shown in Table 1, were analyzed in terms of seasonal availability and flux production in the 32 states (k) that constitute Mexico; since it would be wrong to take into consideration a constant flow of residues throughout the year.

Table 1. Agriculture residues considering in this work

Agave	Broad Beans	Coffee bean	Kidney beans	Peanut	Sorghum
Banana	Chickpea	Cotton	Lentil	Rice	Sugar Cane
Barley	Cocoa	Fodder	Oats	Sesame	Wheat

Hence, using Eq. (1) was possible to calculate the flow of residues of each crop produced $\left(FR_{i,k,w}^{AGr}\right)$ per week (w), biomass flux is directly linked with the respective parameter, yield of ton of residue per a ton of crop cultivated $\left(P_{i,k,w}^{AGry}\right)$ and the area used for the harvesting $\left(A_{i,k,w}^{harv}\right)$.

$$FR_{i,k,w}^{AGr} = A_{i,k,w}^{harv} * P_{i,k,w}^{AGry}, \forall\, i \in I, k \in K, w \in W \tag{1}$$

Considering the agriculture residues availability, these can be collected in each harvesting site and transported to the pelletization plant. The installation of the pelletization plant is considered for each state in Mexico; however, the decision of the existence is constrained by the installation cost, and the residues availability in each state, a binary variable was used to identify in which states are installed a pelletization plant. This binary variable $\left(Y_{i,m}^{bv}\right)$ is multiplied by the fixed installation cost $\left(IC_{i,m}^{fix}\right)$, a value that added to the variable cost $\left(IC_{i,m}^{var}\right)$ multiplied by the process flow capacity $\left(RF_{i,m}^{proc}\right)$, gives the total installation cost of each pelletization plant $\left(TI_{i,m}^{cost}\right)$.

$$TI_{i,m}^{cost} = IC_{i,m}^{fix} * Y_{i,m}^{bv} + IC_{i,m}^{var} * RF_{i,m}^{proc}, \forall\, i \in I,\ m \in M \tag{2}$$

With the flux of pellets $\left(Pe_{i,j,w}^{flo}\right)$ obtained in each installed plant, is possible to produce electricity in each selected power plant, and satisfy the electrical energy demand in each state $\left(EE_{j,w}^{prod}\right)$; however, if a state doesn't have enough solid biofuel, hence the conventional fuel $\left(CF_{l,j,w}^{flo}\right)$ is used to satisfy the required rest, as is described by Eq. (3).

$$EE_{j,w}^{prod} = \sum_i Pe_{i,j,w}^{flo} * FE_i^{gene} + \sum_l CF_{l,j,w}^{flo} * FE_l^{gene}, \forall\, j \in J, w \in W \tag{3}$$

In order to obtain the optimal supply chain configurations, the proposed mathematical model included residues collection, transportation, transformation, and pellet distribution balances as mathematic equations. Additionally, equations as economic balances and emissions balances were proposed. Consequently, optimal solutions were obtained having as multi-objective function the CO_2 emissions and total cost minimization, and the social impact maximization, as is described in Eq. (4).

$$OF = Min\ CO_2^{Emis}, Min\ TOTC, Max\ TSI \tag{4}$$

Therefore, multi-objective problem was codified at GAMS platform and solved using the ε constraint method (Diwekar, 2010), in which one of the objective functions is optimized

using the others objectives functions as constrains, leading to finding the extreme solution at the feasible region. Afterward, the rest of the objective functions are optimizing under the same criteria. In this way, were analyzed five different scenarios, for each solution (scenario) the model calculates 608 discrete variables, 513,774 continuous variables under 68,718 constraints. The CPU execution time per assay was 2.265 seconds in an Intel 336 processor running at 2.40 GHz with 8 GB of RAM memory, using CPLEX as the solver.

4. Results

Mainly results obtained from the five scenarios analyzed are shown in Table 2. Particularly, scenario A represents the solution with the lower amount of CO_2 emissions in comparison with the rest of the solutions, even in comparison with the current situation of electricity production in Mexico. In this way, according to SENER (2016) the supply of 62% of national electric energy demand in Mexico (using only conventional fuels as energy sources) involves the release of 114,911,336 ton CO_2/y; hence, solution A represents 22.15% of emissions reduction, while the emissions in solution E are 6.58% lower in comparison with the current situation. Therefore, a reduction from 7,561,165-25,452,860 ton CO_2/y is possible to achieve if the agriculture residues revalorization is carried out. However, to achieve any of these percent's of emissions reduction is required money investment. As it is possible to notice in Table 2, scenario A needs the major investment (approximately 428,734,538 USD); nevertheless, for all the obtained solutions 80% of the total cost is given by the production cost, which is constituted by the pelletization plant installation and the transformation costs. In this way, the highest cost makes sense due that currently there are not any pelletization plant installed in Mexico; moreover, in solution A 32 new plants are installed to satisfy 62% of the national electrical energy demand. The 20% of the total money invested is given by transportation costs and taxes added to the new economic activity.

Table 2. Mainly results obtained in the different scenarios analyzed

		Difference with scenario A (%)			
	A	**B**	**C**	**D**	**E**
Invested money (USD/y)	428,734,538	-1.525	-4.261	-4.444	-4.574
Human development index (% of invested money)	12.96	-2.022	-6.420	-9.568	-11.574
Emissions (ton CO_2/y)	89,451,365	+7.69	+12.63	+16.29	+20.01

On the other hand, as it is shown in Table 2, the installation of 32 new industrial plants focused on solid biofuel production (scenario A) requires the higher investment; nonetheless, in this scenario is obtained a social benefit, which is represented by the money invested in social security, education and per-capita income variables which has a direct impact on the human development index. In this way, due that a new economic activity is developed; hence, new employees and as consequence new salaries are created, salaries to which taxes are deducted. Additionally, a taxes deduction over the new implanted industries is considered. So, from all the deducted taxes is possible to take into consideration a new national flux of money invested to human development activities; in this context, scenario A achieves 12.96% more invested money than that considered in the current national situation, while scenario E is possible to achieve only 1.38% more invested money.

Table 3 shown the economic comparison with the current conventional fuels used in the Mexican power grid; as it is possible to notice, each tone of pellets is 38% more expensive than the carbon prices (the cheaper conventional fuel alternative). However, in comparison with the Gas LP (the expensive conventional fuel alternative) and fossil diesel, each tone of pellets is 96% and 95% cheaper, respectively. In this way, according to Olguin-Muciel et al. (2020) the biofuel's potential is directly linked with the cost-competitively with conventional fuels; hence, this argument is achieved when pellets are compared with diesel and gas but is not reached with carbon. On the other hand, CO_2 emissions per MWh generated by the pellets are 20% higher than emissions released by the use of gas LP (conventional fuel with lower CO_2 emissions); but pellets emissions are 36% and 27% lower in comparison with the emissions generated by the use of diesel (conventional fuel with higher CO_2 emissions) and carbon, respectively. Nevertheless, emissions listed in Table 3 for conventional fuels do not take into consideration the CO_2 emitted by the extraction of the fuels, only are quantified the emissions released by the combustion (use); the reported CO_2 emissions by the use of pellets take into consideration all the emissions generated in the complete life cycle.

Table 3. Economic and emissions comparison

Energy source	USD/ton	CO_2 ton/MWh produced
Carbon	34	0.603
Diesel	1,275	0.692
Gas	1,398	0.352
Pellets	55	0.44

In the CO_2 emissions context, transportation activities in the pellets supply chain represent the process with the higher released emissions. In solution A, biomass transportation and pellets distribution contribute with 88% of the total generated emissions, while in scenario E these activities contribute with 68% of total emissions. Similarly, the pellets production release 10% of the total emissions in scenario A, while in scenario E this activity contributes with only 7%. However, in scenario E 20% of the energy sources used to satisfy the 62% of electrical energy demand comes from conventional fuels; therefore, 24% of the total emissions in this solution are given by the combustion of conventional fuels, while in scenario A the 62% of electricity national demand is satisfied only with pellets.

5. Conclusions

Agriculture residues revalorization throughout pellets production is a green alternative that can be used in the Mexican electricity grid, satisfying almost 62% of national electricity demand. Additionally, residues transformation into solid biofuel is an alternative to reduce dependence on fossil fuels under a circular economy development, having beneficial impact on the human development index. Fuel pellets production supply chain is a complex system due to all the activities involved to achieve the process goal; hence, the use of the mathematical model approach allows to obtain different system configurations, which results in attractive solutions from the economic, social, and environmental points of view.

Acknowledgements

Financial support provided by Secretaría de Educación Pública (SEP), through grant PRODEP-UAQ/332/19, for Sergio Iván Martínez-Guido is gratefully acknowledged.

References

ANBIO-National Atlas of Biomass, 2018, Atlas Nacional de Biomasa. Available on: https://dgel.energia.gob.mx/ANBIO/

D. Debnath, S. Babu, 2019, Biofuels, Bioenergy andk Food Security; Technology, Institutions and Policies, first ed., ELSEVIER, London, United Kindong

E. Olguin-Muciel, A. Singh, R. Cable-Villacis, R. Tapia-Tussell, H.A. Ruiz, 2020. Consolidated Bioprocessing, an Innovative Strategy towards Sustainability for Biofuels Production from Crop Residues: An Overview, Agronomy, 10, 1834

E. Sandberg, J.G. Kirkerud, E. Tømborg, T.F. Bolkejø, 2019, Energy system impacts of grid tariff structures for flexible power to district heat, Energy, 168, 772-781

N. Santiago-De La Rosa, G. González-Cardoso, J de J. Figueroa-Lara, M. Gutiérrez-Arzaluz, C. Octaviano-Villasana, I.F. Ramírez-Hernández, V. Mugica-Álvarez, 2018, Emission factors of atmospheric and climatic pollutants from crop residues burning, Journal of the Air & Waste Management Association, 68, 849-865

P. Pradhan, P. Gadkari, A. Arora, S.M. Mahajani, 2019, Economic feasibility of agro waste pelletization as an energy option in rural India, Energy Procedia, 158, 3405-3410

R. Kumar-Chhetri, N. Aryal, S. Kharel, R. Chandra-Poudel, D. Pant, 2020, Current Developments in Biotechnology and Bioengineering: Sustainable Bioresources for the Emerging Bioeconomy, first ed., ELSEVIER, Amsterdam, Netherlands (Chapter 5)

R. Tauro, C.A. García, M. Skutsch, O. Masera, 2018, The potential for sustainable biomass pellets in México: An analysis of energy potential, logistic cost and market demand, Renewable and Sustainable Energy Reviews, 82, 380-389

S. Kaza, L. Yao, P. Bhada-Tata, F. Van-Woerden, 2018, What a waste 2.0: A Global snapshot of solid waste management to 2050, World Bank Group, Available on: https://openknowledge.worldbank.org/handle/10986/30317

S. Theppitak, D. Hungwe, L. Ding, D. Xin, G. Yu, K. Yoshikawa, 2020, Comparison on solid biofuel production from wet and dry carbonization processes of food wastes, Applied Energy, 272, 115264

S.I. Martínez-Guido, I.M. Ríos-Badrán, C. Gutiérrez-Antonio, J.M. Ponce-Ortega, 2019, Strategic planning for the use of waste biomass pellets in Mexican power plants, Renewable Energy 130, 622-632

SENER "Energy Secretary" México, 2016, Development program of the national energy system. https://www.gob.mx/cms/uploads/attachment/file/98308/PRODESEN-2016-2030_1.pdf, (accessed on February 2020).

T. Schwarzböck, H. Rechberger, O. Cenic, J. Fellner, 2016, Determining national greenhouse gas emissions from waste-to-energy using the Balance Method, Waste Management, 49, 263-271

U. Diwekar, 2010, Introduction to Applied Optimization, Second ed, Springer, 10.1007/978-0-387-76635-5

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50100-7

Global Supply Chain Optimization for COVID-19 Vaccine under COVAX initiative

Katragadda Apoorva[a], IA Karimi[a]* and Xiaonan WANG[b]*
[a]*Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore 117585 Singapore*
[b] *Department of Chemical Engineering, Tsinghua University, Beijing 100084, China*
wangxiaonan@mail.tsinghua.edu.cn

Abstract

After the onset of the COVID-19 pandemic, World Health Organization (WHO) launched COVAX in April 2020 to bring together countries and vaccine manufacturers and provide innovative and equitable access to COVID-19 vaccines. We developed a global supply chain model to optimize the production of the vaccines and allocation to different countries worldwide. The global COVID-19 vaccine supply chain of COVAX should be resilient to various disruptions and risks. In this work, we develop an optimization model with risk mitigation strategies to determine the procurement of vaccines from production centers and distribution to different countries when the supply chain is subjected to various disruptions. Our case study demonstrates how different risk mitigation strategies would enable COVAX to meet the demand amid multiple disruptions. It indicates that it is feasible to meet the vaccine demand and help participating countries overcome the global pandemic.

Keywords: add three to five keywords here.

1. Introduction

The only way to overcome the ongoing pandemic and achieve herd immunity is to vaccinate people worldwide. After the accelerated vaccine development, several vaccines have been approved globally. However, to overcome the challenge of coordinating the procurement and distribution of vaccines globally, the COVAX initiative was set up. It is co-led by WHO, Gavi (an organization that works towards getting people from developing countries vaccinated), and Coalition for Epidemic Preparedness Innovations (CEPI), a Gates Foundation-funded project that aims to make more vaccines available during an outbreak, was setup. Under the COVAX initiative, funding from rich countries was supposed to pool to invest in multiple vaccine development to increase the chances of effective vaccine development and approval. In addition, the funding was intended to provide vaccines for poorer countries. COVAX initiative handles the procurement and distribution of COVID-19 vaccines without discriminating between participating countries based on income [1] (Figure 1). However, after the approval of several COVID-19 vaccines, the global delivery has been successful but limited to rich countries. Due to unprecedented demand for vaccines, there has been a shortage of vaccines to the COVAX initiative as the rich countries procured vaccines through bilateral deals. [2]. COVAX deliveries have been hindered but slowly accelerating. COVAX has been making efforts to address and mitigate various risks. Therefore, optimizing the supply chain under

various uncertain disruptions is crucial to minimize the overall cost and utilize the resources best.

Stochastic supply chain optimization has been studied to optimize the supply chains of vaccines in the past. However, a recent study [3] reported a lack of existing academic publications on vaccine supply chain resilience. Few papers have been published partially addressing the challenges for the COVID-19 vaccine supply chain [4,5]. To account for the potential disruptions of vaccine production and delivery, we establish a stochastic model to capture different scenarios and minimize the overall cost. The objective function also has a term quantifying the supply chain's resilience. This work first provides insights into various setbacks faced by the COVAX facility in section 2. Section 3 describes the mathematical model in detail, and the solution for optimistic deterministic cases and scenarios with various disruptions are presented in section 4. Finally, the conclusion and future work is mentioned in Section 5.

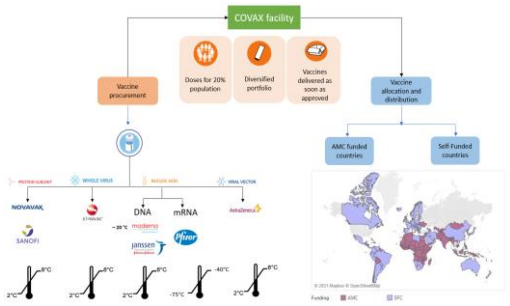


Figure 1. The schematic of the COVAX facility involves vaccine procurement and global distribution.

2. Challenges faced by COVAX

To achieve the objectives of COVAX, they made investments for the development of several vaccines during the development phase and then signed deals with various manufacturers to deliver vaccines once approved. Once the vaccines are procured, COVAX is responsible for transporting the vaccines from the manufacturing center to the Countries' central hub. However, even after the successful approval of the vaccines, COVAX has failed to meet its objectives. Since its launch, it has been subjected to various setbacks, and the most crucial ones are as follows: 1) Production facilities are not adhering to the promised doses. For example, COVAX depended mainly on the Serum Institute in India for vaccines for 2021 (~50%). The second wave of COVID-19 in India resulted in a halt in the supply of doses to COVAX [6], 2) Lack of transportation capacity and various transportation disruptions: With the unprecedented demand for vaccines, available transportation capacity is not enough. In addition, when the rich countries came forward for donation, the freight and storage had to be taken care of by the COVAX [7], 3) Vaccine nationalism: Rich countries have procured the vaccines, and none are left for the low- and middle-income countries under the COVAX facility. After 18 months of the

launch of the COVAX initiative, 98% of people in low-income countries remain unvaccinated [8].

3. Problem Statement

This paper considers a two-echelon production-distribution network for vaccine distribution under COVAX. The COVID-19 vaccine delivery from the manufacturer to the country's central hub is carried through direct shipment to the point of use without any distribution centers in the middle. As part of the initiative, there are various manufacturers m, manufacturing set of vaccines v and delivering them to different countries c in the world through transportation links (TLs) using storage containers type s based on the storage requirements of vaccines. The risk mitigation strategies for different disruptions are as follows: 1) In case of a risk involved with the production facility, we employ two mitigation strategies: a) More investment should be made in scaling up the production capacity. If the production capacity is reduced, investment should be made to restore the capacity for the next round of allocation. In addition, instead of ordering the exact amount initially, COVAX should have a deal for extra doses as a buffer, b) Streamline the donation process: Coordinate with rich countries and plan the delivery of the vaccines. Rich countries are wasting several million unused doses. COVAX should tap into these countries and sign deals for donations. In other words, putting efforts to connect donating countries (backup nodes) when the primary manufacturer is not available. These mitigation strategies will help deal with the above-mentioned (Section 2) challenges with the procurement of doses. During the disruption in the cold chain transportation of the vaccines or sudden requirement of transportation fleet when a country agrees to donate, the COVAX facility should employee backup, i.e., 3PL, to take care of the vaccine distribution. This would ensure that the facility is prepared for the transportation of vaccines without any capacity constraints, would also enable that no vaccines are damaged during transport, and ensure that the donations are well-received.

4. Supply chain optimization

Decision-making in vaccine production and allocation is formulated as a mixed-integer linear programming (MILP) model. The objective functions and constraints are mentioned in table 1. The objective function is divided into two parts: one corresponding to minimizing the design costs based on pre-disruption decisions, and the second is the cost during expected worst-case after the realization of the disruptions. The worst case is incorporated in our model via the conditional Value at Risk (cVaR) measure. Stochastic optimization aims to minimize the total cost. Pre-disruptions design costs include establishing transportation links with the fleet and investment in procurement, such as signing more deals with donors to scale up the production capacity and contracting 3PL as backup transportation. The second part of the objective function is the expected worst-case cost, including transportation costs, storage costs, vaccine procurement costs, and recovery costs for restoring production and transportation capacities after disruptions. The optimization constraint includes capacity constraints, supply-demand balance, and other logical constraints. (Eq 6 – 9, Table 1).

5. Case study

We have considered the demand for nine countries (India, Pakistan, Nigeria, Mexico, Ethiopia, Egypt, DRC, Iran, and Thailand), 50% of COVAX demand. The vaccine portfolio of 9 vaccines is also considered. These vaccines broadly fall into four categories (Figure 1). The vaccines differ in price per dose, the number of doses per person, storage requirement, and production sites where each vaccine is produced. The production centers are fixed based on the deals with the vaccine manufacturers. Various costs such as storage cost, transportation per unit distance, and selling price of vaccines are fixed and taken from literature. However, few costs such as contracting 3PL, cost of restoration of production capacity, and setting up a contract between a production center and a country for vaccine delivery are assumed. Through our case study, we want to demonstrate the effectiveness of our model over the basic model employed by the COVAX facility currently. The production and transportation capacities are subjected to various disruptions and are expressed in the percentage of available capacity. We define four

Table 1. Mathematical model formulation as designed for the optimization problem

	Index	Mathematical Formulation	Description
Objective function			
	(1)	$\sum_{p}\sum_{c}\sum_{m} C_{p,c,m}^{install} Z_{p,c,m}$	Design Costs
	(2)	$\sum_{p} C_{p}^{extra\ capacity} w_p + \sum_{c} C_{c}^{contract} Z_{c}^{contract}$	Resilience Enhancing Investment
	(4)	$\sum_{m} C_{v}^{vaccine}\left(\sum_{c}\left(\sum_{p}\sum_{t}\sum_{m} C_{v}^{vaccine} y_{v,p,c,t,m}^{s} F_{c}^{AMC}\right)\right)$	Cost of vaccines
	(5)	$\sum_{v}\sum_{p}\sum_{c}\sum_{t}\sum_{m} tr_{v,p,c,m} y_{v,p,c,t,m}^{s} + \sum_{v}\sum_{p}\sum_{c}\sum_{t}\sum_{m} C_{m}^{storage} y_{v,p,c,t,m}^{s}$	Cost of transportation and storage
	(6)	$\sum_{m}\sum_{p}\sum_{c}\sum_{t} C_{p}^{restore} rp_{p,t,v}^{s}$	Cost of transportation by 3PL
Constraints			
	(6)	$y_{v,p,c,t,m}^{s} \leq M Z_{p,c,m}$	Logical constraints
	(7)	$\sum_{c}\left[\sum_{m}(y_{v,p,c,t,m}^{s}) + (yb_{v,p,c,t}^{s} + rb_{v,p,c,t}^{s})\right] \leq x_{v,p,t}^{s}$	Balance constraints
	(8)	$x_{v,p,t}^{s} \leq M * VP(v,p)$	Vaccine-production matching
	(9)	$D_{c,t} - \sum_{p}\sum_{m}\sum_{v}\frac{y_{v,p,c,t,m}^{s}}{Dose_v} - \sum_{p}\frac{yb_{v,p,c,t}^{s}}{Dose_v} - \sum_{p}\sum_{m}\sum_{v}\frac{rb_{v,p,c,t}^{s}}{Dose_v} = q_{c,t}^{s}$	Overall Supply-demand balance

independent scenarios for production and transportation disruptions each i.e., 100%, 80%, 60% and 40% capacities. Therefore, we have a total of 4^4 = 64 scenarios, each with an equal probability of occurrence. The demand for each country is fixed at 20% of its population, which was the goal of COVAX in 2021 [1].

6. Results

Evaluating the deterministic base supply chain model without any risk mitigation strategies: First, we optimized the base supply chain model without risk mitigation strategies. We have solved the problem for the ideal (optimistic) scenario which COVAX had expected. Under the ideal scenario, all the deals with the manufacturers are delivered without any delay and the transportation is not subjected to any disruptions. Then we optimized the model for the actual situation faced by the COVAX facility, accounting for all the production and transportation disruptions. Results demonstrate how the supply

chain model failed when subjected to disruptions. This is validated by the fact that COVAX is far behind in fulfilling its objectives. Based on the results, it was found that these disruptions have led to around 64% of the unmet demand (Figure 2b). Based on the deals made by the facility, the allocation results indicate that countries are served by more than one production center. The allocation is not only based on the transportation and storage cost but based on the selling price of the vaccines. Vaccines provided to the AMC-funded countries are cheaper than the ones provided to the self-funding countries. On the other hand, the cost distribution indicates that vaccine cost is the main contributing factor compared to other costs (Figure 2a).

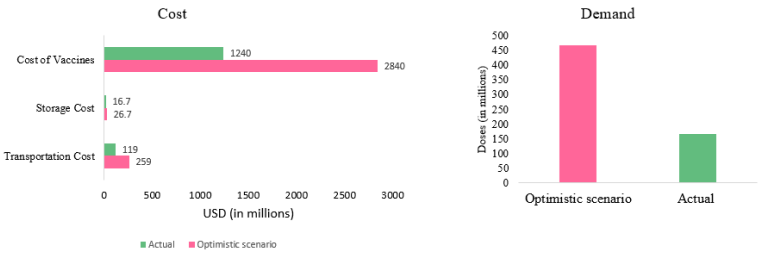


Figure 2. a) Cost distribution for basic supply chain model for optimistic scenario and actual case, b) The demand satisfied under each scenario.

5.2 Evaluating the performance of our model for the 2021 case study: We have successfully demonstrated the performance of our model as 100% demand is met and the cost is optimized (Table 2). Results support the claim that the vaccine demand can be met by having well-coordinated donations as backup production centers, restoring the production capacity after disruption, and having deals for buffer doses. We also studied the effect of disruption on different costs and found out that disruptions lead to an increase in the overall cost (Table 2). However, which cost factors will increase is not certain and is subjected to the nature of disruption. The cost distribution is illustrated in Figure 3. As we see that the major cost driver is the cost of vaccines, followed by the 3PL transportation cost. This is because 3PL handles the transportation of donations from various countries. Surprisingly, the results also demonstrate that the optimized cost for our model is only 13% higher than the ideal scenario. This proves that the COVAX facility does not have to spend a lot of extra money to meet the demand, rather should focus on coordinating and procuring donations effectively.

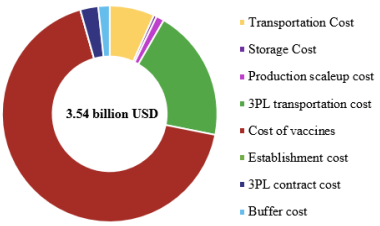


Figure 3. Cost distribution for our supply chain model for actual scenario of 2021 case study

Table 1. Optimization results of scenarios considering various disruptions.

Scenario	Total Cost (USD in billions)	Demand Satisfied
Optimistic scenario (Base model)	3.13	100%
Actual (Base model)	1.48	35.7%
Actual (Our model)	3.54	100%

7. Conclusions

The case study of 2021 has been used to study the performance of our supply chain model. Through our proposed model, we demonstrated that different risk mitigation strategies are crucial in order to successfully deliver the vaccines worldwide. In the absence of such measures, it is observed that the demand is not met, and the low- and middle-income countries are suffering the most. To handle various disruptions, a two-stage scenario-based MILP s programming model is presented. Scaling up the production facilities, coordinating donations, contracting 3PL to manage sudden transportation requirements are considered as the mitigation (i.e., resilience-enhancing) strategies. Furthermore, the COVAX facility also should invest in restoring the disrupted capacitates. The model demonstrates how COVAX could have battled various challenges it faced during 2021, and these strategies should be employed for 2022 to effectively deliver the vaccines to all the countries. Future work is to use this model and determine the strategies that COVAX should adopt for 2022 to meet the delivery promises. The future work is to forecast the dynamics of the virus in different countries through compartmental modeling and determine the vaccine demand to attain herd immunity and use our model to plan the production and distribution of the vaccines.

References

1. World Health Organization. "Fair allocation mechanism for COVID-19 vaccines through the COVAX Facility." Final working version-9 September (2020).

2. ECCLESTON-TURNER, M. A. R. K., and Harry Upton. "International Collaboration to Ensure Equitable Access to Vaccines for COVID-19: The ACT-Accelerator and the COVAX Facility." The Milbank Quarterly (2021).

3. Golan, Maureen S., et al. "The vaccine supply chain: a call for resilience analytics to support COVID-19 vaccine production and distribution." COVID-19: Systemic Risk and Resilience. Springer, Cham, 2021. 389-437.

4. Georgiadis, Georgios P., and Michael C. Georgiadis. "Optimal planning of the COVID-19 vaccine supply chain." Vaccine 39.37 (2021): 5302-5312.

5. DeRoo, Sarah Schaffer, Natalie J. Pudalov, and Linda Y. Fu. "Planning for a COVID-19 vaccination program." Jama 323.24 (2020): 2458-2459.

6. Burki, Talha Khan. "Challenges in the rollout of COVID-19 vaccines worldwide." The Lancet Respiratory Medicine 9.4 (2021): e42-e43.

7. Spadaro, Benedetta. "COVID-19 vaccines: challenges and promises of trials, manufacturing and allocation of doses." (2020): FDD.

8. Asundi, Archana, Colin O'Leary, and Nahid Bhadelia. "Global COVID-19 vaccine inequity: The scope, the impact, and the challenges." Cell Host & Microbe 29.7 (2021): 1036-1039

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50101-9

Optimal Liquefied Natural Gas (LNG) Annual Delivery Program Reflecting both Supplier and Customer Perspectives

Dnyanesh Deshpande[a], Mohd Shahrukh[a],Rajagopalan Srinivasan[a*] , I.A Karimi[b]

[a]*Department of Chemical Engineering , Indian Institute of Technology Madras, Chennai - 600036 , India*
[b]*Department of Chemical & Biomolecular Engineering , National University of Singapore, 119077, Singapore*
raj@iitm.ac.in

Abstract

Liquefied Natural Gas (LNG) is a convenient way of storing and transporting natural gas. Traditionally, LNG is traded via Long Term Contract (LTC) signed between the suppliers and the customers. But recently share of the spot market has increased significantly. This makes the development of a suitable delivery strategy of LNG shipments complex for the supplier. Therefore, the supplier develops an annual schedule of delivery for the upcoming year known as the Annual Delivery Program (ADP). In this paper, we present an ADP planning problem for an LNG supplier. A mixed-integer linear programming (MILP) model is developed with the objective to minimize the cost of delivering LNG shipments.

Keywords: Liquefied Natural Gas, Annual Delivery Program, Scheduling, Mixed Integer Linear Programming

1. Introduction

The demand for natural gas is increasing rapidly due to environmental considerations. However, transportation of natural gas from producing regions to customers around the world via pipelines is economically unattractive. Therefore, natural gas is liquefied to -162 °C at atmospheric pressure to form Liquefied Natural Gas (LNG), making it convenient for storage and transportation. Traditionally, LNG is traded through Long Term Contract (LTC) signed between supplier and customer. But due to an increase in the number of suppliers and customers around the world, a spot market for LNG trade has emerged in recent years.

LTCs guarantee supply security to the customers but they do not offer flexibility due to the presence of strict clauses in the contract (Shahrukh et al.,2021). Ensuring customer satisfaction and their reliability on the supplier in future is crucial. This motivates the supplier to come up with a strategy of delivering LNG shipments to its customers at the accepted delivery dates throughout the planning horizon. In order to plan for the LNG shipment deliveries, the supplier creates an Annual Delivery Program (ADP). An ADP is the list of scheduled voyages with information about the long-term contract served, ships nominated to serve the contracts, customer terminals nominated to receive delivery, date of ship loading etc. In this paper, we present an ADP planning problem

with an objective to create an ADP which satisfies customers' demands at minimum cost.

1.1. Literature Survey

The ADP planning problem has been thoroughly studied in the literature. Stålhane et al. (2011) developed a Mixed Integer Programming (MIP) formulation and a construction and improvement heuristic (CIH) to solve the problem. Rakke et al. (2011) developed a rolling horizon heuristic (RHH) for creating ADPs. Andersson et al. (2010) solved two problems, one for a supplier of LNG and one for a vertically integrated company. Mathematical models for each problem are presented, and solution methods are discussed, but the computational results are not given. Mutlu et al. (2015) presented an ADP planning problem that allowed split delivery for delivering LNG shipments. The proposed model was computationally very expensive to solve. So, they proposed an efficient vehicle routing heuristic (VRH) which gave cost-effective solutions and outperformed commercial optimizers.

Generally, ADP planning problems in literature have focused on inventory and berth management at the supplier's terminal. They have considered that the supplier can deliver only one grade of LNG to a customer. These models are unable to satisfy customers' demands as they allow under-supply. Therefore, in this paper, we present a MILP model for the ADP planning problem considering inventory and berth management at the customer terminals, allowing the supplier to produce and deliver multiple grades of LNG and satisfying customer's demands with a reasonable over-supply.

2. Problem Description

The ADP planning problem has a supplier of LNG. The supplier produces multiple grades of LNG at a variable production rate. The supplier serves multiple LTC customers. Every customer has a slot-wise demand for one or more than one grade of LNG. When the supplier delivers more than the demand, the customer is said to be over-supplied. The supplier is penalized for over-supplying the customer and ensures that the over-supply is minimal. The supplier owns a heterogeneous fleet of ships. The supplier takes decisions regarding the determination of the production rate of every LNG grade and scheduling of the fleet. Scheduling of fleet involves taking decisions regarding loading and maintenance of ships. Loading of a ship takes place at a berth of the supplier terminal. Maintenance of ships is carried out at a maintenance terminal located near the supplier terminal. After maintenance, the ship undergoes a purge and cool down operation at the berth of the supplier's terminal. The objective of the problem is to create an ADP to satisfy the customer demands at minimum cost. The cost incurred by the supplier in delivering LNG shipments is the sum of transportation and penalty cost.

The sets of LTC customers, ship capacities in the heterogeneous fleet, planning horizon, initial inventory at the supplier and customer terminals, minimum and maximum storage capacities at the supplier and customer terminals, minimum and maximum production rates for each grade of LNG are known. Transportation time, slot wise demand at all customer terminals, time window and duration of maintenance, transportation cost of ships in \$ per slot and penalty in \$ per m^3 of oversupply to a customer are also known.

3. Mathematical Model

3.1. Modelling inventory and berth management at the customer terminals

The proposed model for ADP planning problem considers inventory and berth management at the supplier terminal, production of multiple LNG grades, over-supply and maintenance of ships. These aspects were modelled based on the models developed by Andersson et al. (2010), Rakke et al. (2011), Stålhane et al. (2011) in literature. But models in literature do not consider inventory and berth management at the customer terminals and delivery of multiple grades of LNG. For modelling inventory and berth management, we consider a set of customer terminals LTC (LTC = 1, 2, ...C) with every terminal having demand for one or multiple grades of LNG. D_{tgc} is the slot-wise demand of grade g LNG at customer $c's$ terminal. Inventory for every grade of LNG is stored in separate storage tanks. At every customer terminal c inventory level in storage tanks is monitored for every grade g at the end of every slot t and is denoted by I_{cgt}. So, inventory balance at customer terminals is written as follows:

$$I_{cgt} = I_{cg(t-1)} - D_{tgc} + \sum_{v \in V} C_v . x_{cgv(t-TT_{cv}-1)} \quad \forall\, c \in LTC, g \in G, t \in T \tag{1}$$

where x_{cgvt} is equal to 1 if ship v loads at the start of slot t to serve customer c having demand for grade g LNG and C_v is the capacity of ship v in m^3

The inventory of each grade of LNG must be maintained between the minimum and the maximum storage capacities of the terminal. This is ensured by writing the following constraint:

$$I_{cg}^{min} \leq I_{cgt} \leq I_{cg}^{max} \qquad \forall\ c \in LTC, g \in G,\ t \in T \tag{2}$$

The entire fleet of ships is divided into Q-Flex and Q-Max ships based on their capacities. Q-Flex and Q- Max ships are allotted different berths for unloading at the customer terminal. Now every customer terminal has a fixed number of berths available for unloading of ships. So, the number of ships unloading in every slot should not exceed the number of berths available at the customer terminal. This is called as the berth capacity constraint and is modelled as follows:

$$\sum_{g \in G} \sum_{v \in QF} x_{cgv(t-TT_{cv}-1)} \leq B1_c \quad \forall\, c \in LTC,\ t \in T \tag{3}$$

$$\sum_{g \in G} \sum_{v \in QM} x_{cgv(t-TT_{cv}-1)} \leq B2_c \quad \forall\, c \in LTC, t \in T \tag{4}$$

Here QF and QM are the set of Q-Flex and Q-Max ships in the heterogeneous fleet and $B1_c$ and $B2_c$ are the number of berths available for unloading of Q-flex and Q-max ships at customer terminal c respectively.

Every ship can serve a single customer in a given voyage. We call this the regular delivery constraint. It is modelled as follows:

$$\sum_{c\in LTC}\sum_{g\in G}\sum_{k=t}^{t+2TT_{pv}+1} x_{cgvk} \leq 1 \quad \forall\, p \in LTC, v \in V, t \in T \tag{5}$$

Whenever the supplier delivers LNG more than the customer's demand, the customer is said to be oversupplied. Oversupply of grade g LNG for customer c is denoted by os_{cg} and is modelled as follows:

$$os_{cg} = \sum_{v\in V}\sum_{t\in T} C_v . x_{cgv(t-TT_{cv}-1)} - \sum_{t\in T} D_{tgc} \quad \forall\ c \in LTC, g \in G \tag{6}$$

Over-supply should be non – negative. This is ensured by the following constraint:

$$os_{cg} \geq 0 \quad \forall\ c \in LTC, g \in G \tag{7}$$

3.2. Objective Function

The objective is to minimize the cost incurred by the supplier in delivering LNG shipments to its customers over the planning horizon. The cost incurred by the supplier is the sum of transportation cost and penalty cost. CR is the transportation cost of ships in \$ per slot, TT_{cv} is the one-way transportation time in slots required for ship v to reach customer $c's$ terminal, P is the penalty in \$ per m^3 of oversupply to the customer. Then the objective function is formulated as follows:

$$\begin{aligned} Cost = & \sum_{c\in LTC}\sum_{g\in G}\sum_{v\in V}\sum_{t\in H} CR * (2TT_{cv} + 1) * x_{cgvt} \\ & + \sum_{c\in LTC}\sum_{g\in G} P * os_{cg} \end{aligned} \tag{8}$$

4. Results

4.1. Illustrative Example

The example presented in this section considers a supplier of LNG serving two LTC customers. The supplier has a heterogenous fleet of 12 ships. The supplier produces two grades of LNG, namely Lean LNG (LNGL - G1) and Rich LNG (LNGR – G2). The production rate of each grade of LNG is known and varies between known limits. The supplier serves 2 LTC customers. In this example, we have taken that customer 1 (C1) has demand for both the grades, customer 2 (C2) has demand for LNGL only. The planning horizon is of 60 days. The entire planning horizon is divided into 60 uniform slots having a duration of 1 day. The duration for both loading and unloading of ships at the supplier and customer terminals is considered as 1 slot. The chartering rate considered for evaluating the transportation cost for delivering LNG is $ 80,000 per slot. The penalty per m^3 of oversupply is taken as $25.The transportation and penalty cost incurred by the supplier over the planning horizon is $37,760,000 and $1,79,000, respectively. The supplier delivers total 15 LNG shipments out of which 10 shipments are delivered to C1(5 shipments of LNGL and 5 shipments of LNGR) while 5 shipments are delivered to C2. Table 1 shows the ship schedule obtained by solving the illustrative example. The production rate for both the grades of LNG is almost constant over the planning horizon. The supplier operates at the upper limit of production rate for LNGL (G1) and then operates at the lower limit while for LNGR (G2) it almost operates at the lower limit. The supplier is able to satisfy the LTC customers demand with less than 1 % oversupply. The percentage over-supply for C1 for LNGL and LNGR is 0.2 % and 0.13%, respectively and that for C2 for LNGL is 0.67 %. The LNG shipment deliveries are fairly evenly spread over the planning horizon, which maintains the inventory at the customer terminals within the operational limits.

Table 1:Table giving the ship schedule for both the customers over the planning horizon

Customer	Grade of LNG	Ship Number	Loading Slot	Volume of Shipment (m^3)
C1	LNGR	12	1	259,789
C2	LNGL	8	1	181,454
C2	LNGL	11	3	252,597
C1	LNGL	9	4	237,919
C1	LNGR	10	5	244,826
C1	LNGR	4	6	157,154
C2	LNGL	7	11	173,859
C1	LNGL	5	15	163,395
C1	LNGL	6	28	168,550
C2	LNGL	8	32	181,454
C1	LNGR	1	37	138,270
C2	LNGL	12	39	259,789
C1	LNGL	11	41	252,597
C1	LNGR	10	42	244,826
C1	LNGL	9	44	237,919

4.2. Computational Results

The illustrative example was implemented in IBM ILOG CPLEX Optimization Studio 12.10.0 on a Dell Workstation with Intel(R) Xeon(R) Silver 4114 CPU@2.20 GHz with 32 GB RAM. While solving the example, we have considered a relative convergence criterion of 0.5 % with a global time limit of 86,400 seconds. Also, we have considered Depth First Search (DFS) as the node selection strategy and enabled solution polishing after a relative gap of 10 %. It took 31.82 seconds to solve the 2-month example with a relative gap of 0.47 %. The example had 3,668 binary and 490 continuous variables, 3,578 constraints, and the optimal cost obtained was $ 37,939,000.

5. Conclusions

In this paper, we present an ADP planning problem for a supplier of LNG. The objective of the problem was to develop a ship schedule which would satisfy LTC customers' demand at minimum cost. In this regard, a mixed integer linear programming (MILP) model is developed. ADP planning problems in literature did not focus on inventory and berth management at the customer's terminal which may lead to the violation of customer inventory resources which in effect will lead to a change in the developed ADP. Hence, our model reduces the probability of this change in ADP by considering customer inventory and berth availability. This consideration of resources at customer terminals also leads to a reduction in oversupply as compared to models proposed in the literature. Reduction in oversupply reduces the penalty cost incurred by the supplier. These incentives of our model are explained by solving an illustrative example, and the computational results of the example are also reported. In the future, we aim to develop an ADP considering spot market sales along with LTCs.

References

F. Mutlu, M. K. Msakni, H. Yildiz, E. Sönmez, and S. Pokharel, "A comprehensive annual delivery program for upstream liquefied natural gas supply chain," *European Journal of Operational Research*, vol. 250, no. 1, pp. 120–130, Apr. 2016.

H. Andersson, M. Christiansen, and K. Fagerholt, "Transportation Planning and Inventory Management in the LNG Supply Chain," in *Energy, Natural Resources and Environmental Economics*, E. Bjørndal, M. Bjørndal, P. M. Pardalos, and M. Rönnqvist, Eds. Berlin, Heidelberg: Springer, 2010, pp. 427–439.

J. G. Rakke *et al.*, "A rolling horizon heuristic for creating a liquefied natural gas annual delivery program," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 5, pp. 896–911, Aug. 2011.

M. Shahrukh, R. Srinivasan, and I. A. Karimi, "Optimal Procurement of Liquefied Natural Gas Cargos from Long-Term Contracts and Spot Market through Mathematical Programming," Ind. Eng. Chem. Res., vol. 60, no. 9, pp. 3658–3669, Mar. 2021.

M. Stålhane, J. G. Rakke, C. R. Moe, H. Andersson, M. Christiansen, and K. Fagerholt, "A construction and improvement heuristic for a liquefied natural gas inventory routing problem," *Computers & Industrial Engineering*, vol. 62, no. 1, pp. 245–255, Feb. 2012.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50102-0

The Waste-to-Resource Game: Informed Decision-Making for Plastic Waste Transformers

Fabian Lechtenberg[a*], Antonio Espuña[a], Moisès Graells[a]

[a] *Department of Chemical Engineering, Universitat Politècnica de Catalunya, Barcelona 08019, Spain*
[*]fabian.lechtenberg@upc.edu

Abstract

Upgrading waste to recovered resources requires a (chemical) transformation process. Thus, waste transformers, waste providers and potential customers of such added-value resources, are essential actors in the game of circular economy. Hence, this work presents a decision-making framework addressing the definition and solution of multi-leader-follower games to find bargaining outcomes between these actors. The potential of the approach is illustrated through the case of a transformer who operates a pyrolysis process producing char, oil and gas fractions from plastic waste. A novel process model is used to aid the agents in making informed decisions. Results reveal that satisfactory natural bargaining outcomes may not exist in the interaction between the agents. Hence, the need of an authority altering the rules of the game and ensuring win-win situations is stressed.

Keywords: Plastic Waste; Pyrolysis; Game Theory; Surrogate Model; Decision-Making

1. Introduction

Rethinking waste products as suboptimally allocated resources is an essential part of sustainable development. Taking advantage of this unused potential to enable circularity in the supply chains often needs a waste transformer that performs the upgrading of wastes to added-value materials. Multiple objectives need to be considered accordingly. However, the fact that these objectives represent agents that are competing and making decisions based on their individual profit cannot be overlooked. Thus, the PSE community has adopted Game Theory as a powerful complement to multi-objective optimization and derive realistic bargaining outcomes (Avraamidou et al., 2020).

This work studies the bargaining process between a Waste Provider (WP), a Waste Transformer (WT) and a set of potential customers, and develops a strategy to obtain satisfactory situations in which no objective can be improved by unilaterally changing a decision. To that end, multi-leader-follower games (Aussel and Senvsson, 2020) are defined and solved through their corresponding bi-level optimization problems (Beykal et al., 2021).

Thus, this work proposes (1) a general modelling framework for finding “equilibrium” bargaining outcomes in which waste transformers are integrated to extend the lifespan of a product (resource) or to close a supply chain and (2) a process model for the pyrolysis process for plastic conversion. The latter is related to the case study, dealing with the conversion of plastic waste into added-value products. The knowledge captured in this model discloses the existence of satisfactory bargaining outcomes between the waste transformer and its customers under some restricted conditions.

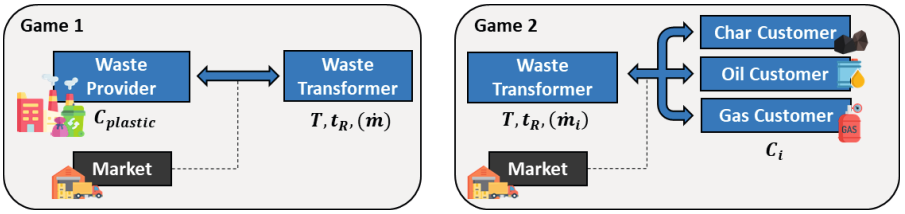


Figure 1. Bargaining interaction between the considered stakeholders in this work: Waste Provider (WP), Waste Transformer (WT) and customers.

2. Problem Statement

The problem and system boundaries considered in this contribution are illustrated in Fig. 1 and can be summarized as follows. Given:

- A Waste Provider (WP) producing, collecting and/or pre-treating plastic waste
- A Waste Transformer (WT) that has the technology and facilities for processing waste into added-value resources (e.g. the WT is willing to buy the pre-treated plastic waste and transform it into valuable products such as char, oil and gas)
- A set of customers (CUST) that buy the products produced by the WT
- A set of alternative strategies that can be taken by each of these agents.

Find optimal strategy sets of the agents and bargaining outcomes. All agents intend to maximize an economic objective. Here, two sequential Stackelberg competition scenarios are considered that, once solved, yield optimal pricing strategies and operational schemes. The pricing strategy, if existent, can be understood as the "equilibrium cost" of plastic waste and added-value products, which is conditioned by the process model of the WT and the market conditions.

3. Bargaining Game

The bargaining between WT, WP and customers is expressed as multi-leader follower games. A general form of the corresponding bi-level structure is given below:

$$\begin{aligned} &\max_x f_i^{leader}(x,y) \\ &\text{s.t.} \quad g_i^{leader}(x,y) \geq 0 \qquad i \in I^{leaders} \\ &\quad y \; solves \; \max_y f_j^{follower}(x,y) \\ &\quad\quad \text{s.t.} \quad g_j^{follower}(x,y) \geq 0 \qquad j \in J^{followers} \end{aligned} \tag{OP1}$$

The objective function of the WP is the maximization of its own profit:

$$f^{WP} = \dot{m}_{Waste}^{WP} \cdot C_{Waste} \tag{1}$$

With $\dot{m}_{Waste}^{WP}$ being the mass bought by the WT and C_{Waste} the price of the waste proposed by the WP. The objective function of the customers can be similarly expressed through the maximization of the savings over buying from the market:

$$f_k^{CUST} = \dot{m}_k \cdot \left(C_k^{market} - C_k\right) \qquad k = \text{customers} \tag{2}$$

Where $\dot{m}_k$ is the amount of product the WT sells to the customers and C_k is the price proposed by the customers. Since the amount $\dot{m}_k$ depends on C_k, the customers must strategize on how to choose a price that maximizes their benefit.

Finally, the objective of the WT is expressed as the maximization of its revenue:

$$f^{WT} = \dot{m}_{Waste}^{WP} \cdot \left(\sum_{k} \eta^{k} \cdot C_{k} - C_{Waste}^{WP} \right) \tag{3}$$

Here, η^k is the conversion of plastic waste into added-value products k. Utility costs (electricity, pumping …) could be included. The limits on the decision variables (strategies) and other constraints are stated in $g_{i/j}(x, y)$. The possible leader-follower relations in this three-actor system are summarized in Tab. 1.

Table 1. Bargaining games constellation

Game	Leader	Follower	Type
1.1	WT	WP	Single-Leader Single-Follower (SLSF)
1.2	WP	WT	Single-Leader Single-Follower (SLSF)
2.1	WT	CUST	Single-Leader Multi-Follower (SLMF)
2.2	CUST	WT	Multi-Leader Single-Follower (MLSF)

4. Pyrolysis Model

For each case (i.e. for each different WT considered), a specific model for the transformation process will be required. This model may be available or may need to be developed. The particular case study addressed in this work corresponds to the chemical upcycling of plastic waste through a pyrolysis process. There are only few models that describe the pyrolysis of plastic waste and those that are available are often limited to a single set of operating conditions. The distribution of pyrolysis products strongly depends on key operating conditions like the residence time and reaction temperature (Miandad et al., 2016). This allows the plant operator to change the operating conditions to maximize the expected value of the product portfolio.

The pyrolysis process model developed herein extends the one presented by Fivga and Dimitriou (2018) through the incorporation of temperature and residence time dependent product yields in the reactor model. The reactor is assumed to be a fluidized bed reactor that can vary the residence time t_R and reaction temperature within some limited bounds by manipulating the recycle stream and the furnace operating conditions. A cyclone acts as solid separator to obtain the char product. The non-recycled gaseous product is cooled down to 25 ºC to obtain the oil and gas products. Part of the gas product is burned in a furnace to supply the heat of reaction, so no external fuel is needed to drive the reaction.

In order to be able to evaluate the economic objective function of the WT it is necessary to quantify how much char, oil and gas is produced (η^k). Experimental data for the non-catalytic conversion of polyethylene (PE) waste was taken from Quesada et al. (2019). The authors report oil and char yield of experiments performed at 15 different temperature, residence time and heating rate combinations. Based on this data, a set of metamodels is trained to predict oil and char yields at the non-measured conditions. It is found that Gaussian Process Regressor (GPR) models and second- and third-degree polynomials exhibit desirable prediction and interpolation capabilities, exceeding those of the ANFIS and second-degree polynomials considered by Quesada et al. (2019).

For a fixed mass flow $\dot{m}^{plastic}$ the oil and char products can be directly calculated from the yield correlations $\eta^{char}(t_R, T)$ and $\eta^{oil}(t_R, T)$ described by the trained metamodels. Note, that the heating rate dependence has been omitted here due to the small influence

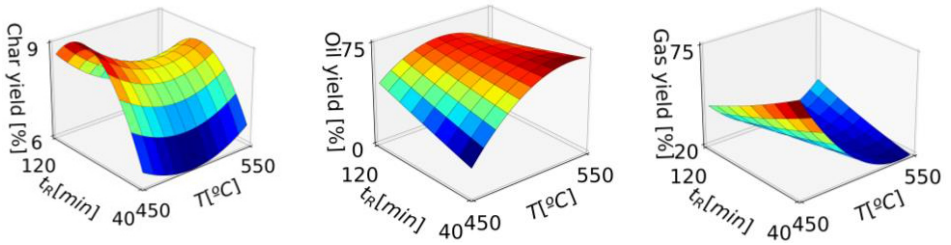


Figure 2. Char, oil and gas yield response surfaces (3rd degree polynomials).

on the yields. A similar correlation for the gaseous product has been obtained from repeated simulation of the flowsheet at different operating conditions. See Fig. 2 for the response surfaces.

By fixing the reference enthalpy of the plastic inlet, it is imposed that at nominal operating conditions (500 ºC, 80 min) the heat of reaction equals 1316 kJ/mol (Fivga and Dimitriou, 2018). Following the same authors' example, it is assumed that the pyrolysis product can be sufficiently well represented by carbon and six selected hydrocarbons in the range of C_2-C_{30}. Then, a first-order decomposition (Eq. 4) was assumed and the kinetic constants were fitted to include the reported compositions.

$$\text{Plastic} \xrightarrow{k1} \text{n-C30} \xrightarrow{k2} \text{n-C25} \xrightarrow{k3} \text{n-C18} \xrightarrow{k4} \text{n-C14} \xrightarrow{k5} \text{n-C8} \xrightarrow{k6} \text{C2} \quad (4)$$

By doing so, a lookup table for the reactor outlet composition is obtained correlating the pyrolysis product composition with the temperature and residence time.

5. Case Study

We consider an operational problem of a plant that has been designed for treating 4 t/h of municipal PE plastic waste. The amount bought by the WT is not part of its strategy. The bounds on the prices proposed by the WP are lower bounded ($C_{plastic}^{low}$) by the cost that he would usually pay/receive for treating its waste and upper bounded ($C_{plastic}^{up}$) by the cost of virgin PE or the revenue of the highest value product of the WP. The prices that can be proposed by the customers are bounded ($C_k^{low/up}$) by some lower bound that is subject of investigation (> 0 €/kg) and the market price of the virgin raw material. To illustrate the effect of different (ratios of) values of the products we arbitrarily chose their market prices to be equal at 1 €/kg. The operational limits of the WT (T, t_R) stem from the range used to train the metamodels (450 – 550 °C, 40 – 120 min) and can be justified by the kinetics of the reaction and technical limitations of the reactor.

6. Results and Discussion

The bi-level problems can be solved through various approaches such as multi-parametric programming or data-driven optimization (Beykal et al., 2021). Since the complexity of the problem strongly depends on the configuration and constraints considered, here, different strategies are used for each game.

The solutions to games 1.1 and 2.1 are straightforward: The follower's problem (WP and customers) comprise a linear problem with box constraints on the decision variable which can be substituted by its KKT optimality conditions. Since the follower has no conjecture about the leader's decision, its safest bet is to decide on the price that would maximize its objective for a fixed $\dot{m}_i$. In the case of the WP this limit is the market price raw plastic price $C_{plastic}^{up}$ and for the customers the low bound C_k^{low} (see Tab. 3).

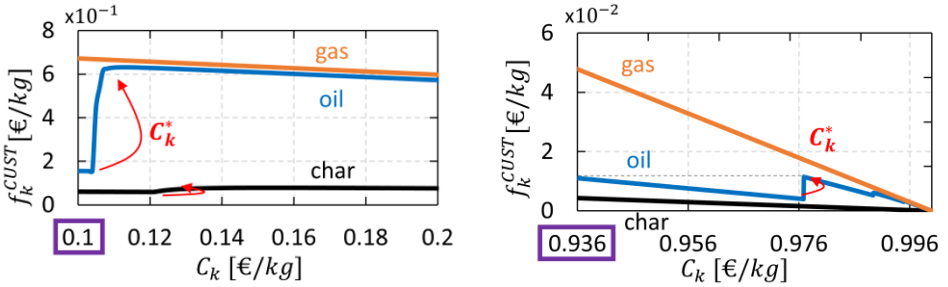


Figure 3. (left) Customer objectives depending on unilateral changes of proposed prices at C_k^{low}= 0.1 €/kg. (right) Customers do not improve their objectives by deviating from C_k^{low}=0.936 €/kg.

It has to be acknowledged that these solutions cannot be considered satisfactory. The WT would not buy the waste plastic at the same price as virgin plastic and it would also not sell the products for 0 €/kg.

When changing the role of the leader and the follower (WT as follower), it is assumed that the leader has full knowledge about how the follower will react to its decisions. In the case of game 1.2 this knowledge does not add any additional value to the bargaining since we assume that the exchanged mass is fixed and not part of the follower's strategy. As a result, the WT has no way of influencing the WP's decision, leading him to choose again the highest possible price. Neither game 1.1 nor 1.2 yield satisfactory bargaining outcomes. A potential authority setting rules or redirecting taxes and/or incentives could resolve this situation (e.g. by introducing an incentive that stems from the avoided cost of municipal waste treatment).

In the case of game 2.2 the knowledge of the customers about the WTs reaction can be exploited by them to make an informed decision on the price that they propose. Fig. 3 (left) illustrates how changing the proposed price influences the agent's objective. It can be seen that when every agent proposes the same price (0.1 €/kg) the WT will operate at those conditions with the overall highest yield (least amount of gas burned). This situation favours the gas customer. However, the oil and char customers can react by proposing a slightly higher price that convinces the WT to operate at different conditions and produce more oil or char respectively.

The bi-level game was reformulated into an NLP problem following the description by Leyffer and Munson (2010) and solved using BARON in GAMS. Not all properties (e.g. convexity of follower's problem) are fulfilled to qualify the solution as an equilibrium point. Thus, a numerical test has been conducted. It is found that, due to the specific characteristics of the WT production process (model), a strict mathematical equilibrium point only exists when the lower bound (C_k^{low}) is forced to be ≥ 0.936 €/kg for all agents. In this situation, any bid from the char or oil customers will not improve their objective despite the WT changing its operating conditions (C_k^*) (Fig 3. (right)).

Table 2. Summary of equilibrium solutions in the studied bargaining games.

Game	Leader	Follower
1.1	$T^{opt}, t_R^{opt} = f(C_{plastic}^{up})$	$C_{plastic} = C_{plastic}^{up}$
1.2	$C_{plastic} = C_{plastic}^{up}$	$T^{opt}, t_R^{opt} = f(C_{plastic}^{up})$
2.1	$T^{opt}(C_k^{low}), t_R^{opt}(C_k^{low})$	$C_k = C_k^{low}$
2.2	$T^{opt}(C_k^{eq*}), t_R^{opt}(C_k^{eq*})$	$C_k = C_k^{eq*}$ (Fig 3)

This points again towards the need of an authority that sets the initially arbitrarily defined bound (C_k^{low}) to this "forced equilibrium" bound. The solution could be regarded as satisfactory from all agents' points of view: The WT can sell its recovered added value products at a price slightly below the market price of virgin or fossil-sourced materials while the customers win by paying less than the market price.
Table 2 summarizes the results of the studied case. These results are conditioned by the underlying process model and market assumptions and can be very different for other cases (e.g. different chemical recycling process, recovery of used solvents …).

7. Conclusions

The presented work introduces a general bargaining framework between waste transformers, waste providers and customers of the added-value materials. Applying this framework to a chemical recycling of waste plastic case study reveals that one-to-one bargaining between a waste provider and a waste transformer leads to outcomes that are not acceptable for either of the agents. The interaction between waste transformer and customers neither yields acceptable results if the customers choose their minimum acceptable price too ambitiously. Both situations point towards the need of an authority that sets the rules of the games so that win-win situations are enforced, despite the agents' effort to maximize their own profit. The knowledge captured in the proposed WT's process model enables the determination of a reasonable lower bound on the customers' pricing strategies. Future work should explicitly consider such an authority in the modelling framework. Moreover, this authority should follow strategies that focus not only on economic but also environmental and social sustainability.

Acknowledgements: Financial support from the Spanish "Ministerio de Ciencia e Innovación" and the European Regional Development Fund, both funding the Project CEPI (PID2020-116051RB-I00) is fully acknowledged. F.L. gratefully acknowledges the Universitat Politècnica de Catalunya for the financial support of his predoctoral grant FPU-UPC, with the collaboration of Banco de Santander.

References

D. Aussel, A. Svensson, 2020, A Short State of the Art on Multi-Leader-Follower Games, Bilevel Optimization, 53-76

S. Avraamidou, S.G. Baratsas, Y. Tian, E.N. Pistikopoulos, 2020, Circular Economy - A challenge and an opportunity for Process Systems Engineering, Computers & Chemical Engineering, 133

B. Beykal, S. Avraamidou, E.N. Pistikopoulos, 2021, Bi-level Mixed-Integer Data-Driven Optimization of Integrated Planning and Scheduling Problems, Computer Aided Chemical Engineering , 50, 1707-1713

R. Miandad, M. Barakat, A. Aburiazaira, M. Rehan, A. Nizami, 2016, Catalytic pyrolysis of plastic waste: A review, Process Safety and Environmental Protection, 120, 822-838

A. Fivga, I. Dimitriou, 2018, Pyrolysis of plastic waste for production of heavy fuel substitute: A techno-economic assessment, Energy, 149, 865-874

L. Quesada, A. Pérez, V. Godoy, F.J. Peula, M. Calero, G. Blázquez, 2019, Optimization of the pyrolysis process of a plastic waste to obtain a liquid fuel using different mathematical models, Energy Conversion and Management, 188, 19-26

S. Leyffer, T. Munson, 2010, Solving Multi-Leader-Common Follower Games, Optimisation Methods & Software, 601-623

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50103-2

Implications of Optimal BECCS Supply Chains on Absolute Sustainability

Valentina Negri[a], Gonzalo Guillén-Gosálbez[a*]

[a]*Institute for Chemical and Bioengineering, Department of Chemistry and Applied Biosciences, ETH Zürich, Vladimir-Prelog-Weg 1, 8093 Zürich, Switzerland*
gonzalo.guillen.gosalbez@chem.ethz.ch

Abstract

Carbon dioxide removal technologies are expected to play a decisive role in meeting the target of 1.5 °C, yet their broad sustainability implications remain unclear. Among those, bioenergy with carbon capture and storage (BECCS) has attracted growing interest, as it can remove CO_2 while providing energy. This study presents an optimization approach to design and evaluate BECCS supply chains based on absolute sustainability criteria. We analyze the solution to a minimum cost scenario removing 0.61 $GtCO_2$ in the European Union, assessing the impacts with the Environmental Footprint method and interpreting the results using the planetary boundaries linked to the United Nations Sustainable Development Goals. We find that BECCS could indeed be implemented within the safe operating space but would consume large amounts of global ecological shares due to burden-shifting on some categories.

Keywords: bioenergy with carbon capture and storage, supply chain, optimization, planetary boundaries, Sustainable Development Goals.

Introduction

The IPCC pathways that limit the global temperature rise by 2100 to well below 2 °C above pre-industrial levels indicate that a large deployment of negative emissions technologies and practices (NETPs) is required to compensate for the emissions from hard-to-abate sectors (IPCC, 2018).

Bioenergy with carbon capture and storage (BECCS) has been identified as the most promising nature-based NETPs because it contributes to CO_2 removal (CDR) while providing a clean, reliable energy source. BECCS, defined as the coupling of bioenergy production with carbon capture and storage (CCS) (Canadell and Schulze, 2014), has been extensively studied in the literature from a techno-economic standpoint.

From an environmental perspective, BECCS has proven to have the capacity to be implemented at a large scale and contribute substantially to achieving the CDR required to meet the Paris Agreement goal. Nevertheless, it depletes resources such as water and land (Heck et al., 2018), ultimately competing with food production (van Vuuren et al., 2018). The severity of these impacts can be assessed following a life cycle assessment (LCA) approach. However, standard LCA studies lack reference values to interpret the results and, consequently, they provide limited insights into the broad implications of deploying NETPs at a large scale.

In recent years, absolute sustainability assessments based on LCA principles were put forward to quantify impacts relative to the planet's carrying capacity. Such methods are based on the planetary boundaries (PB) concept (Rockström et al., 2009), which defines a set of biophysical limits of the Earth system that should never be surpassed to operate our planet safely. Recently, Sala et al. (2020) built on existing literature to map 16

indicators of the Environmental Footprint (EF) method to five Sustainable Development Goals (SDGs). Despite sustainability assessments are emerging in the literature, especially for what concerns supply chains (SCs) (Mota et al., 2013 and 2017, Barbosa-Póvoa, 2009), their application to CDR technologies is at the very beginning.
In this contribution, we quantify the performance of BECCS on five SDGs, previously studied by qualitative approaches alone (Honegger et al., 2020 for CDR options in general and Smith et al., 2019 for land-based solutions in particular).

Methods

In a previous work (Negri et al., 2021), we presented a highly detailed BECCS SC model (NETCOM, Negative Emissions Technologies COoperative Model) that identifies the optimal SC configuration in the European Union (EU) for a net global yearly CDR target retrieved from Peters and Geden (2017). The model includes all EU Member States as of 2018 (28) and assumes full cooperation among the countries to meet the climate target. Following a life cycle optimization approach, the model optimizes costs, emissions and impacts on the Recipe 2016 endpoints. The input data are given specifically for the five echelons of the SC included in NETCOM, connected by transportation. They comprise unitary costs, crop yield and carbon intensities, physical limits for biomass cultivation and growth, and environmental impacts.
Here we enlarge the scope of a cost-driven optimization in NETCOM to include a set of metrics that evaluates the impact of deploying BECCS on the PB linked to five SDGs, providing a more detailed picture of the implications of deploying CDR at a large scale on sustainable development. Yet, the model is still a linear programming, which guarantees a global optimum solution.
Figure 1 provides a sketch of the updated NETCOM. This mathematical model consists of mass and energy balances at each stage of the SC and calculates the total cost, emissions and life cycle impacts.

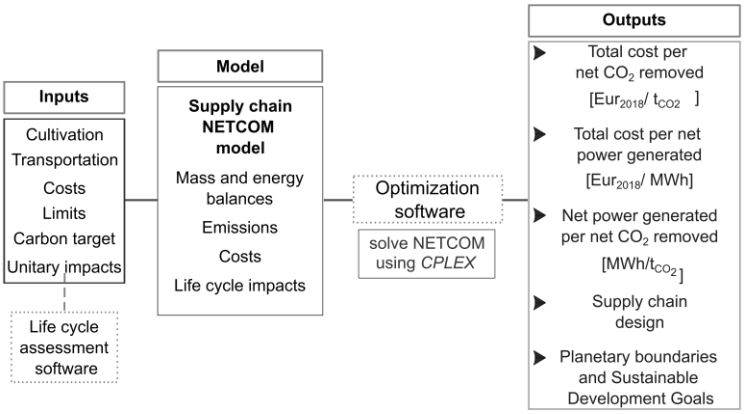


Figure 1. Structure of the NETCOM model. Input data, equations, solver, and outputs included in the model.

To this end, we quantify the total impact on a set of LCA metrics k connected to the PBs and five SDGs, summing up the impacts of all the activities in the SC. Each of them is determined by multiplying an ecovector defined for every activity with the corresponding functional unit, as shown in Eq.(1).

$$\mathrm{eimp_k} = \sum_{\mathrm{i}} \mathrm{fu_i}\ \mathrm{IMP_{ik}}\ \forall\ \mathrm{k} \in \mathrm{K} \tag{1}$$

where $\mathrm{eimp_k}$ is the total environmental impact of the BECCS SC on metric k, calculated as the sum of each activity i characterized by its functional unit ($\mathrm{fu_i}$). The ecovector $\mathrm{IMP_{ik}}$ denotes the impact per unit of activity i in each metric k, calculated by implementing a full LCA of the activity in SimaPro v.9.0.0.48.
We refer to the original work (Negri et al.) for detailed information about NETCOM. In the analysis presented hereafter, we update the cost of CO_2 transportation via pipeline to \$ 3.60/tCO_2/250 km from Budinis et al. (2018).

Eq.(1) provides the impact of the SC, which needs to be evaluated relative to the global carrying capacity (i.e., safe operating space). Sala et al. provided life cycle impact assessment-based limits referred to the EF method. The human health-related limits, namely human cancer and non-cancer effects, particulate matter and ionizing radiation, depend on the population and have been updated for 2018. In order to assess the impact of BECCS in the EU, we downscale the global limits by applying an egalitarian-based sharing principle, similarly to Wheeler et al. (2020), obtaining shares of the safe operating space for EU ($\mathrm{PB^{EU}}$). Then, the transgression level ($\mathrm{tl_k}$) of the SC on the metric k is calculated as in Eq.(2).

$$\mathrm{tl_k} = \frac{\mathrm{eimp_k}}{\mathrm{PB_k^{EU}}}\ \forall\ \mathrm{k} \in \mathrm{K} \tag{2}$$

Later, the transgression level is incorporated into the original optimization model's objective function, which includes mass balances, capacity constraints, and other techno-economic and environmental equations of the BECCS SC for each EU country.

The problem, comprising 756,298 variables and 712,381 equations, was solved in GAMS 35.1 with CPLEX on an Intel CoreTM i7-10510U machine at 1.80 GHz and 16 GB RAM running Windows 10.

Figure 2. Heat map of the feedstock distribution in the EU-28 countries. The light grey cells represent no use of the corresponding biomass type in that country. Straw, woody and forest residues dominate, while only Miscanthus is selected among the energy crops available.

Results

We ran NETCOM to minimize the total cost of the SC subject to a minimum net removal of 0.61 $GtCO_2$. The minimum cost is 74 billion Euros, with the CDR constraint being active. The largest cost contributions include biomass combustion (58 %), transportation (13 %) and pelleting (10 %). The latter two are also the main contributors to the SC emissions, each one accounting for 30 % of the total amount. The net electricity production is 511 TWh, which accounts for the energy penalty associated with the CCS system at the biomass combustion stage. Here we do not integrate BECCS in the EU energy system; therefore, the electricity generation is considered in the functional unit together with the annual CO_2 removal.

The total feedstock is 562 Mt of biomass, consisting of 88 % of residues (straw cereals, woody and forest) complemented by the energy crop Miscanthus. The latter is cultivated in Bulgaria, Germany and Romania, taking up only 8 % of the total land available in the EU due to its high carbon content (%C in wet biomass) and higher yield compared to other crops. The feedstock distribution is visualized in the heat map in Figure 2. The optimal SC is centralized in Bulgaria and Poland. The solution is driven by the local costs, computed using the purchasing power parity metric, which stand below the European average. The biomass is firstly transported from most EU countries to Bulgaria and Poland mainly by train and, for what concerns the latter, also by ship. At these locations, the biomass is converted into pellets and then combusted. The CO_2 captured at the power generation plants is then distributed in the EU via pipeline and injected in suitable storage sites (depleted carbon fields or aquifers). The transportation of biomass from the land to the processing site and the CO_2 pipeline network are given in Figure 3.

The total impact on the metrics linked to the SDGs is obtained by summing up the impacts in Eq.(1) of the activities modeled in the entire supply chain. Then, we compare the performance of the SC with respect to the safe operating space assigned to the EU (PB^{EU}). We find that none of the PBs is transgressed in the EU, but significant burden-shifting occurs when reducing climate change impacts, as shown in Figure 4. Notably, respiratory inorganics (particulate matter) is the most critical impact, occupying roughly 50 % of the EU safe operating space. Non-cancer human health also shows a 17 % of the EU threshold. The negative implications on these categories can be mainly attributed to the pelleting activity and transportation, which were already identified as hotspots in Negri et al. by performing a standard LCA analysis. We recall that the model relies primarily on residues feedstock, leading to low impacts on water consumption and land use

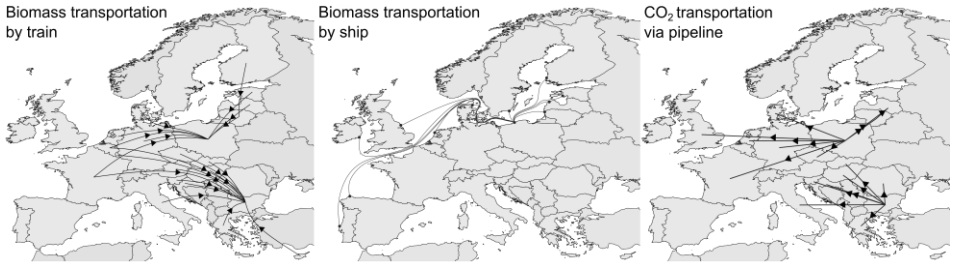


Figure 3. Links of transportation by train, by ship and CO_2 by pipeline among countries in the optimal supply chain. Intra-country connections are not represented.

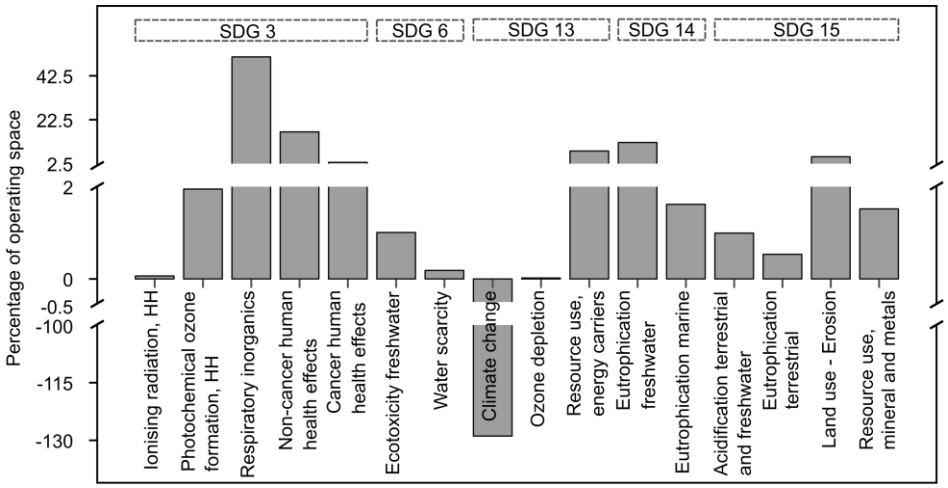


Figure 4. Total impact of the BECCS supply chain on the planetary boundaries considered. At the top we group the impacts per corresponding SDG according to Sala et al.

categories. Nonetheless, eutrophication freshwater also presents a non-negligible burden-shifting, with 14 % of the PB^{EU} occupied.

Contextualizing the results from an SDGs perspective, we find that improving climate action (SDG 13) worsens good health and wellbeing (SDG 3) due to respiratory inorganics and non-cancer human health effects. Similarly, life below water (SDG 14) worsens due to the impact on eutrophication freshwater. Lastly, the most considerable contributions affecting life on land (SDG 15) are land erosion and resource use, minerals and metals.

Conclusions

Here we studied the broad implications of deploying BECCS at a large scale on sustainable development. The minimum cost solution of a BECCS supply chain in the European Union was sought using the NETCOM optimization model, which was enlarged in scope to include planetary boundaries-based metrics. The total cost amounted to 74 billion Euros for a net removal of 0.61 tCO_2 in one year, without considering the credits from a net electricity production of 511 TWh. The optimal solution relies on residues and a centralized supply chain. The environmental assessment showed that the BECCS supply chain could operate within the safe operating space. Yet, delivering negative emissions to mitigate climate change inevitably exacerbates other indicators. This is particularly evident in the case of respiratory inorganics, where the BECCS supply chain took half of the share allocated to the European Union. Note that this safe operating space should be shared among all economic activities, such as chemical production and the energy sector, which raises concerns about the feasibility of BECCS at a large scale. Additionally, given that the yearly CO_2 target is reached using biomass residues, future research should investigate more ambitious removal and different types of feedstock, which could eventually lead to the transgression of the safe operating space. When looking at the effect on the Sustainable Development Goals, negative emissions benefited climate action (SDG 13) at the expense of good health and wellbeing (SDG 3) and life below water (SDG 14). Overall, this work shows that impacts should be quantified from

a life cycle perspective and considering absolute limits to interpret the results from a global sustainability viewpoint.

References

A. P. Barbosa-Póvoa, 2009, Sustainable Supply Chains: Key Challenges, Computer Aided Chemical Engineering, 27, 127-132

S. Budinis, S. Krevor, N. Mac Dowell, N. Brandon, A. Hawkes, 2018, An assessment of CCS costs, barriers and potential, Energy Strategy Reviews, 22, 61-81

J. G. Canadell and E. D. Schulze, 2014, Global potential of biospheric carbon management for climate mitigation, Nature Communications, 5,1-12

V. Heck, D. Gerten, W. Lucht, A. Popp, 2018, Biomass-based negative emissions difficult to reconcile with planetary boundaries, Nature Climate Change, 8, 151-155

M. Honegger, A. Michaelowa, J. Roy, 2020, Potential implications of carbon dioxide removal for the sustainable development goals, Climate Policy, 21, 678-698

IPCC, 2018, Mitigation Pathways Compatible with 1.5°C in the Context of Sustainable Development. In Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty

B. Mota, M. I. Gomes, A. Barbosa-Póvoa, 2013, Towards supply chain sustainability: balancing costs with environmental and social impacts, Computer Aided Chemical Engineering, 32, 895-900

B. Mota, A. Carvalho, M. I. Gomes, A. Barbosa-Póvoa, 2017, Sustainable supply chain design and planning: the importance of life cycle scope definition, Computer Aided Chemical Engineering, 40, 541-546

V. Negri, A. Galán-Martín, C. Pozo, M. Fajardy, D. M. Reiner, N. Mac Dowell, G. Guillén-Gosálbez, 2021, Life cycle optimization of BECCS supply chains in the European Union, Applied Energy, 298, 117252

G.P. Peters and O. Geden, 2017, Catalysing a political shift from low to negative carbon. Nature Climate Change, 7,619–621

J. Rockström, W. Steffen, K Noone, Å. Persson, F. Stuart III Chapin , E. Lambin, T. M. Lenton, M. Scheffer, C. Folke, H. Joachim Schellnhuber, B. Nykvist, C. A. de Wit, T. Hughes, S. van der Leeuw, H. Rodhe, S. Sörlin, P. K. Snyder, R. Costanza, U. Svedin, M Falkenmark, L. Karlberg, R. W. Corell, V. J. Fabry, J. Hansen, B. Walker, D. Liverman, K. Richardson, P. Crutzen, J. Foley, 2009, Planetary boundaries: Exploring the safe operating space for humanity, Ecology and Society, 14

S. Sala, E. Crenna, M. Secchi, E. Sany-Mengual, 2020, Environmental sustainability of European production and consumption assessed against planetary boundaries, Journal of Environmental Management, 269, 110686

P. Smith, J. Adams, D. J. Beerling, T. Beringer, K. V. Calvin, S. Fuss, B. Griscom, N. Hagemann, C. Kammann, F. Kraxner, J. C. Minx, A. Popp, P. Renforth, J. Luis Vicente Vicente, S. Keesstra, 2019, Land-Management Options for Greenhouse Gas Removal and Their Impacts on Ecosystem Services and the Sustainable Development Goals, Annual Review of Environment and Resources, 44, 255–86

United Nations, Sustainable Development Goals https://sdgs.un.org/goals.

D. P. van Vuuren, E. Stehfest, D. E. H. J. Gernaat, M. van den Berg, D. L. Bijl, H. Sytze de Boer, V. Daioglou, J. C. Doelman, O. Y. Edelenbosch, M. Harmsen, A. F. Hof, M. A. E. van Sluisveld, 2018, Alternative pathways to the 1.5 °C target reduce the need for negative emission technologies, Nature Climate Change 8, 391–397

J. Wheeler, Á. Galán Martín, F. D. Mele, G. Guillén-Gosálbez, 2020, Supply Chain Design Optimization within Planetary Boundaries, Computer Aided Chemical Engineering, 48, 1489-1494

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50104-4

A Multi-disciplinary Assessment of Innovations to Improve Grocery Bag Circularity

Kevin Dooley[a*], Vyom Thakker[b], Bhavik Bakshi[b], Matt Scholz[a], Fatima Hafsa[a], George Basile[a], Raj Buch[a]

[a]*The Global Kaiteki Center, The Julie Anne Wrigley Global Futures Lab, Arizona State University, Tempe, Arizona 85287, USA*
[b]*Department of Chemical & Biolmolecular Engineering, The Ohio State University, Columbus, Ohio 43210, USA*
Kevin.Dooley@asu.edu

Abstract

While demand for plastic increases because of its broad application base, the negative environmental consequences of plastic production must be minimized through effective value chain design. Plastic production creates GHG emissions, and its inadequate disposal can generate water or air pollution. Plastic packaging makes up over 40 percent of all plastic made, and within that category, plastic grocery bags have been a focal point for reduction of impacts. This paper explores the types of innovations needed to make grocery bags more circular, i.e., increased recycling and reuse. In similar studies, researchers have used one type of model or theoretical frame to address the question, such as life cycle assessment or economics. In this paper, we use the multi-disciplinary approach of convergence science to address this question. We consider a baseline scenario involving single-use plastic grocery bags, and then explore alternatives from the perspectives of life cycle assessment (LCA), policy, economics, and supply chain management. Integration of these perspective highlights the necessary interdependency of circular innovations needed to bring about systemic improvement.

Keywords: Plastic; Packaging; Sustainability; Circular Economy; Recycling

1. The need for more circular grocery bags

Society calling for the plastic economy to be more circular by reducing plastic usage and reusing or recycling the plastic that we consume (Ellen MacArthur Foundation, 2017). Plastics posea particularly significant circular economy challenge given their pervasiveness. Plastic production creates GHG emissions (Zheng & Suh, 2019), and its inadequate disposal can pollute local water sources and generate air pollution (Verma et al., 2016). These externalities can be mitigated by establishing a more circular economy for plastics, but less than 10 percent of plastic of all plastic ever produced has been recycled and data on plastic reduction and reuse efforts are scarce (Geyer et al., 2017).

Among different plastic products, plastic packaging constitutes 42 percent of plastic production and 47 percent of the plastic waste stream (Ritchie, 2018). Less than 20 percent of plastic packaging is recycled globally, the rest is either landfilled, incinerated, or leaked. Our previous research has identified seven different classes of circular-economy innovation: polymer design and production, packaging design, packaging reuse, packaging disposal, waste collection and sorting, waste treatment, and recycled content

use. An eighth type, policy, is added and may have implications at any of the value chain stages (Hafsa et al., 2020).

The question remains, which innovations will be most impactful or necessary to make the plastic packaging value chain more circular? In many similar studies, researchers have used one type of model or theoretical frame to address the question, such as life cycle assessment or economics. In this paper, we use the multi-disciplinary approach of convergence science to address this question.

To scope our research, we will apply this research question to the domain of grocery bags. We consider a baseline scenario involving single-use plastic grocery bags, and then explore alternatives, including different materials. We will use the following lenses to address the research question, and then synthesize by identifying common themes and disciplinary differences.

- Life cycle assessment (LCA) will be used to model the environmental trade-offs between different types of grocery bag materials, and the impact of different levels of reuse and recycling (Thakker and Bakshi, 2021).
- Policy analysis will be used to identify the regulatory and market mechanisms that have been used to reduce grocery bags' negative impacts and assess the effectiveness of these policies.
- Economic analysis will be used to assess the downstream market for recycled plastic grocery bags, which will vary by plastic type and attributes as well as region of the market.
- Supply chain management research will be used to understand the opportunities and constraints concerning the logistics of plastic bag collection, recycling, and remanufacturing.

2. Baseline model

The model that underlies our analysis is shown in Figure 1 (Thakker and Bakshi, 2021). At the top left, plastic feedstock is extracted, processed, and manufactured into grocery bags of four different possible plastic types: high density polyethylene (HDPE), low density polyethylene (LDPE), polypropylene (PP), or bioplastic polylactic acid (PLA). These are distributed to consumers (households) and may be one-use or multi-use. They are then collected either through waste management (trucks) or consumers (cars), and a certain number of bags are assumed lost to the environment. Then they are transported to a material recovery facility (MRF) where they are either landfilled; chemically recycled via pyrolysis; incinerated for energy; or segregated for composting, recycling, or upcycling to plastic lumber or cement clinker. This model has three decision points: which material portfolio to choose, how to collect the bags, and how to treat bags at end of life. From a system standpoint, one can use bag design to increase the number of uses per bag; policy and consumer education to incentivize more flow of bags into appropriate recovery channels; technologies at waste processing that minimize environmental impacts, especially related to energy use; more infrastructure to handle increased volume of flow through recycling channels; and more efficient production processes to make the bag itself.

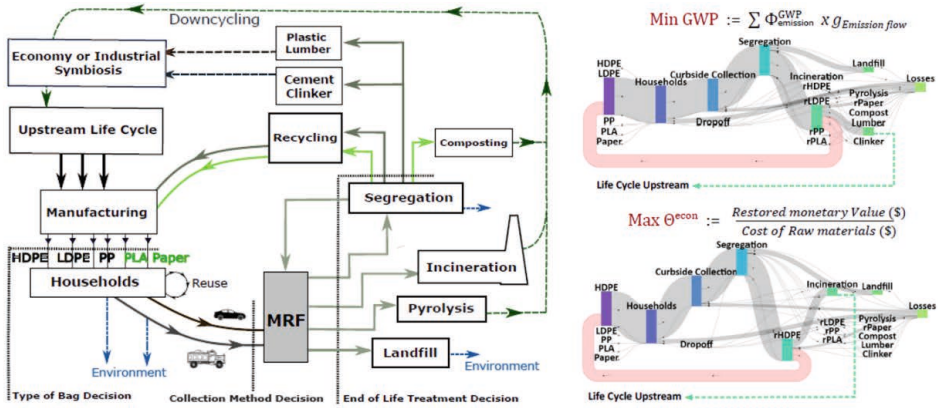


Figure 1 LCA optimization model of grocery bags

3. Life cycle assessment

Life cycle assessment and optimization methods were used to quantify the environmental impacts associated with different pathways and system designs. For any given objective function, the modeling yields an optimal bag combination and subsequent waste collection and management strategies, as shown in the Sankey diagrams in Figure 1. The optimal pathways are chosen among all life-cycle alternatives for objectives from three dimensions: circularity (Θ), environmental impact (Global Warming Potential, GWP), and cost (not shown). The trade-offs between objectives are quantified using pareto fronts, and optimality is identified. The conclusions from our modeling are:

- The scenario with least global warming potential and economic cost is the manufacture of reusable LDPE bags, followed by household littering to the environment without collection or treatment. This is a somewhat trivial answer, in that it would make little sense to accept a system that only littered to the environment; but since littering is an uncaptured externality in most LCA models, this solution highlights society's own incentives to minimize leakage.
- After considering the trade-offs between objectives of the three dimensions, the following value chain is found to be optimal under current technosphere conditions: Manufacture and use of reusable PP and LDPE, curbside collection by a truck, followed by segregation and recycling. The surplus generated recycled resin is used as cement clinker and incinerated.
- If adequate infrastructure existed to process more compostable material, the optimal shifts to a combination of corn-based PLA and PP bags. The shift of technosphere from current waste management scenario to the flexible technosphere yield win-win solutions in all three dimensions.

4. Policy

An array of policy interventions, both mandatory and voluntary, have been enacted or considered for reducing the consumption of plastic grocery bags and subsequent litter and waste. Policy makers have targeted governments, producers, retailers, consumers, and

waste managers with diverse measures that include prescribed product design specifications, taxes, subsidies, bans, quotas, awareness campaigns, and extended producer responsibility (EPR) programs, wherein producers individually or collectively subsidize the cost of managing the final disposition of their products. A substantial body of empirical literature—and indeed much theoretical literature—has been developed around these topics, and several reviews can be consulted (e.g., Abbott, 2019).

Taxes on plastic carrier bags have been implemented by at least 30 countries spread across all habitable continents and are often coupled to campaigns to promote the adoption of reusable bags. While taxes can be assessed at the producer level, they typically take the form of levies that are either imposed upon or voluntarily adopted by retailers, who may pass the fees on to consumers at the point of sale. Programs vary in the cost of the fee, its frequency, and on what thickness of bag it is assessed. An oft-cited example of a popular and successful levy program is the Irish PlasTax, which was originally levied in 2002 and led to use reduction on the order of 90% and marked decreases in plastic bag litter (Convery, 2007).

Bans are equally ubiquitous and appear to be more common than taxes in Africa, Australia, and the US. These can take several forms, but outright bans on plastic bags at point-of-sale appear most common (United Nations Environment Programme, 2018). Often these bans permit thicker bags, which are presumed reusable, and there are frequent exemptions for certain uses, such as wrapping fresh meat. Soft bans also exist where consumers only receive bags when they specifically request them. Another type of ban is the landfill ban that prohibits the disposal of plastic waste in landfills—a type of ban that has become popular in many northern European and Scandinavian countries (Steensgaard, 2017).

Perhaps the most concise statement of the effectiveness of various policy interventions for reducing the environmental impacts of plastic carrier bags is provided by the IMF: “No single policy approach is ideal for all contexts, and regulatory and economic instruments can serve as complements as well as substitutes” (Matheson, 2019). The available data seem to support the effectiveness of bans and taxes on plastic bag use at reducing their consumption, litter, and waste.

5. Economics

The economic incentives to collect and process used grocery bags must be sufficient to overcome the default of simply disposing them in landfill. For other types of products or packaging materials, such a metal or paper fibre, the economic incentives are high, in that recovering the material is so much less expensive than the cost of making it anew; in other words, recycled content is less expensive than virgin content. There are also functional requirements – the recycled content must be of high enough quality that it can meet the engineering needs of its use. Finally, recycled content is more valuable if it has fewer impurities

These desired objectives are not well met by the current plastic recycling market. In part because of the pandemic and in part due to a reduced cost of petroleum – the feedstock most often used to make virgin plastic – plastic recycling prices for some types of plastics became inverted in 2021. Data was collected from the publication Plastic News, and in June 2021, the price of virgin HDPE in the U.S. was around 85 cents per pound, while scrap clear HDPE was 110 cents per pound. On one hand, this does increase profit per pound to companies in the recycling stream. However, the much lower cost of purchasing

virgin HDPE leads to less use and demand for recycled HDPE, so overall profits decline as volumes decrease.

Our LCA studies suggest that because of HDPE's larger environmental impact, a more sustainable portfolio would consist of either PP and LDPE bags or PLA bags. Unfortunately, these solutions run up against economic constraints. PP and LPDE are both materials where recycled content costs less than virgin content, as expected. However, their recovery is so low that it is a disincentive to communities to collect it. In the U.S., PP has historically been collected curbside, but the poor recycling value has caused many cities to stop collection.

Economics also do not currently support a move to PLA bags. Until there is broad infrastructure to separate and collect compostable material, too many compostable bags would end up in recycling bins, thus acting as a contaminant, and reducing the margins associated with collecting recyclable material.

6. Supply chain management

Modern waste management systems are designed and operated on the basis that large volumes need to be collected and aggregated to take advantage of strong economies of scale. Large capital investments are required for infrastructure, and the impact of the large, fixed cost is reduced as volume of waste managed increases. On the revenue side, especially for plastic recycled content, margins are small, so large volumes must be amassed to waste management businesses to make sufficient total profit.

There is also a conflict between the material pathways that LCA modelling prescribes and the constraints that supply chain management issues pose with prescribed materials. Because of the relatively higher rate that HDPE is collected curb side in the US, compared to LDPE and PP, the latter suffer from the need to aggregate even greater volumes of recycled content, which leads to increasing logistics costs and environmental impact. Plastic in general, and LDPE specifically, are also less attractive to material handling companies such as warehouses because the value of the recycled content relative to its physical footprint is poor. Per tonne, it takes up more volume in the truck or warehouse but has less value per storage unit.

7. Conclusion

In conclusion, as one would expect in any complex system, there is no single or dominant optimal solution. LCA modelling demonstrates that different stakeholders, represented by different utility or objective functions, will assess different circular solutions and material pathways in different ways. Maximizing circularity will not necessarily minimize environmental impacts or direct material costs. Likewise, the optimality of many solutions depends on a bundle of innovations to be implemented. For example, the attractiveness of PLA based solutions increases as we have more composting infrastructure available.

When we combine multiple disciplinary perspectives to this problem, we further highlight the trade-offs that exist between any decision pathways. LCA modelling suggests that PP and LDPE may be strong alternatives to HPDE for grocery bag. But economic and supply chain analysis suggest that because of recycled HDPE's larger price per tonne, it may still be a better solution for companies in the recycling sector. Policy analysis suggests that there are pre-competitive mechanisms that can yield either consumers or manufacturers

to be more committed to collecting recyclable packaging, but that the success of such policy mechanisms may be different for different packaging applications, formats, material types, and regions. We encourage other scholars and practitioners to bring an interdisciplinary lens when exploring solutions for a more circular economy.

In this study, we developed a proprietary LCA optimization model to examine the impact of different circular innovations, and then qualitatively examined whether other perspectives aligned with or contradicted the conclusions from the LCA model. In the future, the LCA model can be expanded to take these additional dimensions (e.g., economics, policy) directly into account within the computational model.

References

Abbott, J.K., Sumaila, U.R., 2019. Reducing Marine Plastic Pollution: Policy Insights from Economics. Review of Environmental Economics and Policy 13, 327–336. https://doi.org/10.1093/reep/rez007

Convery, F., McDonnell, S., Ferreira, S., 2007. The most popular tax in Europe? Lessons from the Irish plastic bags levy. Environ Resource Econ 38, 1–11. https://doi.org/10.1007/s10640-006-9059-2

Ellen MacArthur Foundation, 2017. The New Plastics Economy: Rethinking the Future of Plastics & Catalysing Action. Ellen MacArthur Foundation.

Geyer, R., Jambeck, J., & Law, K. L., 2017. Production, use, and fate of all plastics ever made. Science Advances, 3(7), e1700782. https://doi.org/10.1126/sciadv.1700782

Hafsa, F., Dooley, K., Basile, G., and Buch, R., 2020, Innovation pathways for more circular plastic packaging, AIChE Sustainable Packaging Conference.

Matheson, T., 2019. Disposal is Not Free: Fiscal Instruments to Internalize the Environmental Costs of Solid Waste. IMF Working Papers 19. https://doi.org/10.5089/9781513521589.001

Steensgaard, I.M., Syberg, K., Rist, S., Hartmann, N.B., Boldrin, A., Hansen, S.F., 2017. From macro- to microplastics - Analysis of EU regulation along the life cycle of plastic bags. Environmental Pollution 224, 289–299. https://doi.org/10.1016/j.envpol.2017.02.007

Thakker, V. and Bakshi, B., 2021, Toward sustainable circular economies: A computational framework for assessment and design, Journal of Cleaner Production, Vol. 295, https://www.sciencedirect.com/science/article/abs/pii/S0959652621005734

United Nations Environment Programme, 2018. Single-Use Plastics: A Roadmap for Sustainability. Available at https://www.euractiv.com/wp-content/uploads/sites/2/2018/06/WED-REPORT-SINGLE-USE-PLASTICS.pdf.

Verma, R., Vinoda, K. S., Papireddy, M., & Gowda, A. N. S., 2016. Toxic pollutants from plastic waste-a review. Procedia Environmental Sciences, 35, 701–708. https://doi.org/10.1016/j.proenv.2016.07.069

Zheng, J., & Suh, S., 2019. Strategies to reduce the global carbon footprint of plastics. Nature Climate Change, 9(5), 374–378. https://doi.org/10.1038/s41558-019-0459-z

Acknowledgements

The authors gratefully acknowledge the support of the research that led to this publication, from "The Global KAITEKI Center" at Arizona State University (ASU), a university-industry partnership between ASU and The KAITEKI Institute of Mitsubishi Chemical Holdings Corporation.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50105-6

Process Sustainable Supply Chain: integrating monetization strategies in the design and planning

Cátia da Silva[a]*, Ana Barbosa-Póvoa[a], and Ana Carvalho[a]

[a]*CEG-IST, Instituto Superior Técnico, University of Lisbon, Lisbon 1049-001, PORTUGAL*
[*]*catia.silva@tecnico.ulisboa.pt*

Abstract

Several wrong human actions have compromised future generations leading to growing concerns on environmental and social issues. As a result, the industry became aware of the need to integrate these concerns into its decision-making process. This is crucial within the process supply chains (that deal with process design with the aim of converting raw materials into final products), due to the type of process and products that they deal with. Though, quantifying both environmental and social impacts of the supply chain is not easy and understanding their units is challenging, particularly for decision-makers. In this way, this work intends to develop an optimization-based decision support tool that: i) models possible decisions taken throughout the supply chain, while considering demand uncertainty; ii) translates both environmental and social impacts of the supply chain into the same monetary unit so as to optimize the design and planning of economic, environmental, and social performances of the supply chain in the same unit. Considering that decision-makers are used to dealing with money when managing their supply chain, this innovative decision-support tool simplifies the decision-making process as all supply chain performances are quantified in an understandable monetary unit, which can constitute an asset to inform decision-makers. This innovative decision support tool is validated considering a real case study of a process supply chain.

Keywords: Sustainable supply chain; Design and planning; Optimization; Monetization; Uncertainty.

1. Introduction

All companies want to improve their supply chain (SC) as only through its effective management is it possible to obtain beneficial results for the company. Furthermore, in the current situation of great competition among companies, an optimized management of the SC is essential for complete customer satisfaction. In this scenario the main focus of companies ceased to be exclusively an economic performance, but environmental and social issues are very important to the proper functioning of SCs. In fact, the integration of the three pillars of sustainability (i.e., economic, environmental, and social) in the SC management has become essential (Barbosa-Póvoa et al., 2018). The World Commission on Environment and Development recognized sustainable SCs as a form of optimizing customer value and reach a market's competitive advantage. However, SC management concerning sustainable and effective goals is challenging, particularly when its design and planning are considered. Considering that sustainability involves a focus on the

economic, environmental, and social performance, the most challenging are the ones associated with environmental and social quantification. In fact, the economic performance of the SC has been evaluated by decision-makers for a long time and is the one that they find it most easy to understand. On the other hand, environmental and social performances are not easily perceived by decision-makers, and its quantification often does not help this understanding. Thus, monetization can help in understanding these abstract impacts, as it allows to translate environmental and social impacts into monetary units, which facilitates the perception of the value of these abstract impacts in the SC and can facilitate the management of decision-making. It is also important to highlight that besides studying SC economic and social impacts, the major focus of this work will be on environmental aspects as they are truly critical within the process SC. This paper presents a MILP that accounts for the economic, environmental, and social concerns in the same objective function by monetizing both environmental and social impacts and considering demand uncertainty.

2. Environmental and social monetization methodologies

Regarding the environmental impacts' assessment, it appears that the life cycle assessment (LCA) is the most used methodology in the literature. LCA is composed by four main phases. The first one includes the goal and scope definition, where the context of the study is set out. The second phase is related to inventory analysis, which involves creating flows' inventory from and to nature for a system. The third step is the life cycle impact assessment (LCIA) phase that aims to evaluate the significance of potential environmental impacts based on the life cycle inventory flow results. However, this is a critical phase, where decision-makers must assign weights to factors, which may not be easy for most decision-makers. In addition, there is many environmental quantification methods, which also makes it difficult to choose the best one to use. Among the methods that exist, it appears that some of them quantify environmental impacts in abstract units or scores, while others monetize them. Monetizing means quantifying the impact in monetary units. Although there are many methods capable of translating environmental impacts into money, the European Commission (2010) considers EPS 2000 to be a very adequate and complete method when compared to other LCIA methods, having its uncertainties fully specified (Steen, 2000). The fourth phase is the life cycle interpretation that allows the identification, quantification, check and evaluation of the information from the LCA results.

Considering the social impacts' quantification, many works related to sustainable SC focus mainly on the environmental pillar while neglecting the social one. In fact, social performance of the SC has been the least explored one, which resulted in a relevant research gap in this area. In this set, companies have difficulties to assess and quantify their social performance (Beske-Janssen, Johnson, and Schaltegger 2015). Generally, social indicators are associated with safety, health, human rights, community initiatives, child labour, labour issues, and employment benefit. Global Reporting Initiative (GRI) identified labour practices and decent work, human rights, society, and product responsibility as important categories in the social component. It is important to mention that there is no consensus in the scientific community, particularly when defining social indicators. In this way, the identification of the suitable social indicators to be applied within the SC and its quantification is urgent to support the decision-making process. For this reason, translating social impacts into monetary units so that they are easily

understood by decision-makers can be an important contribution to explore these research gaps.

3. Problem description and model characterization

This problem considers a generic SC that includes the flow of suppliers in which raw materials are sent to factories and the final products are obtained. Final products can move to warehouses or directly to markets. At markets, the end-of-life products can be recovered and sent to warehouses or directly remanufactured in factories. Given the possible set of locations of SC entities, production and remanufacturing technologies, possible transportation modes between entities, and products within the SC, the main objective is to obtain the SC network structure, supply and purchase levels, entities' capacities, transportation network, production, remanufacturing and storage levels, supply flow amounts, and product recovery levels, to maximize profit and social performance, while minimizing environmental impact. To solve this problem a MILP model was developed, which is based on da Silva et al. (2020). This model was extended to also consider the impact of SC social performance. Eq. (1) represents the first objective function, which is the maximization of the expected economic, environmental, and social performances of the SC, namely the expected net present value (eNPV), the expected social impact (eSoImpact), and the expected environmental impact (eEnvImpact). The economic performance is assessed through the eNPV (represented by Eq (2)), which is obtained through the sum of each node's probability multiplied by the discounted cash flows (CF_{Nt}) in each period *t* and for each node *N* at a given interest rate (*ir*). These CF_{Nt} are obtained from de net earnings (difference between incomes and costs). There are several costs included, namely raw material costs, product recovery costs, production/remanufacturing operating costs, transportation costs, contracted costs with airline or freight, handling costs at the hub terminal, inventory costs, and labour costs. In addition, for the last time period, it is considered the salvage values of the SC (FCIγ). The environmental performance is assessed through the eEnvImpact (represented by Eq. (3)), LCA is performed on the transportation modes and on entities installed in the SC boundaries, using EPS 2015. The Life Cycle Inventory is retrieved from the Ecoinvent database (through SimaPro 8.4.0 software). The LCA results are expressed in Environmental Load Units (ELU) and used as input data (*ei*) in Eq. (3), particularly in the environmental impact of transportation (first term), and entity (second term). Moreover, considering that the main focus of European Commission is to bet on promoting job creation and regional development, the social performance here considered is represented by Eq. (4), where $\frac{LProd_t}{GProd_t}$ corresponds to the ratio between labour productivity and global productivity; GDP_i is the gross domestic product of a country or sector where the SC is inserted; α_i represents an impact regional factor, which can assume different values according to the intended purpose of the study; and Y_i is a binary variable (which returns 1 if entity i is opened). Global productivity is given by $GProd_t = \sum_j \frac{GDP_j}{NEmpl_j}$ and labour productivity is given by $LProd_t = \sum_i \sum_j \frac{€prod_{ij}}{Nwor_{ij}}$ where j is a given country or sector where the SC is inserted; GDP_j corresponds to the gross domestic product of a country or sector j where the SC is inserted; and $NEmpl_j$ is the number of people employed in country or sector j where the SC is implanted and $Nwor_{ij}$ is the number of workers that the company or the SC owns. The model also considers several constraints regarding mass balance,

capacity, transportation, and technology (for more details, please see da Silva et al. (2020)).

$$\max\ (e\,\mathrm{NPV}_N + e\,\mathrm{So\,Impact}_N - e\,\mathrm{Env\,Impact}_N) \tag{1}$$

$$\mathrm{eNPV_N} = \sum_N pb_N \left(\sum_{t\in T} \frac{CF_{Nt}}{(1+ir)^t} - \sum_\gamma FCI_\gamma\right) \tag{2}$$

$$\mathrm{eEnvImpact_N} = \sum_N pb_N \left(\sum_{\substack{t\in T \\ (a,m,i,j)\in NetP}} ei_{ac}\, pw_m d_{ij} X_{maijtN} + \sum_{i\in I_f \cup I_w} ei_{ic} YC_i\right) \tag{3}$$

$$\mathrm{eSoImpact_N} = \sum_N pb_N \left(\sum_i \sum_j \frac{L\,\mathrm{Prod}_t}{G\,\mathrm{Prod}_t} \cdot \frac{1}{GDP_j} . \alpha_i . Y_i\right) \tag{4}$$

4. Case study

The model is applied to a chemical components' producer located in Lyon, France (Silva et.al., 2020). The company's suppliers are local and placed in Lyon. It supplies three main markets that are in different European countries: Portugal, France, and Germany. France is the market that owns the highest percentage of company's sales (38.9%), followed by Germany (33.7%), and Portugal with 27.4% of company's sales. Considering the willingness of the company's decision-makers to expand its SC, to three new markets (Ireland, Spain, and Canada), the company wants to know what changes could result from this expansion in financial terms. Currently, transportation is only performed by road, namely by truck. However, with the expansion, there is the need to combine road transportation with air and sea transportation modes. Regarding company's characterization, it is important to mention that it sells three different types of chemical products (fp1, fp2, fp3) that can be sold within the chemical industry and to other industries. In the current production, three technologies (pr1, pr2, pr3) are used that produce respectively products fp1, fp2, and fp3. Furthermore, end-of-life products can be recovered and remanufactured into final products. This work accounts for product's demand uncertainty through a stochastic approach since this method allows the discretization of stochastic data over the time horizon and can be adjusted during the planning horizon. A scenario three was considered (Figure 1), where node N characterizes a possible state and the arcs represent the evolutions it may have. Each node has a specific probability and a path from the root to a leaf node represents a scenario.

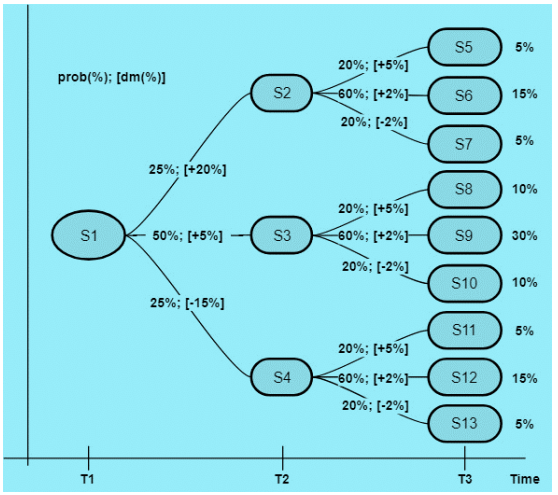


Figure 1- Scenario tree - values for probability and demand (D) variation are represented

4.1. Results

Initially, the model was solved considering each objective function individually, which means that the three goals are considered

separately. In addition, the three goals were also analysed within the same function (single objective function). Thus, four cases are studied to understand the results obtained with the proposed approach:

- Case A: analyses the case-study results that represents the optimal economic performance as major goal;
- Case B: analyses the case-study results that represents the optimal social performance as major goal;
- Case C: analyses the case-study results that represents the optimal environmental performance as major goal, evaluated through EPS;
- Case D: analyses the case-study results that represent the optimal trade-off between social, environmental, and economic performances.

As a result, the values for the total expected net present value, social impact, and environmental impact obtained in each case are shown in Table 1.

Table 1 Outcomes for the economic, social, and environmental impacts.

Obj. Function	**Case A** Max eNPV	**Case B** Max eSoImpact	**Case C** Min eEImpact	**Case D** Max (eNPV + eSoImpact - eEnvImpact)
eNPV	**$1.5203x10^{9}$ €**	$1.4905x10^{9}$€	$1.1979x10^{9}$ €	$1.5098x10^{9}$ €
eSoImpact	$8.1700x10^{-12}$€	**$3.1150x10^{-11}$€**	$1.6100x10^{-11}$€	$5.2567x10^{-12}$ €
eEImpact	$1.0149x10^{9}$ €	$1.1057x10^{9}$€	**$9.3470x10^{8}$ €**	$9.8220x10^{8}$ €

Regarding the results, it is noted that case A is associated with the biggest value for eNPV, which arises from the fact that its objective function is maximized, while returns the lowest social benefit. Case B considers the maximization of the social performance and has the highest value for eSoImpact that corresponds to $3.1150x10^{-11}$€ and the lowest value of economic performance. In contrast, case C that consider the minimization of the eEImpact (assessed through EPS method) has the lowest eNPV value and an expected environmental impact that corresponds to $9,3470x10^{8}$ €. Finally, case D that considers the three performances simultaneously, which is only possible thanks to the social and environmental impacts' monetization process. It can be seen that global profit is decreased in 1 million euros when comparing cases A with D, but the impact of environment increased by nearly 32.7 million euros that can only be accomplished by using the single objective approach. On the other hand, the value achieved for social performance in case D, is between the values of case A and B. Furthermore, it is possible to note that global profit improved further than 300 million euros between cases C and D, whereas the impact of environment is worst at nearly 48 million. In other words, with the aim of maximizing the global profit, the single objective approach allows to improve 3.33% in environmental performance and to improve 55% in the social performance, while ENPV only decreases 0.7%. Table 2 shows SC design decisions considering the results depicted in Table 1 for case D. Additionally to the existent factory, one new factory is installed. The results also show that there is the need to expand the existing capacity by opening new warehouses. Regarding suppliers' allocation, all factories are supplied by all suppliers, which results from the balance between the lower costs of raw materials and fewer transportation costs. Regarding transportation, the truck with more capacity (Truck2) is preferred in all cases since this has a lower environmental impact. In terms of intermodal transportation, sea option is preferred, while air transportation is not used.

Table 2 Supply chain network structure for case D

Allocation of suppliers		Closest suppliers supply all factories plus Bremen supplier
Factories		Lyon and Galway
Warehouses		Lyon
Production		The majority of fp1 production is in the current factory
		The majority of fp2 production is shared between factories
		Most production of fp3 is in the opening factories
Remanufacture		Most of rp1 is in the existing facility (around 70%)
		rp2 and rp3 are only performed in the existing facility
Technologies		Alternative one is preferred for fp1 and fp2 in the opening factories
Inventory	fp1	43% in Lyon and 57% in Galway
	fp2	65% in Lyon and 35% in Galway
	fp3	30% in Lyon and 70% in Galway
Transportation		Road: 8 trucks; Air: Not adopted; Sea: Adopted

5. Conclusions

This study intended to develop a decision support tool for the design and planning of a generic SC, where the economic, environmental, and social performances are considered under uncertainty on the product's demand. Economic performance is assessed by using the expected NPV. In addition, environmental performance was evaluated through the LCA methodology. Considering the social performance, job creation was the indicator considered in our first social approach. This analysis was only possible due to the monetization, which was able to quantify both environmental and social impacts in a monetary unit. This allowed to include in the same objective function the economic, the environmental and the social impacts. Moreover, from the analysis made, it was clear that results are influenced by social and environmental impacts, and this proves the importance of considering these issues in solving real-life problems. For future work, further research should be done on this topic to better explore monetization approaches to be a reliable alternative to evaluate environmental and social impacts. Also, an extension of this work should consider different social indicators and a more comprehensive study of uncertainty to conclude on its adequacy even better.

Acknowledgements

The authors acknowledge the support provided by FCT and PORTUGAL2020 under the project PTDC/EGE-OGE/28071/2017, Lisboa -01.0145-Feder-28071.

References

A. Barbosa-Póvoa, Cátia da Silva, and Ana Carvalho. 2018. “Opportunities and Challenges in Sustainable Supply Chain: An Operations Research Perspective.” European Journal of Operational Research 268 (2): 399–431.

P. Beske-Janssen, Matthew Phillip Johnson, and Stefan Schaltegger. 2015. “20 Years of Performance Measurement in Sustainable Supply Chain Management – What Has Been Achieved?” Supply Chain Management: An International Journal 20 (6): 664–80.

C. da Silva, Ana Barbosa-Póvoa, and Ana Carvalho. 2020. “Environmental Monetization and Risk Assessment in Supply Chain Design and Planning.” Journal of Cleaner Production, 270, 121552.

European Commission. 2010. “Joint Research Centre – Institute for Environment and Sustainability: International Reference Life Cycle Data System (ILCD) Handbook – General Guide for Life Cycle Assessment – Detailed Guidance. Publications Office of the European Union”.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50106-8

Systematically Identifying Energy-Efficient and Attractive Multicomponent Distillation Configurations

Tony Joseph Mathew[a], Mohit Tawarmalani[b], Rakesh Agrawal[a*]

[a]*Davidson School of Chemical Engineering, Purdue University, West Lafayette, IN-47907, USA*
[b]*Krannert School of Management, Purdue University, West Lafayette, IN-47907, USA*
agrawalr@purdue.edu

Abstract

Thousands of configurations exist for multicomponent distillation, making it laborious to use standard process simulators for identifying which among this plenitude are energy-efficient for a given separation. Shortcut models quickly screen the wide search space, but their development has been limited by various obstacles. In this work, we overcome three challenges: assumptions of constant relative volatilities and constant molar overflow, and utilizing heat integration. We incorporate our solutions into an optimization formulation and subsequently demonstrate its ability to identify energy-efficient and heat-integrated configurations on a case study. We also demonstrate how process intensification can be used to raise the value of the selected configuration.

Keywords: Multicomponent Distillation; Optimization; Process Synthesis; Process Intensification.

1. Introduction

Ubiquitous in the chemical and petrochemical industries, distillation is a staple unit operation in the separation of various mixtures such as crude petroleum, air, natural gas liquids, alcohols, and aromatics. However, these separations come with substantial energy expenses which have a by-product of greenhouse gas emissions. Therefore, it is of vital importance to identify new distillation arrangements with lower energy consumption for separations.

For multicomponent separations, there is a vast search space of configurations possible. For example, 6,128 configurations exist for separating just five components in a mixture (Shah and Agrawal, 2010). A systematic analysis (with optimization) and comparison is required to determine which of these are energy-efficient for a given separation. But performing such an analysis in a standard process simulator is impractical due to the computational challenges of optimizing with the complex thermodynamics and exorbitant time spent in evaluations (Madenoor Ramapriya et al., 2018). Hence, we use a simplified model for energy optimization to quickly screen through the vast search space of configurations, identifying a handful of candidates for further analysis in the process simulator.

Although several simplified models exist for distillation in the literature (Caballero and Grossmann, 2006; Nallasivam et al., 2016; Tumbalam Gooty et al., 2018; Ryu and Maravelias, 2020), they make various assumptions which, while granting them

computational simplicity, limit their accuracy. For example, the Underwood method (Underwood, 1948) is often used to estimate the minimum reflux in columns. However, the method was derived based on two vital assumptions, constant relative volatilities (CRV) and constant molar overflow (CMO), which do not hold for many real separations. Moreover, these models face computational difficulties incorporating important elements of process flowsheets. For example, heat integration (the re-use of heat) is invaluable for energy reductions. However, its feasibility checks can hamper global optimization when complex equations are used to calculate temperatures. In this paper, we present our solutions to the aforementioned limitations and demonstrate the optimization formulation in which they have been incorporated to identify energy-efficient and heat-integrated distillation configurations. We also demonstrate process intensification techniques to improve the attractiveness of the chosen configuration.

2. Optimization Model

2.1. Base Model

First, we describe the base model of Nallasivam et al. (2016), upon which we have incorporated our advances. Implemented in a nonlinear program, it determines the minimum vapor duty (sum of the reboiler vapor duties) of regular-column (n columns for n-component separation) distillation configurations. It employs assumptions of ideal and zeotropic mixtures, CRV, and the standard McCabe-Thiele assumptions (which result in CMO). Each configuration is uniquely identified by the set of streams present. Columns with multiple feeds are modelled as a collection of pseudo-columns for the split of each feed stream. These pseudo-columns are stacked vertically, with the common products of adjacent splits being withdrawn as sidedraws.

The variables in the base model are component and vapor flowrates of each stream and column section, and the Underwood roots of each split. Mass and vapor balances are applied at the product and feed ends of each stream. The Underwood method (Underwood, 1948) is utilized to constrain minimum vapor flows in columns and ensure feasible component distributions in sloppy splits. The vapor flowrate of a stream is constrained according to its nature, viz., associated with a heat exchanger, sidedraw, or thermally coupled.

2.2. Estimating Better Relative Volatilities

The relative volatility of component i, α_i, is a measure of how this component distributes between the phases of a vapor-liquid equilibrium (VLE) compared to the heaviest component n in the mixture. When α_i are constants (the direct result of the CRV assumption), the following equation (written to predict vapor composition, y_i, from liquid composition, x_i) is sufficient to characterize the VLE.

$$y_i = \frac{\alpha_i x_i}{\sum_{j=1}^{n} \alpha_j x_j} \tag{1}$$

But in general, component relative volatilities vary with the specifications of the mixture, i.e., composition, temperature, and pressure. For this reason, CRV (and thereby the above simplified VLE) has been viewed as a poor assumption in distillation models. However, even if the physical values of relative volatilities vary widely, there can exist constant mathematical values which capture these variations. Instead of the common perception that Eq.(1) is only valid for ideal mixtures, we advocate interpreting it as a

surrogate VLE. α_i are then to be determined by parameter estimation, so that this surrogate VLE best approximates the true VLE (Anderson & Doherty, 1984).

The method we propose for estimating α_i is as follows (Mathew et al., 2020). Generate a systematic spread of VLE training data in x_i and y_i, by considering each component to be either lean or rich, as well as the pressure to be low or high. Then regress α_i in Eq.(1) via non-linear least squares. This yields values for α_i which can capture the VLE through Eq.(1), and therefore be employed in methods assuming CRV, such as the Underwood method, while accounting for the variations in relative volatilities.

2.3. Relaxing Constant Molar Overflow

CMO is valuable in shortcut models as it simplifies complex energy balances over enthalpy into simple linear balances over vapor and liquid flowrates. A critical requirement for CMO is that all components have the same latent heat of vaporization, λ_i. However, λ_i can have a wide spread in many separations, such as for crude oil. To account for different latent heats, researchers such as Mole (1950) have derived the following simple variable transformation from molar variables to latent heat variables ($^{\mathrm{LH}}$) for component flowrate (f), total flowrate (F), and composition (z).

$$f_i^{\mathrm{LH}} = f_i\lambda_i, \quad F^{\mathrm{LH}} = F\sum_{j=1}^{n} z_j\lambda_j, \quad z_i^{LH} = \frac{z_i\lambda_i}{\sum_{j=1}^{n} z_j\lambda_j} \tag{2}$$

The benefit of this particular transformation is that, when the different latent heats are accounted for, the resulting model written in latent heat variables is mathematically equivalent to the usual CMO model in molar variables. Thus, the transformation allows us to implicitly account for different latent heats but while retaining the computational simplicity of the CMO equations. We applied the above transformation in our formulation to not only relax the CMO assumption but also determine the minimum heat duty of configurations, which is a more accurate proxy for energy consumption than vapor duty which the base model determines.

2.4. Heat Integration to Reduce Energy Consumption

One opportunity for heat integration in distillation is to transfer heat from the condenser of one column to the reboiler of another. Such a heat transfer is allowed only if the temperature of the source (condenser) is greater than that of the sink (reboiler). This implies that temperatures would need to be calculated during optimization while checking for feasible heat integrations. However, such checks can hamper convergence to global optimality when complex thermodynamic equations are used. To address this, we developed a shortcut criterion for feasibility of heat integration (Mathew et al., 2021).

Using Raoult's and Dalton's laws for partial pressures, along with Eq.(1), we derived a new metric (pressure-scaled pseudo relative volatility, ρ) that monotonically decreases with temperature. Therefore, we utilize ρ as an inverse proxy for temperature which is computationally cheaper to calculate during optimization. The shortcut criterion is shown below. ρ is first calculated, in Eq.(3), for each stream at its bubble and dew points, using only component relative volatilities (α_i), composition (z_i), and pressure (P). Then, heat transfer (Q) is permitted via Eq.(4) only if ρ of the condensing stream (c) is less than that of the boiling stream (b), implying that the condensing stream is hotter than the boiling stream.

Table 1: Feed data for case study. Feed is saturated liquid at 1 atm.

	Components				
	A	B	C	D	E
Species	Benzene	Toluene	Ethylbenzene	p-Xylene	o-Xylene
Feed Flowrate, f (kmol/h)	30	30	5	5	30
Relative Volatility α	5.61	2.43	1.25	1.17	1
Latent Heat λ (kJ/mol)	30.0	32.5	34.7	34.9	35.7

$$\rho^{\text{bub}} = \frac{1}{P}\sum_{j=1}^{n} \alpha_j z_j, \quad \rho^{\text{dew}} = \left[P \sum_{j=1}^{n} \frac{z_j}{\alpha_j}\right]^{-1} \tag{3}$$

$$Q\left(\rho_c^{\text{bub}} - \rho_b^{\text{bub}}\right) \leq 0, \quad Q\left(\rho_c^{\text{dew}} - \rho_b^{\text{dew}}\right) \leq 0 \tag{4}$$

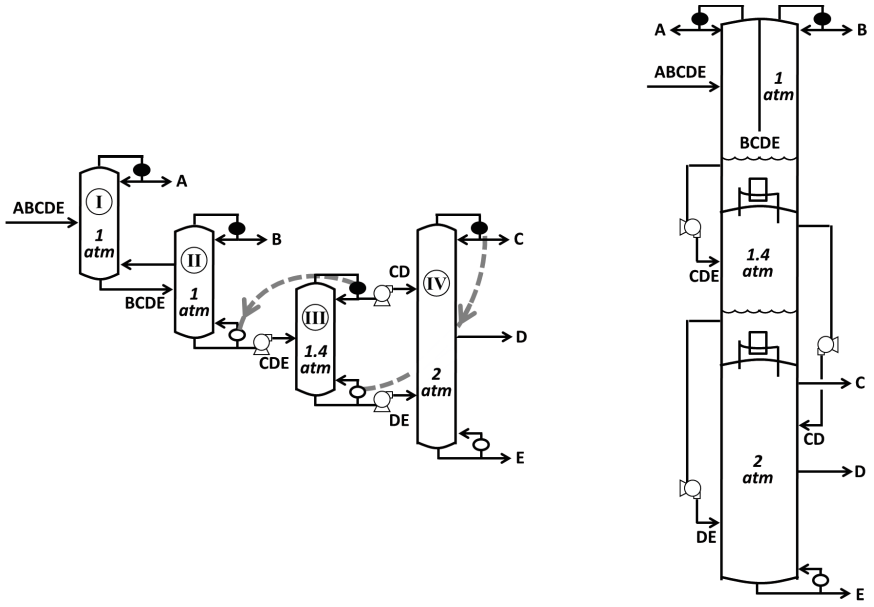
3. Case Study

We now demonstrate our formulation on a five-component aromatics separation, with feed data shown in Table 1. We solved our shortcut model for each of the 6,128 different regular-column configurations possible for this separation (Shah and Agrawal, 2010), while considering heat integration opportunities and allowing column pressures to vary between 1 atm and 2 atm.

The configuration in Figure 1(a) was predicted to have the least heat duty of 1.82 MW among all the configurations optimized. It uses two heat integrations of condenser (filled circle) of stream CD supplying heat to the reboiler (hollow circle) of stream CDE, and condenser C supplying heat to reboiler DE. To make these heat integrations feasible, columns 1 and 2 are operated at the feed pressure of 1 atm while columns 3 and 4 are operated at higher pressures of 1.4 atm and 2 atm respectively.

To improve the structural attractiveness of this configuration, we apply a variety of transformation techniques which fall under the broad umbrella of process intensification (PI) (Jiang and Agrawal, 2019). First, we horizontally consolidate columns I and II into a single shell with a vertical partition in between, forming a dividing wall column (DWC) (Agrawal, 2001; Madenoor Ramapriya et al., 2018b). Then, the three column shells are vertically consolidated into a single shell (I-II, III, then IV), with the heat integrations performed internally using multi-effect technology (Agrawal, 2000).

The final configuration is shown in Figure 1(b). An important feature of the transformation techniques we employed is that they maintain thermodynamic equivalence. Thus, the process-intensified configuration in Figure 1(b) has the same least heat duty of 1.82 MW as that of Figure 1(a) but with greater structural attractiveness. We remark that the configuration performs a five-component separation in a single shell. It is a triple-effect column, with the topmost column being a DWC. The vapor-split at the bottom of the partition in this part can be indirectly controlled via pressures in the condensers of A and B. Thus, this type 2 DWC is more operable compared to the conventional type 1 DWC where the partitions lie in the middle of the shell (Chen and Agrawal, 2020).



(a) Heat-integrated configuration with least heat duty

(b) After process intensification

Figure 1: Configurations identified by formulation for case study. Column number and pressure are indicated in roman and arabic numerals respectively.

4. Conclusions

Shortcut models are invaluable in screening the vast search space of possible configurations for energy demand, since a systematic analysis and comparison using standard process simulators is impractical. However, such models have been limited in their accuracy and the flowsheet features that they could capture. In this work, we presented solutions to three challenges faced by current shortcut models for distillation during global optimization: 1) accounting for the variations in component relative volatilities, 2) accounting for different latent heats of vaporization, and 3) searching for heat integration opportunities. We incorporated these advances in our optimization formulation, and it is now capable of quickly identifying energy-efficient and heat-integrated distillation configurations with greater accuracy, as demonstrated on a five-component separation of aromatics. Furthermore, we demonstrated how various PI techniques can improve the structural attractiveness of selected configurations without comprising on their energy benefits through thermodynamic equivalence. The PI techniques we used, and several others, are general and can be applied to a wide number of configurations in addition to the one we showed. This work will aid the process engineer in more reliably finding valuable and novel configurations for their separations.

Acknowledgments

We thank the US Department of Energy (Award number: DE -- EE0005768) for financial support. We also thank Dr. Radhakrishna Tumbalam Gooty for his input during discussions of the work in this paper.

References

R. Agrawal, 2000, Multieffect Distillation for Thermally Coupled Configurations. AIChE Journal, 46(11), pp.2211–2224.

R. Agrawal, 2001, Multicomponent Distillation Columns with Partitions and Multiple Reboilers and Condensers. Industrial Engineering & Chemistry Research, 40(20), pp.4258-4266.

N. J. Anderson, and M. F. Doherty, 1984, An Approximate Model for Binary Azeotropic Distillation Design, Chemical Engineering Science, 39(1), pp.11-19.

J. A. Caballero, and I. E. Grossmann, 2006, Structural considerations and modeling in the synthesis of heat-integrated thermally coupled distillation sequences, Industrial Engineering & Chemistry Research, 45(25), pp.8454-8474.

Z. Chen, and R. Agrawal, 2020, Classification and Comparison of Dividing Walls for Distillation Columns. Processes, 8(6), p.699.

Z. Jiang, and R. Agrawal, 2019, Process intensification in multicomponent distillation: A review of recent advancements, Chemical Engineering Research Design, 147, pp.122-145.

G. Madenoor Ramapriya, A. Selvarajah, L. E. Jimenez Cucaita, J. Huff, M. Tawarmalani, and R. Agrawal, 2018a, Short-Cut Methods versus Rigorous Methods for Performance-Evaluation of Distillation Configurations, Industrial & Engineering Chemistry Research, 57(22), pp.7726–7731.

G. Madenoor Ramapriya, M. Tawarmalani, and R. Agrawal, 2018b, A Systematic Method to Synthesize All Dividing Wall Columns n-Component Separation — Part I, AIChE Journal, 64(2), pp.649–659.

T. J. Mathew, R. Tumbalam Gooty, M. Tawarmalani, and R. Agrawal, 2020, Quickly Assess Distillation Columns, Chemical Engineering Progress, 116(12), pp.27-34.

T. J. Mathew, R. Tumbalam Gooty, M. Tawarmalani, and R. Agrawal, 2021, A simple criterion for feasibility of heat integration between distillation streams based on relative volatilities, Industrial & Engineering Chemistry Research, 60(28), 10286-10302.

P. Mole, 1958, A unified theory for ideal separation processes. Chemical Engineering Science, 8(3), pp.236-243.

U. Nallasivam, V. H. Shah, A. A. Shenvi, J. Huff, M. Tawarmalani, and R. Agrawal, 2016, Global optimization of multicomponent distillation configurations: 2. Enumeration based global minimization algorithm, AIChE Journal, 62(6), pp.2071-2086.

J. Ryu and C. T. Maravelias, 2020, Computationally efficient optimization models for preliminary distillation column design and separation energy targeting, Computers and Chemical Engineering, 143, 107072.

V. H. Shah and R. Agrawal, 2010, A matrix method for multicomponent distillation sequences, AIChE journal, 56(7), pp.1759-1775.

R. Tumbalam Gooty, M. Tawarmalani, and R. Agrawal, 2018, Optimal Multicomponent Distillation Column Sequencing: Software and Case Studies, Computer Aided Chemical Engineering, 44(2), pp.660–672.

A. J. V. Underwood, 1948, Fractional distillation of multi-component mixtures, Chemical Engineering Progress, 44, pp.223–228.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50107-X

Synthesis of Advanced Reactive Distillation Technologies: Early-Stage Assessment Based on Thermodynamic Properties and Kinetics

Isabel Pazmiño-Mayorga[a*], A.A. Kiss,[a] M. Jobson[a]

[a] *Centre for Process Integration, Department of Chemical Engineering, The University of Manchester, M13 9PL, Manchester, United Kingdom*

isabel.pazminomayorga@manchester.ac.uk

Abstract

In the early stages of process design, the design space is enormous, often requiring simplified models that are based on conventional unit operations with their inherent strengths and limitations. In this paper, we present a novel conceptual framework and a synthesis methodology applied to a specific design problem featuring an equilibrium-limited reaction or a separation that can benefit from the introduction of a reactive separating agent to use intensified technologies following the success of reactive distillation. Two case studies (related to lactic acid and dimethyl ether production) illustrate the application of the synthesis methodology, and the outcomes are verified with conceptual design studies reported in the literature.

Keywords: Process synthesis; Reactive distillation; Process intensification; Feasibility methodology; Fluid separations

1. Introduction

Chemical industries face multiple challenges due to the increasing competitiveness, tighter regulations, and more stringent technical, commercial, safety and environmental requirements. Conventional process synthesis and design approaches to serve chemical industries generally apply established unit operations that are continuously challenged to reach new targets. Also, the increase in computing power has benefited the development of complex superstructure optimization methods combining synthesis and design simultaneously and including economic and sustainability indicators. As a result, the increased complexity requires numerous assumptions to simplify the models and often neglect important characteristics of the system that may impact process design at later stages. On the other hand, understanding a chemical system is crucial in the early stages of process design, for example, during flowsheet development.

Therefore, this work focuses on a much smaller subset of the design problem and proposes a methodology for early-stage process design to quickly evaluate advanced reactive distillation technologies that incorporate additional intensification features to reactive distillation (RD). These advanced RD technologies can be applied to equilibrium-limited reactive systems in the liquid phase or fluid separations that can benefit from the introduction of a reactive separating agent (RSA) to facilitate the separation. Decision making for process synthesis is guided by high-level questions about the chemical system of interest, which can be answered by knowing basic thermodynamic properties and kinetic data and using experience or experts' knowledge from conventional RD applications.

2. Advanced reactive distillation technologies

The technologies included in the scope of the synthesis methodology were developed following the success of RD by evaluating additional intensification features to expand the operating window and find potential new applications. The technologies included in this study are reactive dividing-wall column (R-DWC), catalytic cyclic distillation (CCD), reactive internally heat-integrated distillation (R-HIDiC), reactive high-gravity distillation (R-HiGee), and membrane-assisted reactive distillation (MA-RD). Figure 1 illustrates the key features of the five technologies included in the scope of this study and highlights their advantages compared to RD.

3. Research approach, conceptual framework and scope

3.1. Research approach

Kiss et al. (2019) noted that advanced RD technologies have not reached maturity due to the lack of robust methods and tools for process design and simulation, process dynamics and control, lack of pilot and industrial tests, and because of practical challenges related to ease of implementation. However, the range of applications of conventional reactive distillation is well established for equilibrium-limited reaction systems that have been studied on a laboratory, pilot or industrial scale (Luyben, 2013; Skiborowski, 2018). The ongoing research and developments about advanced distillation technologies (Kiss, 2013) and the knowledge and understanding of various chemical systems – feed and product specification, kinetics, phase equilibria, catalysts and operating conditions – are the basis for the decision-making methodology that aims to include advanced RD technologies in the early stages of process design.

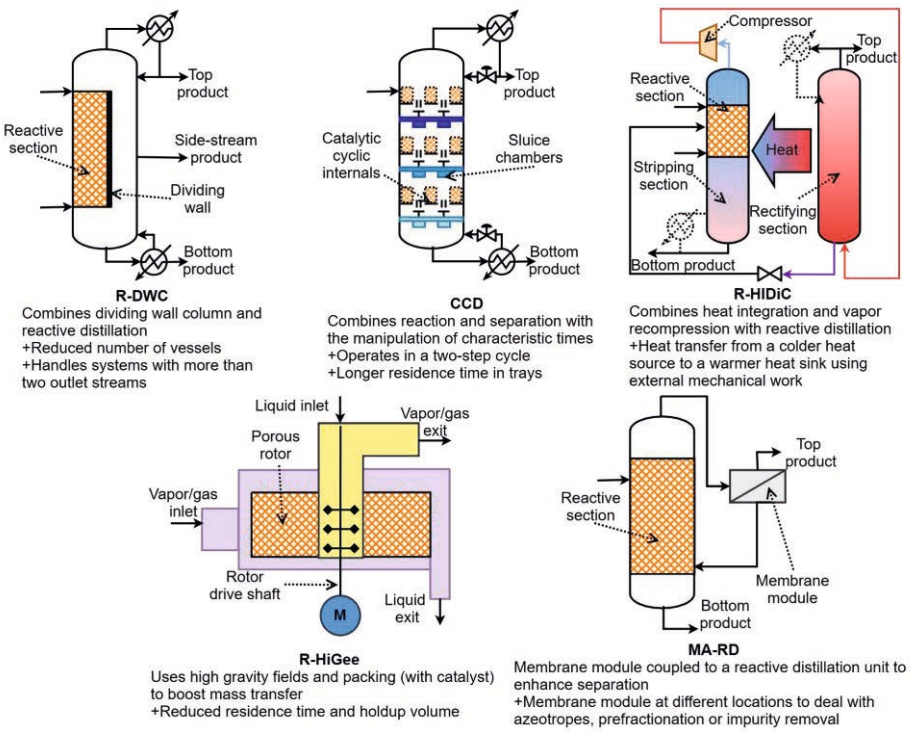


Figure 1. Main features of the five advanced reactive distillation technologies

3.2. Conceptual framework and scope

The methodology aims to be general enough to cover a wide range of equilibrium-limited reaction systems and their characteristics relevant to large-scale industrial applications (e.g., azeotropes, impurities and trace components, difficult separations). These characteristics are often simplified in laboratory-scale investigations, such as considering pure feeds, a large amount of solvent, and not accounting for the type of utilities needed. As a result, these characteristics usually lead to the need for multiple processing steps. However, intensified technologies that combine reaction and separation may overcome these challenging characteristics resulting in more compact and efficient processing units.

Figure 2 shows the conceptual framework of the decision-making methodology proposed in this work, which prompts high-level questions to the user to know the chemical system better. These questions can be answered given the thermodynamic properties and the kinetic parameters. The most common sources of information are also indicated.

Thermodynamic property data of pure compounds and mixtures include process stream compositions (feeds and products), basic physical properties and phase equilibrium information. These data can be easily gathered from databases and handbooks, from experiments or robust equations of state and activity coefficient models.

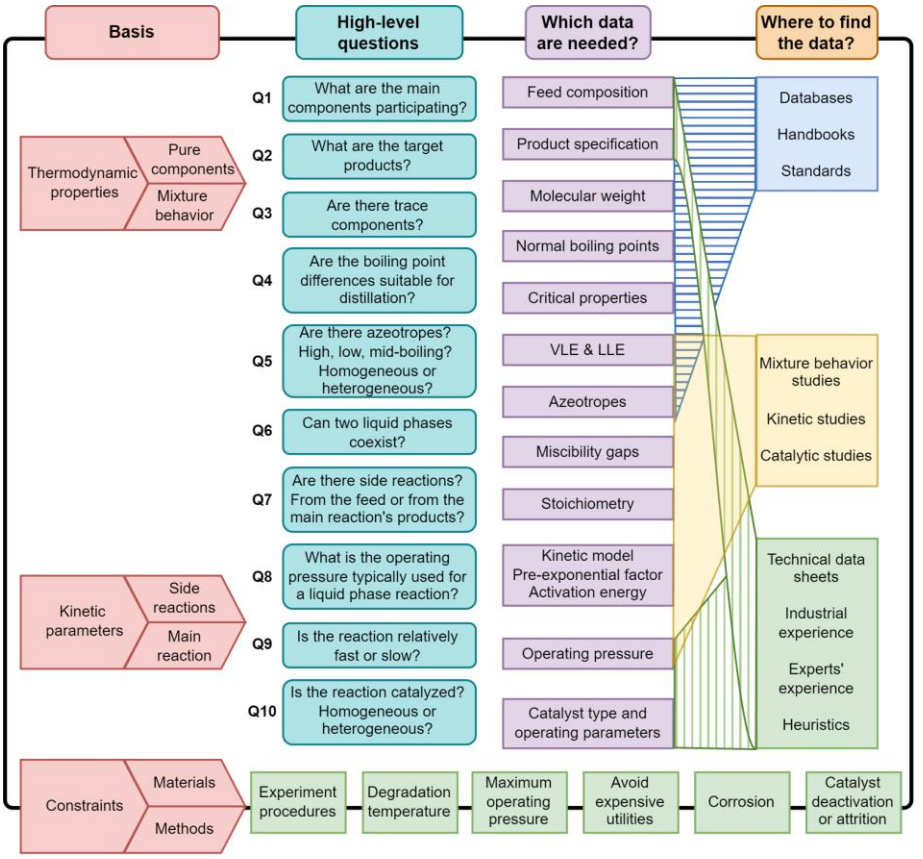


Figure 2. Conceptual framework for the synthesis of advanced reactive distillation technologies

The kinetic parameters of the main reaction must be carefully identified along with relevant side reactions that are often neglected in the conceptual design phase. For example, a parallel reaction can compete with the main reaction (consuming a reactant), or a series reaction can consume a product. Therefore, potential side reactions and their impact must be thoroughly understood at early stages because neglecting them can drive the decision-making process towards infeasible designs that may not be able to handle or overcome such complexities.

Information about thermodynamic properties and kinetics parameters can guide selection, but decisions also need to take into account constraints imposed by materials (i.e., catalysts, materials of construction) and methods (e.g., laboratory experiments). The materials constraint accounts for the chemicals used, the catalyst (functions, robustness against deactivation, thermal and mechanical performance, availability for large scale applications), and the materials used for vessels, column internals, membranes, and fluid handling system. The methods constraint includes laboratory procedures to evaluate kinetics or mixture behavior, or methods to conduct equipment sizing especially when severe conditions are required: operating pressure and temperature, moving parts, corrosive environment, or abrasive materials.

4. Decision-making methodology applied to two case studies

The high-level questions shown in Figure 2 are the backbone of the decision-making methodology. These questions firstly address the feed composition and the expected products and by-products, considering dilution, trace components and impurities. The next set of questions refers to the operating conditions: operating pressure and the range of temperatures expected. The next step deals with the reactions identified in the system. The forward reaction rate constant of the main reaction is evaluated at 363.15 K to categorize its rate as relatively slow, average, or fast. Side reactions are then evaluated according to the source of the reactants: from the feed stream, which can drive parallel reactions; or from the main reaction products, which can trigger a series reaction. Finally, the mixture behavior is evaluated by understanding the phases present and the existence of azeotropes.

We illustrate the application of the decision-making methodology in two case studies. A brief introduction about each chemical system allows the high-level questions to be answered. In each case study, relevant decision points that guide the technology selection are examined, along with an excerpt of the decision-making flowchart.

5. Results and discussion

5.1. Concentration and purification of lactic acid

The feed consists of a pre-treated fermentation broth that contains lactic acid (LA) (30 % wt.), succinic acid (5 % wt., reactive impurity), and a large amount of water. The design problem involves separating water and other acid impurities to obtain food-grade LA (88 % wt.). We introduced a reactive separating agent (RSA), methanol, to promote the esterification LA to produce methyl lactate and water with a heterogeneously catalyzed equilibrium-limited reaction of the type $A + B \rightleftharpoons C + D$. Succinic acid also reacts to produce succinates. Once the heavier succinates are removed, the reverse hydrolysis reaction is promoted by providing sufficient water in the liquid phase to obtain LA on specification and free from the impurity.

Relevant answers used in the decision-making flowchart are that the feed is diluted (Q1); the streams identified are product, by-product, excess water, and the RSA (Q2); the operating pressure should be sub-atmospheric to avoid thermal degradation (Q8); the existence of parallel side reaction from the impurity in the feed (Q7); the existence of homogenous mid boiling azeotropes (Q5), which are presented in an excerpt of the decision-making flowchart in Figure 3a.

The main outcomes are that a preconcentration step is required to remove excess water. Among the advanced RD technologies evaluated, the R-DWC is deemed suitable due to the number of outlet streams required, the side reaction by-products and the presence of mid-boiling azeotropes that can be consumed in the reactive sections. In addition, R-HIDiC and CCD are discarded because of their limitations to operate under vacuum. These outcomes were supported with conceptual design studies about thermally coupled configurations (Kim et al., 2017) and a dual R-DWC (Pazmiño-Mayorga et al., 2021), which demonstrated energy savings compared to a conventional RD.

5.2. Production of dimethyl ether (DME)

A feed of pure methanol reacts to produce water and dimethyl ether (99.5% mol). Conventional DME production requires high temperatures and pressures and several unit operations. The design problem involves a heterogeneously catalyzed liquid-phase reaction of the type A ⇌ B + C.

The main characteristics driving decisions for technology selection are illustrated in Figure 3b. The reaction rate constant is relatively slow in the framework of RD applications (Q9), the presence of a small two-phase liquid region (Q6) and the absence of azeotropes (Q5).

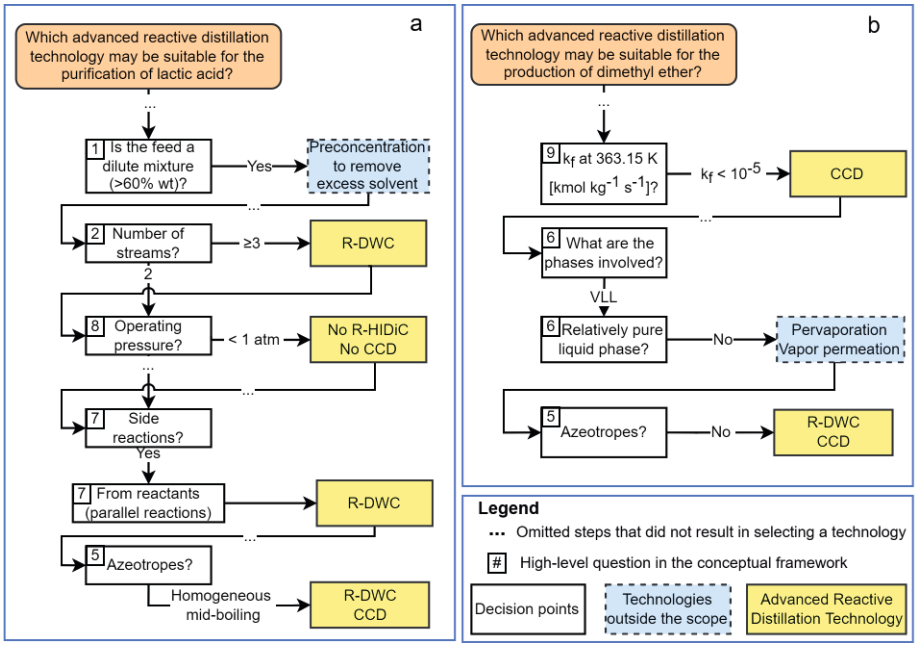


Figure 3. Excerpt of the decision-making methodology and outcomes a) lactic acid purification b) dimethyl ether production

The technologies deemed suitable for DME production are CCD as larger holdups allow longer residence times suitable for the relatively slow reaction. The absence of azeotropes enables the application of R-DWC and CCD without risking the purity of the outlet streams. However, the existence of a small two-phase region suggests that auxiliary equipment may be needed to remove water, such as a membrane. These outcomes were demonstrated with the conceptual design study of an R-DWC that showed energy saving compared with conventional RD (Kiss and Suszwalak, 2012). Also, DME production in a CCD has been studied at the conceptual level (including a patent), indicating that higher purities can be achieved with fewer stages and a lower vapor flow rate (Pătruţ et al., 2014; Kiss et al., 2015).

6. Conclusions

This work showcases a decision-making methodology for the synthesis of advanced RD technologies using chemical system data that can be easily obtained at early stages during flowsheet development. The highlights are the inclusion of novel intensified technologies that are often dismissed in traditional process design settings, the use of thermodynamic properties and kinetic parameters of the chemical systems and a set of heuristics from recent research about RD and advanced distillation. The methodology is demonstrated in two case studies. The results of published studies verify the outcomes of the methodology. A more detailed description of the methodology is under preparation with further case studies to evidence its range of applicability.

Acknowledgements

IPM gratefully acknowledges the full fund support from SENESCYT-Ecuador.

References

Kim, S.Y., Kim, D.M., Lee, B., 2017. Process simulation for the recovery of lactic acid using thermally coupled distillation columns to mitigate the remixing effect. Korean J. Chem. Eng. 34, 1310–1318.

Kiss, A.A., 2013. Advanced distillation technologies: design, control, and applications. Wiley, Chichester, West Sussex, United Kingdom.

Kiss, A.A., Bîldea, C.S., Pătruţ, C., 2015. Process and installation for the production of dialkyl ether. CA2936291A1.

Kiss, A.A., Jobson, M., Gao, X., 2019. Reactive Distillation: Stepping Up to the Next Level of Process Intensification. Ind. Eng. Chem. Res. 58, 5909–5918.

Kiss, A.A., Suszwalak, D.J.-P.C., 2012. Innovative dimethyl ether synthesis in a reactive dividing-wall column. Comput. Chem. Eng. 38, 74–81.

Luyben, W.L., 2013. Distillation design and control using Aspen simulation, 2nd ed. ed. Wiley, Hoboken, N.J.

Pătruţ, C., Bîldea, C.S., Kiss, A.A., 2014. Catalytic cyclic distillation – A novel process intensification approach in reactive separations. Chem. Eng. Process. 81, 1–12.

Pazmiño-Mayorga, I., Jobson, M., Kiss, A.A., 2021. Conceptual design of a dual reactive dividing wall column for downstream processing of lactic acid. Chem. Eng. Process. 164, 108402.

Skiborowski, M., 2018. Process synthesis and design methods for process intensification. Curr. Opin. Chem. Eng., Biotechnology and bioprocess engineering 22, 216–225.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50108-1

Process Synthesis and Intensification for Upgrading Natural Gas Liquids in Shale Gas

Zewei Chen[a], Edwin Andres Rodriguez Gil[a], Rakesh Agrawal[a*]

[a]*Davidson School of Chemical Engineering, Purude University, West Lafayette, Indiana, USA*
agrawalr@purdue.edu

Abstract

In the past two decades, natural gas flaring and venting have increased due to the lack of transforming or transportation infrastructure in emerging shale-gas-producing regions. To reduce carbon emissions and wastage of shale resources, we recently reported several innovative flowsheets for natural gas liquid (NGL) to liquid fuel processes on a small scale near the wellhead (Chen et al., 2021). These processes consist of a direct shale gas dehydrogenation reactor followed by an oligomerization reactor. In this work we perform thermodynamic analysis on both dehydrogenation and oligomerization reactors and demonstrate the benefits of our proposed processes over other processes alternatives. This study also provides a systematic procedure for the synthesis of economically attractive processes for small scale shale gas valorization.

Keywords: shale gas, natural gas liquid, process intensification

1. Introduction

Shale gas is a promising energy resource and chemical feedstock for the transition period towards a sustainable economy and has the potential to be a carbon source for the long term. However, huge amount of shale gas at remote shale gas basins is directly flared (Fisher et al. 2019) due to the lack of infrastructure to transport the gas from well heads to the central processing plant. To avoid long distance transportation, it is essential to convert shale gas to value-added and easily transportable products on site at a distributed scale. Liquid fuel with high market value and large market demand, such as gasoline and diesel are our target products. Unlike current shale gas process where large scales are preferred, simple and intensified processes with least processing steps and least pieces of equipment are favoured for remote shale plays.

Natural gas liquids contained in shale gas are especially of our interest since they are comparably easier to activate and transform to liquid products. While conventional shale gas processing usually follows a hierarchy of "Front-end Separation, NGL activation, NGL upgrading" (He and You 2014), we recently showed that a new process hierarchy of "NGL activation, NGL upgrading, Back-end separation" has significant advantages for small scale installation (Chen et al., 2021a). In this work, we carry out a systematic analysis of the proposed process hierarchy to illustrate its benefits and provide an evolution procedure from the conventional process to the novel processes.

Our process designs evolve from NGL co-processing (Ridha et al., 2018) wherein the NGL mixture after recovery is directly activated and upgraded to liquid fuel, to switched NGL activation and recovery, wherein the NGL recovery step takes place after the NGL activation step. Finally at the backend NGL recovery step, where the NGL recovery step takes place after the NGL has been upgraded to liquid fuel. A two-step conversion of NGL to liquid hydrocarbons via dehydrogenation followed by

oligomerization is used as an example to show how these innovative process designs evolve. We synthesize process configurations corresponding to each step in the evolution and illustrate the merits and shortcomings of each configuration through thermodynamic analysis on both dehydrogenation and oligomerization reactors. Higher yield of liquid products, fewer processing steps, reduced numbers of equipment pieces and elimination of energy and capital-intensive units can be achieved. The intensification of these processes would benefit the modularization of shale gas plants and make it possible for onsite distributed production of liquid hydrocarbons for remote shale locations.

2. Base Case Flowsheet

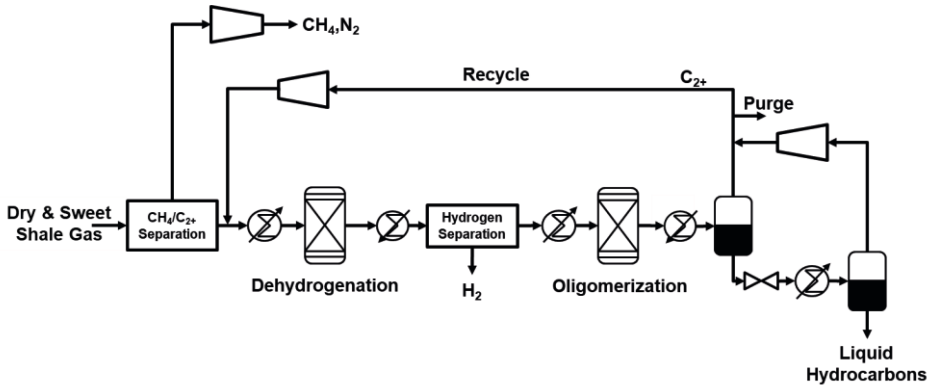


Figure 1: Process I: NGL co-processing proposed by Ridha et al. (2018)

We start with Process I (NGL co-processing) synthesized by Ridha et al., as depicted in Figure 1. In the configuration, the dry and sweet shale gas first goes through a CH_4/C_{2+} separation to separate CH_4 from natural gas liquids. The CH_4/C_{2+} separation unit is a conventional cryogenic demethanizer (Getu et al., 2013). NGL in this stream is then preheated and sent to the catalytic dehydrogenation unit wherein a portion of NGL is converted to their corresponding olefins. The effluent stream is sent to a hydrogen separation unit for hydrogen removal and then the oligomerization unit wherein olefins are converted into longer chain molecules which are liquid in the ambient state. The separation between C_{5+} liquid and light hydrocarbons (C_2-C_5) in stream is performed in a simple two-flash system. In process I (NGL co-processing), NGL and its corresponding olefins are directly sent to the oligomerization reactor without paraffin-olefin separation, and a much simpler separation between C_{5+} liquid and light hydrocarbons is employed. This design is already much simpler than the conventional shale gas process (He and You, 2014).

However, the configuration in Figure 1 is still not economically attractive for a 10 million standard cubic feet per day (MMSCFD) small scale plant at the gas gathering station, due to the following reasons: 1) The front-end separation consists of a series of energy intensive and costly unit operations, especially the cryogenic demethanizer, 2) The catalytic dehydrogenation reactor has limited conversion and selectivity, and 3) There are significant C_{2+} losses at multiple locations of the flowsheet, including the hydrogen membrane, CH_4/C_{2+} separation, and purge stream. In section 3, we will perform a systematic analysis on the dehydrogenation and oligomerization to show how we overcome these drawbacks step by step.

3. Systematic Analysis and Evolution of Process Configurations

We perform a thermodynamic analysis on the dehydrogenation reactor in Figure 1. There are two state-of-the-art methods for dehydrogenation: thermal dehydrogenation, and catalytic dehydrogenation. Catalytic dehydrogenation is operated at a relatively low temperature, usually below 700 ^{0}C and has a high selectivity towards olefins. For example, propane catalytic dehydrogenation is a widely used process for propylene production. Thermal dehydrogenation, on the other hand, is operated at a higher temperature and has a low selectivity due to cracking. A good example of this is steam cracking for olefin production. Although there are multiple efforts in literature studying catalytic ethane dehydrogenation (Dai et al., 2021), this technology is not suitable for a simple plug flow reactor due to equilibrium limitation. Figure 2 shows the equilibrium conversion for a pure ethane feed at different operating conditions. Equilibrium conversion curves show that the maximum conversion of ethane is only around 40% at 700 ^{0}C and 1 bar. The equilibrium conversion marks the highest conversion possible, while in a real catalytic dehydrogenation reactor, the conversion tends to be lower than this value. To achieve a higher conversion, one can either operate the reactor at a lower partial pressure or at a higher temperature. However, when operated at a higher temperature, thermal dehydrogenation is already fast enough and there is no need for the catalyst. The conventional ethane steam cracker only has a residence time of 0.2s and practitioners must quench the effluent stream to stop side reactions (Karimzadeh et al., 2009). Further decreasing the total pressure of the reactor results in vacuum operation, which is also not a good idea for hydrocarbon reactions. While ethane catalytic dehydrogenation may be incorporated in an advanced reactor design, such as membrane reactor to overcome the equilibrium conversion limit (Champagnie et al., 1992), complex reactor designs are not suitable options for small scale installation due to limited capital expenditure.

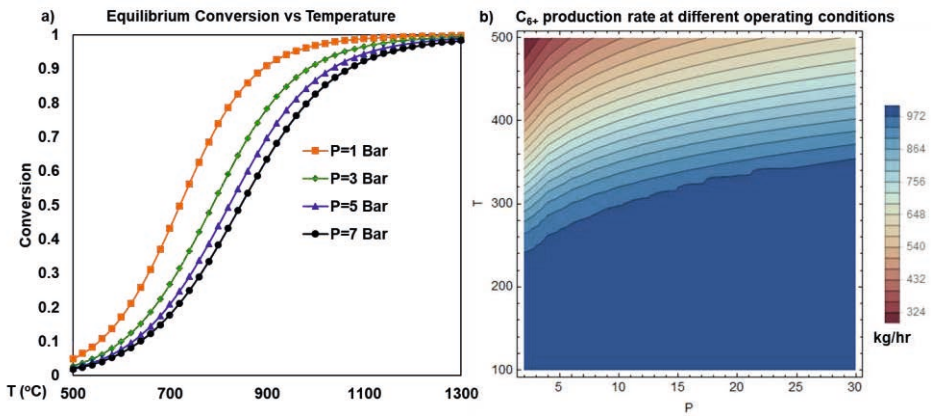


Figure 2: Sensitivity analysis over the dehydrogenation and oligomerization reactors. a) Ethane conversion as a function of operating temperature at different operating pressures, and b) C_{6+} production rate (kg/h) at different temperatures and pressures in the oligomerization reactor.

To achieve high conversion, thermal dehydrogenation is the only option to keep a simple reactor design. The conventional steam cracker is usually operated at 850 ^{0}C with 1:3 mass ratio of water to hydrocarbon in the cracker. The partial pressure of hydrocarbons in the steam cracker is around 1 bar, which indicates around 85%

conversion in Figure 2. In the steam cracking system, steam serves as an inert to decrease the partial pressure of hydrocarbons. However, steam cracker contains a complex system for steam generation, preheating, and post-reaction dehydration, which is also not suitable for small scale installation.

To overcome the thermodynamic as well as economic challenge, it is essential to identify an alternative inert in the system which could decrease the partial pressure of hydrocarbons without introducing additional complexity to the process design. CH_4, which is already mixed with the NGL in the shale gas from a well and a byproduct in the ethane thermal dehydrogenation becomes the perfect choice. Rather than separating CH_4 before dehydrogenation, leaving CH_4 in the feed to the reactor significantly simplifies the entire process. Process II (Switched NGL recovery and activation), the switched NGL recovery and activation, is a process containing such a process sequence (Figure 3). In this configuration, the dry and sweet shale gas is directly sent to thermal dehydrogenation unit and the CH_4/NGL separation takes place after the thermal dehydrogenation step. For a Bakken shale gas containing 57.8% CH_4, 20.0% C_2H_6, 11.4% C_3H_8, and 5.1% heavier components, when the total pressure of the reactor is 2 bar, the partial pressure of NGL components is only 0.6 bar, leading to a higher conversion and selectivity towards olefins.

Furthermore, the conventional steam cracker is followed by a demethanizer to remove CH_4, which is produced in the steam cracker as a byproduct. This demethanizer is a duplication of the front-end demethanizer, resulting in increased capital cost. In Process II (Switched NGL recovery and activation), the repeated demethanizer is now eliminated (Figure 3). The C_{2+} loss in this configuration is also significantly reduced because this process does not have a H_2/NGL separation and a purge stream as in Process I (NGL co-processing), which could be the sources of C_{2+} loss.

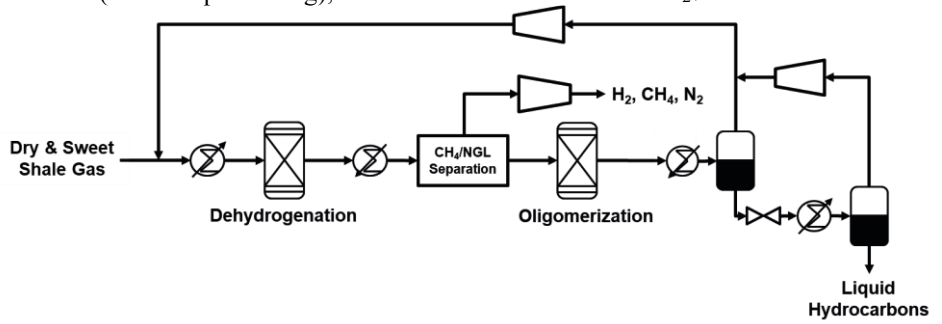


Figure 3: Process II: process configuration with switched NGL recovery and activation

We can further switch the CH_4/NGL separation unit and oligomerization unit. CH_4 and H_2 in the oligomerization reactor serve as thermal mass to somewhat mitigate the temperature increase and as a result, less reactor beds are needed for the process (Chen et al., 2021a). However, it results in reduced partial pressure of olefins, leading to a decreased conversion. Again, we performed a sensitivity analysis on the oligomerization reactor at different temperatures and pressures. The analysis is performed for a 1000 kg/h pure ethylene feed, using RGibbs model in Aspen Plus and assuming olefins from C_2 to C_{20} are produced. From the analysis, we observe that the C_{6+} yield is a weak function of pressure. At an operating temperature of 200 ^{0}C ~ 300 ^{0}C, almost all ethylene is converted to C_{6+}, even if operated at a low pressure. This analysis confirmed

that it is beneficial to further switch the order between the CH_4/NGL separation unit and oligomerization reactor.

Figure 4 shows Process III (Backend separation) wherein the oligomerization takes place before any separations, and it is denoted as back-end NGL recovery configuration. In this configuration, objective of the separation task is to separate the effluent stream from the oligomerization reactor into three streams, a gaseous stream containing H_2, CH_4, N_2 which could be further treated in a membrane separation unit to obtain pipeline natural gas, a recycle stream sent back to the dehydrogenation reactor, and a liquid hydrocarbon product stream containing C_{6+}. An absorption column using a portion of the liquid product stream as the absorbent is used for the separation of CH_4 and NGL. Readers may refer to Chen et al.'s work for detail description of this process. This backend separation system is much simpler and less costly than the conventional front-end demethanizer. The reason being that, as all the separations are performed together in the end, the separation task could become easier, and synergies could be identified among all separations. In Process III (Backend separation), if we denote the feed stream as ABC, in which A is H_2, CH_4, N_2, B is C_2-C_5 hydrocarbons, C is C_{6+} liquid hydrocarbons, then the separation in Process III (Backend separation) could be perceived as sloppy separation A/ABC and ABC/C. While in Process I (NGL co-processing) and II (Switched NGL recovery and activation), all the separations are sharp separations. The sloppy separations could be less energy intensive than the sharp separations, with the only penalty of slightly increasing the recycle ratio.

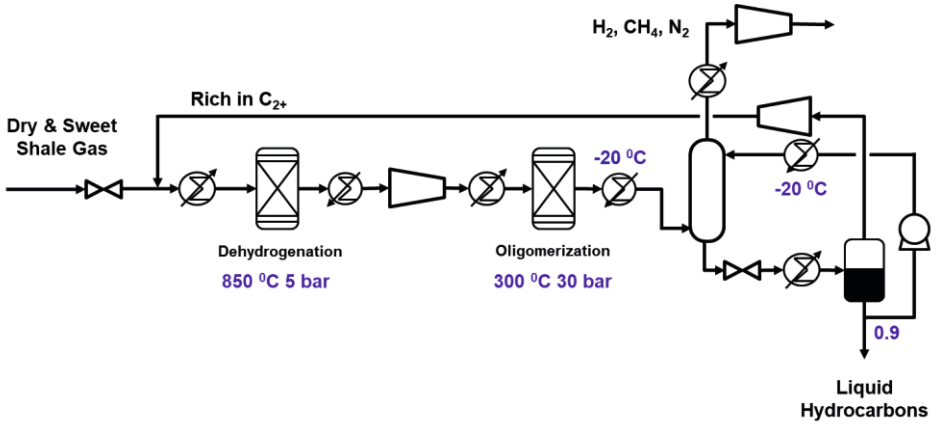


Figure 4: Process III: our process configuration with backend separation (Chen et al., 2021a)

From the simulations and thermodynamic analysis results above, we can conclude several general principles for process synthesis and intensification: when a feed mixture contains inert components and has constituents that are also created in downstream processing, which must be separated downstream, then one should carefully evaluate the merit of (1) avoiding upstream separation of the constituents from the feed mixture, and (2) arranging the processing sequence so the duplication of separation between any two components is avoided, (3) arranging the process sequence for the maximum thermodynamic benefits and (4) arranging most of the separations next to each other and identifying synergies among them. Furthermore, potential advantages of the inert in the feed for the downstream endothermic reactions, which are unfavoured according to

the Le Chatelier's principle, should be carefully explored against the increase in the cost due to an increase in the equipment size. Using these principles, we recently proposed an efficient and cost-effective process for propylene and ethylene production from shale gas (Chen and Agrawal, 2021b).

4. Conclusions

In this work, we described a series of processes for shale gas valorization at remote shale gas basins evolving from NGL co-processing (Ridha et al., 2018) wherein the NGL mixture is directly activated and upgraded to liquid fuel. However, we switched NGL recovery and activation steps and used the CH_4 present in the shale gas as inert for the NGL dehydrogenation step. Finally, unconverted NGLs and methane are separated in one simple separation step after conversion of olefins to liquid fuel. We performed sensitivity analysis on both dehydrogenation and oligomerization reactors and revealed the effects of operating pressure and temperature over the performance of the reactor. In terms of this, we made decisions whether to switch the order between separations and reactions and demonstrated that the backend NGL recovery process has advantages over the other two previous methods and hence more suitable for small scale installation. We also summarized several general principles for process intensification which could be potentially applied to other reaction-separation networks.

References

A. M. Champagnie, T. T. Tsotsis, R. G., Minet, and E. Wagner, 1992, The study of ethane dehydrogenation in a catalytic membrane reactor. Journal of Catalysis, 134(2), 713-730.

Z. Chen, Y. Li, W. P. Oladipupo, A. R. Edwin Gil, G. Sawyer, R. Agrawal, 2021a, Alternative ordering of process hierarchy for more efficient and cost-effective valorization of shale resources. Cell Reports Physical Science, 100581.

Z. Chen, R. Agrawal, 2021b, Alternative Processing Sequence for Process Simplification, Cost Reduction, and Enhanced Light Olefin Recovery from Shale Gas. ACS Sustainable Chemistry & Engineering, 9, 41, 13893–13901

Y. Dai, X. Gao, Q. Wang, X. Wan, C. Zhou, Y. Yang, 2021, Recent progress in heterogeneous metal and metal oxide catalysts for direct dehydrogenation of ethane and propane. Chemical Society Reviews, 50, 5590-5630.

D. Fisher, M. J. Wooster, 2019, Multi-decade global gas flaring change inventoried using the ATSR-1, ATSR-2, AATSR and SLSTR data records. Remote Sensing of Environment, 232, 111298.

M. Getu, S. Mahadzir, N. V. D. Long, M. Lee, 2013, Techno-economic analysis of potential natural gas liquid (NGL) recovery processes under variations of feed compositions. Chemical Engineering Research and Design, 91(7), 1272-1283.

C. He, F. You, 2014, Shale gas processing integrated with ethylene production: novel process designs, exergy analysis, and techno-economic analysis. Industrial & Engineering Chemistry Research, 53(28), 11442-11459.

R. Karimzadeh, H. R., Godini, M. Ghashghaee, 2009, Flowsheeting of steam cracking furnaces. Chemical Engineering Research and Design, 87(1), 36-46.

T. Ridha, Y. Li, E. Gençer, J. J. Siirola, J. T. Miller, F. H., Ribeiro, R. Agrawal, 2018, Valorization of shale gas condensate to liquid hydrocarbons through catalytic dehydrogenation and oligomerization. Processes, 6(9), 139

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50109-3

Energy-Efficient Direct Cyclohexene to Cyclohexanol Process by Heat Pump Assisted Reactive Distillation

Xinyan Liu[a], Yang Lei[a], Hao Luo [a*], Xiaoqin Wu[a] and Rafiqul Gani[b]
[a] *Department of Chemical & Chemical Engineering, Wuhan University of Science and Technology, Wuhan 430081, China*
[b]*PSE for SPEED, Skyttemosen 6, DK-3450 Allerod, Denmark*
haoluo@wust.edu.cn

Abstract

Cyclohexanol is regarded as an important raw material to produce many useful chemicals such as hexanedioic acid, hexamethylenediamine, and caprolactam. However, it's production usually leads to high energy consumption. Traditional technologies include three routes: oxidation of cyclohexane, hydrogenation of phenol and direct hydration of cyclohexene. Avoiding the risk of explosion and a low price of raw material, the direct hydration of cyclohexene route is selected in this paper. To enhance the energy efficiency, a novel heat-pump-assisted reactive distillation process (HPRD) is proposed and simulated through AspenPlus. The discharge compressor pressure is optimized to be 3.2 bar to minimize the cost. For purposes of comparison, two conventional processes, aqueous-phase-refluxed reactive distillation with a stripper (ARDS) and organic-phase-refluxed reactive distillation with a stripper (ORDS), are also simulated and optimized through sensitivity analysis. All these three processes are evaluated through energy and economic analysis. The results show that the ORDS process saves the total energy and annualized cost by 15% and 12%, respectively, compared with the ARDS process, while the heat-pump-assisted process realizes a significant energy saving of 65% and achieves 33% reduction in total annualized cost, demonstrating a high economic feasibility.

Keywords: Cyclohexene to cyclohexanol; Direct hydration; Reactive distillation; Heat-pump; Process optimization.

1. Introduction

Cyclohexanol, as an important intermediate chemical, has been widely used in the production of hexanedioic acid, hexamethylenediamine, caprolactam and many more. These chemicals are the main raw materials for producing polymers such as Nylon 6,6 and Nylon 6. Therefore, cyclohexanol is usually a raw material in huge demand in the chemical industry and receives much attention in terms of their synthesis approaches and catalysts research. One of the main traditional routes for cyclohexanol production is the direct hydration of cyclohexene, overcoming the risk of explosion and high cost, has been identified as the main research direction for cyclohexanol production.

The direct hydration method allows cyclohexene contacting with water by an additive reaction to produce cyclohexanol. However, the reaction occurs in the liquid phase, and the immiscibility of cyclohexene and water leads to a low conversion rate. What's more, the equilibrium reaction would also result in low production efficiency. In order to reduce energy consumption, reactive distillation, as one of the technologies for process

intensification, has been applied in previous studies. The production rate of cyclohexanol could achieve a significant increase through the continuous removal, breaking the limitation of equilibrium reaction and promoting the reaction to the forward direction. Chen et al. (2014) proposed a reactive distillation process making full use of liquid-liquid splitting of the binary cyclohexanol-water azeotrope. Ye et al. (2014) introduced a side reactor configuration. Both strategies could achieve significant energy-efficiency and cost-saving. Although direct cyclohexene hydration as an alternative has been suggested, overall process simulation and evaluation as well as detailed analysis of energy and economic aspects are still scarce. To improve the energy efficiency, some novel technologies such as heat pump, extractive distillation and different flowsheet configurations with different phase reflux options could also be considered.

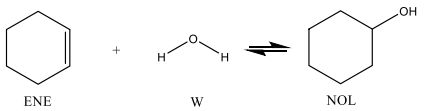
In this work, a novel heat-pump-assisted reactive distillation process (HPRD), combining two technologies: reactive distillation and vapor recompression (VRC) heat pump, is proposed for direct hydration cyclohexene to cyclohexanol, thus allowing a significant reduction of the energy requirements for cyclohexanol purification. The process is simulated with rigorous AspenPlus modules. For a comprehensive comparison, two conventional reactive distillation processes with different phase reflux options (aqueous and organic phase), ARDS (aqueous-phase-refluxed reactive distillation with a stripper) and ORDS (organic-phase-refluxed reactive distillation with a stripper) are also simulated. All these three processes are evaluated by energy and economic analysis.

2. Problem Statement

As a useful material for surfactant production and other industrial utilization, cyclohexanol (NOL) should have a minimum purity of 99 wt%, but this specification requires high energy consumption, leading to a large energy penalty. Using aqueous-phase-refluxed reactive distillation (RD) with a stripper, owing to the enhanced equilibrium conversion and full use of liquid-liquid separation, the energy requirements could be reduced by 46.76% from 2603 kJ/(kg NOL) to 1386 kJ/ (kg NOL) (Ye et al., 2014). However, the energy and economic costs needed for cyclohexanol production are still considerably high, especially when taking into account the price gap between cyclohexane and cyclohexanol. To address this issue, we integrate a heat-pump into the top of the RD-column because, conceptually, less energy would be needed in the reboiler through heat integration, compared with the conventional reactive distillation process. Also, the objective is to confirm previously published results with another thermodynamic model as well as the integrated heat-pump design by studying different integration schemes.

3. Reaction kinetics and phase equilibrium model

The parameters of kinetic models are obtained from Sun et al., (2021). The reaction of direct hydration of cyclohexene (ENE) with water to produce cyclohexanol (NOL) is given by Eq. (1) and the kinetic equation for this reaction is given by Eq. (2):



(1)

$$-r = -\frac{dC_W}{mdt} = 54775.19 \exp\left(\frac{-57894.54}{R_g T}\right) C_{ENE} C_W - 1.7327 \exp\left(\frac{-16451.74}{R_g T}\right) C_{NOL} \quad (2)$$

Where, r is reaction rate in mol/(L s gcat), m is mass of the catalyst in g, T is temperature in K, CENE is water concentration mol/L, CW is the concentration of water mol/L, CNOL is the concentration of NOL mol/L.Rgis the ideal gas constant, 8.314 J/(mol.K). It is assumed that the catalyst occupies half of the total hold of each reactive tray, and 90% of the tray area is considered as the active area to consider the space for downcomer. The column diameter is obtained from Column Internals in AspenPlus. The catalyst is Amberlyst 36 cation exchange resin and its density is assumed to be 770 kg/m^3.

The Universal Quasi–Chemical/Hayden-O'Connell equation of state with Henry's law (UNIQUAC-HOC) property model is selected as an appropriate model in ApsenPlus to describe the vapor-liquid-liquid equilibrium. Table 1 lists the UNIQUAC binary parameters. Table 2 compares the model predicted azeotropic composition with that reported in the literature and also with the NRTL model reported in our published work (Liu et al., 2022), the relative error is smaller than 3% and the two models demonstrate almost the same results. Therefore, it is reasonable to use UNIQUAC-HOC model for process simulation.

Table 1 the UNIQUAC binary parameters in the system

Component *i*	Component *j*	a_{ij}	a_{ji}	b_{ij}	b_{ji}
C_6H_{10}	H_2O	0	0	-1024.1	-466.35
$C_6H_{12}O$	H_2O	0	0	-342.857	26.0981

Table 2 the comparison of boiling point for different systems

Component	Literature values (Gould, 1973)		Calculated values using NRTL (Liu et al., 2022)		Calculated values using UNIQUAC	
	T/°C	Molar composition	T/°C	Molar composition	T/°C	Molar composition
ENE	82.75	-	82.88	-	82.88	-
W	100.00	-	100.02	-	100.02	-
(ENE-W)	70.80	(0.308,0.692)	70.62	(0.314,0.686)	70.58	(0.314,0.686)
NOL	160.65	-	160.84	-	160.84	-
(NOL-W)	97.80	(0.927,0.073)	98.24	(0.929,0.071)	98.46	(0.929,0.071)

Note: ENE denotes cyclohexene; W denotes water; NOL denotes cyclohexanol

4. Process Simulation

In this work, for all processes described below, including the ARDS, ORDS and HPRD. The process simulations were conducted using the rigorous modules (RADFRAC). in AspenPlus. The productivity of cyclohexanol is set to 10016 kg/hr (100 kmol/hr).

4.1. Aqueous-phase-refluxed reactive distillation with a stripper (ARDS)

The ARDS process contains two distillation columns, one is reactive distillation (RD), the other is a distillation stripper in which the product (NOL) is purified. The aqueous phase from the RD-column decanter is fully refluxed to the top of the RD-column, thereby, the organic phase from the decanter is 99.75% cyclohexene, which can be recycled to the RD-column with its make-up, the bottom outlet is the cyclohexanol-water

mixture, which is separated by a decanter after cooling. The aqueous phase from the decanter is recycled back with fresh water into the RD-column, the organic phase is fed into a stripper to obtain 99.9 wt% cyclohexanol. A small heat exchanger is used to heat the recycled water stream to reduce the reboiler duty (691 kW) of the RD-column. In this process, the temperature of the RD-reboiler is 121.45 ºC and that of the stripper is 161.74 ºC, thereby low pressure (LP, 6atm 159 ºC) steam and medium pressure (MP, 11 atm, 184 ºC) steam are required. The original detailed flowsheet is given in Liu et al. (2022).

4.2. Organic-phase-refluxed reactive distillation with a stripper (ORDS)

The flowsheet of ORDS also contains reactive distillation and a stripper, similar to that of ARDS. The main difference is that the organic phase from the decanter is fully recycled back to the top of the RD-column. Therefore, the bottom outlet is a mixture of cyclohexene, cyclohexanol, and H_2O with a lower reboiler temperature (105.59 ºC). Unlike ARDS, where the second decanter is used to split the cyclohexanol-water mixture, in ORDS it is used to split the cyclohexanol-water-cyclohexene mixture, therefore, the remaining organic phase is the mixture of cyclohexene-cyclohexanol, which is further separated by a stripper with an additional decanter. It should be noted that the temperature difference between the RD-column condenser and reboiler in ORDS is smaller than that of ARDS. Also, organic-phase-refluxed method results in a higher cyclohexene concentration in the RD-column, leading to a better energy utilization efficiency of the reactive distillation column. The original detailed flowsheet is given in Liu et al. (2022).

4.3. Heat-pump-assisted reactive distillation (HPRD)

Figure 1 presents the flowsheet of heat-pump-assisted reactive distillation with its mass and energy balance from AspenPlus simulation. The results are almost the same with that reported in our previous work (Liu et al., 2022) in which the NRTL model is used. Compared with the ORDS process, the main difference is that a heat pump is used to compress the vapor phase to 3.2 atm at the top of the reactive distillation column to increase the stream temperature so that it could be used to heat the reboiler at the bottom, thereby an extra heat exchanger is required. A flash tank is also used to provide the bottom vapor phase for the RD-column.

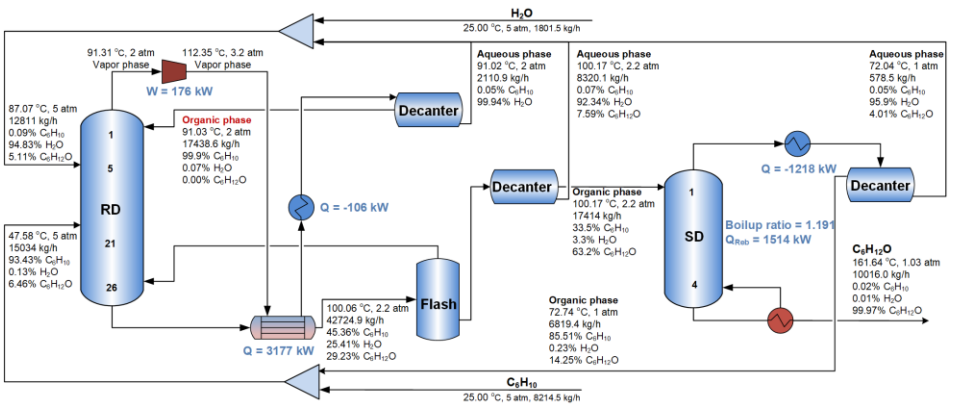


Figure 1 The flowsheet of the novel heat-pump assisted reactive distillation column

5. Process Evaluation

5.1. Evaluation method

Energy analysis: To assess the total energy consumption under the same criterion, the electricity is converted into the equivalent heat duty through a coefficient (see Eq. (3)).

$$TEC = W_e/\eta + Q_{th} \tag{3}$$

Where, We is the electricity consumption in pump and compressor, Q_{th} is the heat duty in flash and stripper, η is the conversion efficiency of thermal energy (heat duty) to electricity, which is usually around 0.3~0.4 (Luo et al., 2015).

Economic analysis: For the economic analysis, the total cost is often calculated from the sum of the annualized capital cost (ACC) and the total operating expenditure (OPEX). The ACC is computed by translating the total capital expenditure (CAPEX) into annualized ones. All the ACC, CAPEX and OPEX are calculated based on the procedure given by (Liu et al., 2021). The purchased equipment cost and utility cost in OPEX are computed from Aspen Process Economic Analyzer (APEA). Considering the same productivity of the different processes, the raw material (cyclohexene and fresh water) is not included in the comparison.

5.2. Evaluation results

For the energy consumption, the total energy consumption (TEC) of the three processes, including electricity (converted by equivalent heat) and heat duty are shown in Figure 2a. Compared with ARDS process, the ORDS process shows a reduction of 15% in TEC, demonstrating the advantage of the process with organic reflux, which could increase the concentration of ENE, enhancing the conversion rate. For the HPRD process, the TEC saving becomes 65%, which is a remarkable figure, implying a very high energy efficiency. Through economic evaluation, As shown in Figure 2b, it can be observed that the annualized capital cost (ACC) contributes little (< 3%) of the total cost compared with the operational expenditure (OPEX) containing the fixed operation cost (FOC) and variable operation cost (VOC) in three processes, indicating the predominance of OPEX in total annualized cost and the importance on energy saving. Compared with the ARDS, ORDS process, the HPRD process achieves a reduction of 23%, 33%, respectively, realizing significant economic-saving. The new analysis confirms the HPRD results reported earlier as the best among the different alternatives studied.

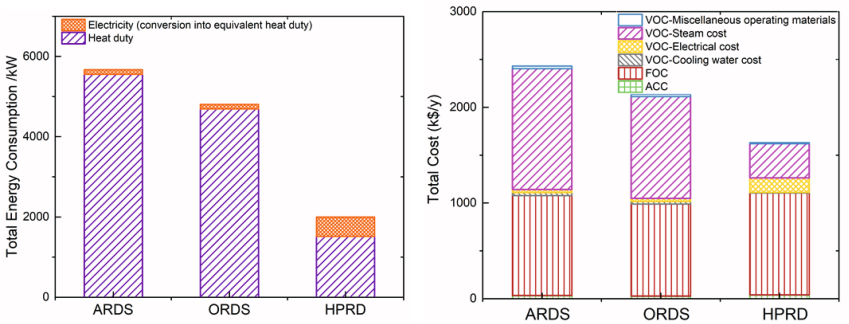


Figure 2 Total energy and economic results in three processes (a) total energy consumption (b) total cost

6. Conclusion

In this work, with cyclohexanol in a huge demand in current industry, the reactive distillation (RD) process for direct hydration of cyclohexene to cyclohexanol has been studied. The parameters of reaction kinetic model and phase equilibrium models are obtained from reported data in published papers and the database in AspenPlus. The prediction of phase equilibrium, which plays a very important role in this study, has been cross-checked not only with available data but also with different property models. The UNIQUAC model has been found to give similar results as the NRTL model. Two reactive distillation processes with conventional flowsheet are simulated and assessed, one is aqueous-phase-refluxed reactive distillation with a stripper (ARDS), the other is organic-phase-refluxed reactive distillation with a stripper (ORDS). To improve the energy efficiency, different heat-pump-assisted reactive distillation process (HPRD) alternatives have been studied and a novel scheme has been identified. The key operation parameters are optimized through sensitivity analysis. Different alternatives have been compared through rigorous process simulation and evaluated in terms of energy and economic data. Results show that the performance of ORDS is better due to a slightly higher conversion of cyclohexene to cyclohexanol. The HPRD process, however, realizes a significant energy saving of 58% and achieves a 23% reduction in total energy consumption and total annualized cost, as compared to ORDS, indicating the possibility of a good economic-saving and environmentally friendly technology.

Acknowledgements

This study was financially supported by the National Undergraduate Training Program for Innovation and Entrepreneurship (No. S202110488032) and the Scientific Research Foundation from Wuhan University of Science and Technology (No.1050038; No.1050040).

References

B. C. Chen, B. Y. Yu, Y. L. Lin, H. P. Huang, I.L. Chien, 2014. Reactive-Distillation Process for Direct Hydration of Cyclohexene to Produce Cyclohexanol. Industrial & Engineering Chemistry Research 53, 7079-7086.

R.F. Gould, 1973. Azeotropic Data—III, Copyright, Advances in Chemistry Series.

X. Liu, Y. Chen, S. Zeng, X. Zhang, X. Liang, R. Gani, G.M. Kontogeorgis, 2021. Separation of NH_3 /CO_2 from melamine tail gas with ionic liquid: Process evaluation and thermodynamic properties modelling. Separation and Purification Technology 274, 119007.

X. Liu, H. Luo, Y. Lei, X. Wu, R. Gani, 2022. Heat-pump-assisted reactive distillation for direct hydration of cyclohexene to cyclohexanol: a sustainable alternative. Separation and Purification Technology 280, 119808.

H. Luo, C.S. Bildea, A.A. Kiss, 2015. Novel Heat-Pump-Assisted Extractive Distillation for Bioethanol Purification. Ind. Eng. Chem. Res. 54, 2208-2213.

D. Sun, H. Tian, J. Sun, W. Xu, 2021. Reactive Distillation for the Production of Cyclohexanol: Experiments and Simulations. Chemical and Biochemical Engineering Quarterly 34, 223-232.

J. Ye, J. Li, Y. Sha, H. Lin, D. Zhou, 2014. Evaluation of Reactive Distillation and Side Reactor Configuration for Direct Hydration of Cyclohexene to Cyclohexanol. Industrial & Engineering Chemistry Research 53, 1461-1469.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50110-X

Sustainable Process Intensification of Refrigerant Mixture Separation and Management: A Multiscale Material Screening and Process Design Approach

Mohammed Sadaf Monjur, Ashfaq Iftakher, M. M. Faruque Hasan*

Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX 77843-3122, USA.
hasan@tamu.edu

Abstract

High global warming potential of refrigerant gases have prompted immediate attention to ensure minimum usage and recovery of hydrofluorocarbons (HFCs). Due to the azeotropic nature of HFC mixtures, advanced separation processes are required for selective separation and recovery. We perform a multiscale separation, intensification and material-to-process systems analysis of R-410A, which is a HFC mixture of R-32 and R-125 refrigerants, using extractive distillation (ED) and ionic liquids (ILs) as solvents. Under different design objectives, we demonstrate improved process performance of the obtained flowsheets in terms of process economics, energy consumption and sustainability. Specifically, when minimization the overall separation cost, we achieve upto 21% lower cost than that of a base design. Additionally, when minimizing CO_2-eqv emission, we achieve process configurations with up to 60% and 50% reductions in energy and emission, respectively. We also performed an IL-process performance mapping considering both cost and equivalent energy consumption as functions of IL-selectivity. Our analysis shows that R-32 selective ILs would lead to processes with better performances. For such cases, the optimal IL candidates would most likely have Henry's constant between 0.4 and 1.6 MPa with energy consumption as low as 500 kJ/kg of HFC mixture.

Keywords: Sustainable Process Intensification; Process Synthesis; Material Screening; Refrigerant Separation; Extractive Distillation

1. Introduction

R-410A belongs to a family of hydrofluorocarbons (HFCs) which are commonly used as refrigerants in domestic and commercial cooling systems. Due to lower ozone depletion potentials, HFCs are predominantly used to replace chlorofluorocarbons (CFCs). However, high global warming potential (GWP) plagues the sustainable usage of HFCs. About 2–3% of the total global greenhouse gas (GHG) emission is due to the millions of tons of worldwide HFC emission (Purohit and Höglund-Isaksson, 2017). To combat this issue, the Kigali Amendment to the Montreal Protocol recommended cutting the global HFC emissions by 80–85% by 2047 (Pardo et al., 2021). In addition, low GWP refrigerants (e.g., hydrofluoroolefins, hydrocarbons, etc.) are proposed to be used in place of HFCs. Therefore, it is desirable to minimize the amount of already existing HFCs. Incineration is not a practical option since it would result in the release of a large

amount of CO_2 into the atmosphere. Recycling, on the other hand, may have a market value of more than a billion dollars in the U.S. At the same time, HFCs can be chemically converted to low GWP hydrofluoroolefins (Pardo et al., 2021). However, recycling HFCs pose a significant challenge as they often form azeotropes or close-boiling mixtures that often behave as single fluid. Consequently, conventional separation techniques, such as cryogenic distillation are highly energy and cost-intensive, and pose significant operational challenges (Pardo et al., 2021). Process intensification techniques, such as extractive distillation (ED), can be used to resolve these challenges (Tian et al., 2018; Demirel et al., 2019). ED processes depend on suitable solvent selection. To combat this issue, ionic liquids (ILs) have garnered significant attention as potential solvents for ED-based separation of the azeotropic HFC mixtures. More specifically, ILs can selectively absorb a particular refrigerant from a mixture of HFCs, thereby improving the efficiency of absorption refrigeration processes. In addition, the negligible vapor pressure of ILs ensures that the toxicity and the contamination of the refrigerant gas with solvent can be prevented (Faúndez et al., 2013).

Although, several works have been reported in the literature on extractive separation applications (Finberg and Shiflett, 2021; Faúndez et al., 2013), there is a need for systematic analysis of ED performance for different ILs as solvents. In this work, we report SPICE_ED (Synthesis and Process Intensification of Chemical Enterprises Involving Extractive Distillation) framework for detailed process synthesis, intensification and optimization. We demonstrate the framework using two ILs, namely $[C_4C_1im][PF_6]$ and $[C_4C_1im][BF_4]$ and carry out rigorous ED modeling and analysis to investigate the process performance in terms of energy requirement, sustainability, and process economics. We extend our analysis to hypothetical ILs by utilizing the temperature dependent binary interaction parameters of vapor-liquid equilibrium (VLE) thermodynamics and perform a material-property-process-performance mapping. We anticipate that such a mapping will direct the experimental efforts in synthesizing new ILs with superior performance.

2. Problem Representation

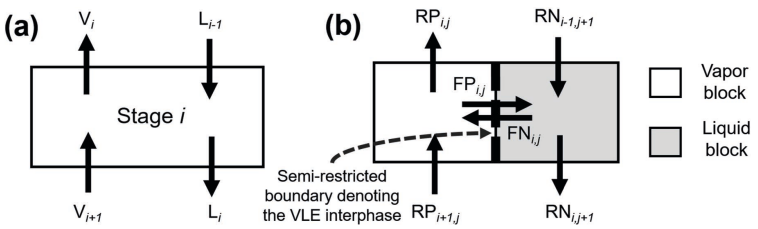


Figure 1: Distillation column tray: (a) conventional representation, (b) equivalent building block-based representation.

In this section, we provide a brief overview of the SPICE_ED framework which is then used for the detailed synthesis and optimization of R-410A separation process. In this framework, we express the physicochemical phenomena through the building-block based representation (Demirel et al., 2017; Monjur et al., 2021a,b), where sets of blocks are arranged in a two-dimensional grid to represent a superstructure. Eqs. 1–4 provide a

simplified (not exhaustive) formulation of the superstructure. We denote the position of blocks in the superstructure by i (row number) and j (column number). Figure 1a depicts the conventional representation of a single ED column tray, whereas Figure 1b depicts building block-based representation of a single tray. Two blocks are used to represent the two phases, and the dashed vertical line between the blocks represent the phase boundary. In this manner, we require two series of blocks to represent the whole ED column. We represent the material balance for each component k (Eq. 1), where material flow rates in vertical and horizontal directions are denoted by $R_{i,j,k}$ and $F_{i,j,k}$, respectively. We denote flow rates of fresh feed f, product p, and jump streams from block (i, j) to block $(i^{'}, j^{'})$ by $M_{i,j,k,f}$, $N_{i,j,k,p}$, and $J_{i,j,i',j',k}$ respectively. We express the energy balance for each block (Eq. 2), which considers stream enthalpies along with work and heat from external utility sources.

$$F_{i,j-1,k} + R_{i-1,j,k} - F_{i,j,k} - R_{i,j,k} + \sum_{f \in FS} M_{i,j,k,f} - \sum_{p \in PS} N_{i,j,k,p} + \sum_{(i',j') \in Link} J_{i',j',i,j,k} - \sum_{(i',j') \in Link} J_{i,j,i',j',k} = 0, \quad \forall\ i,j,k \tag{1}$$

$$EF_{i,j-1} + ER_{i-1,j} - EF_{i,j} - ER_{i,j} + EM_{i,j} - EN_{i,j} + EJ^{f}_{i,j} - EJ^{p}_{i,j} + W^{comp}_{i,j} - W^{exp}_{i,j} + W^{pump}_{i,j} - W^{val}_{i,j} + Q^{h}_{i,j} - Q^{c}_{i,j} = 0, \quad \forall\ i,j \tag{2}$$

$$\mathbf{f}^{ph}_{i,j,k,ph} = \mathbf{f}^{equil}(K^{eq}_{i,j,k,ph}, T_{i,j}, P_{i,j}, y_{i,j,k}, y_{i,j+1,k}, z_{i,j,ph}), \quad \forall\ i,j,k, \quad ph \in Equil \tag{3}$$

$$\mathbf{f}^{ph}_{i,j,k,ph} - M(1 - z_{i,j,ph}) \le y_{i,j,k} \le \mathbf{f}^{ph}_{i,j,k,ph} + M(1 - z_{i,j,ph}), \quad \forall\ i,j,k,ph \tag{4}$$

For the R-410A separation process, we rely on VLE based ED process. Therefore, the phenomena set ph in Eq. 3 contains VLE-based separation phenomena. The equilibrium based separation depends upon equilibrium rate constant, temperature, pressure, and composition, which are denoted by $K^{eq}_{i,j,k,ph}$, $T_{i,j}$, $P_{i,j}$, and $y_{i,j,k}$, respectively. Based on the value of the binary variable, $z_{i,j,ph} \in \{0, 1\}$, the appropriate separation phenomena is activated (Eq. 4). Following previous work (Shiflett and Yokozeki, 2006), we model the solubility of the HFCs in the IL using the Gamma-Phi based method, where the liquid phase activity coefficient is estimated by the NRTL model.

3. Extractive Distillation Process Synthesis for R-410A Separation

3.1. Base Design

For the separation of R-410A mixture, we consider a typical ED based process configuration as base case design, which is shown in Figure 2. $[C_4C_1im][PF_6]$ is used as the solvent to break the azeotrope. R-410A enters the ED column at stage 21 with a flow rate of 100 kg/h, while the recirculating IL enters at stage 2. The ED column operates at 10 bar with a reflux ratio of 0.25 and has total 28 stages. As R-32 is more soluble in $[C_4C_1im][PF_6]$, it leaves the column with the IL at the bottom while the less soluble R-125 separates out as distillate product at the top. The rich IL from the bottom of the ED column is sent to two sequential flash separators for regeneration, which operate at 1 bar and 0.1 bar, respectively. The combined vapor product from the flash separators has a R-32 purity of 99.9 mol%. The IL from Flash 2 is sent back to the ED after being pumped and cooled.

The overall process has equivalent energy consumption of 2078 kJ/kg R-410A, 0.073 kg CO_2-eqv/kg R-410A emission, and separation cost of \$0.081/kg R-410A.

3.2. Process Optimization

We optimize the base design by taking into consideration the detailed phenomena level process synthesis. Here, we limit our process-scale analysis to two commonly used ILs as absorbents, $[C_4C_1im][PF_6]$ and $[C_4C_1im][BF_4]$. For each of the cases, we vary the number of trays between 15 and 40 and allow R-410A to enter at any tray. We also consider heat integration between the hot and the cold process streams. The minimum purity that must be achieved is set to be at least 99.5 mol% for both R-32 and R-125.

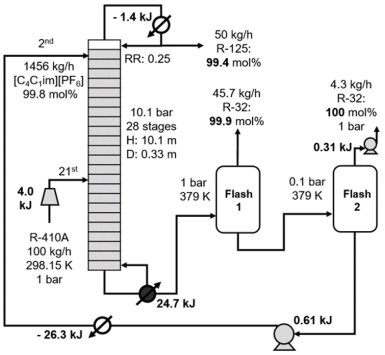


Figure 2: Process flowsheet of base design.

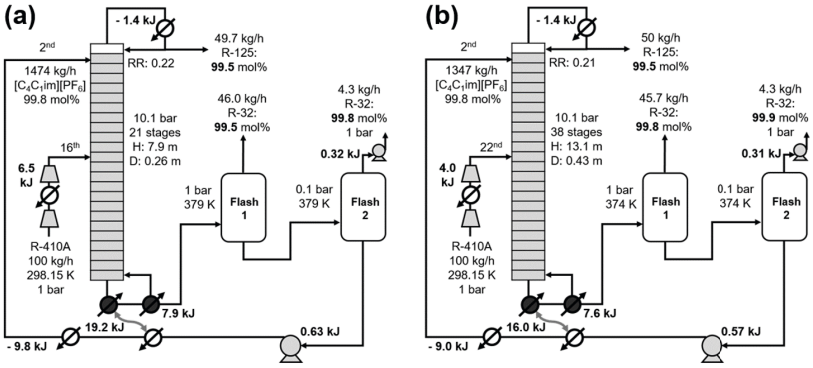


Figure 3: Optimized extractive distillation process with $[C_4C_1im][PF_6]$ as solvent for R-410A separation. Design targets are (a) minimum separation cost, and (b) minimum CO_2-eqv emission.

When $[C_4C_1im][PF_6]$ is used as solvent, Figure 3a shows the optimized process flow sheet with 21 total stages for the design objective of minimization of separation cost (\$0.071/kg R-410A). As shown in Table 1, it offers 12% lower separation cost than that of the base design. It also achieves 60% reduction in energy consumption. Due to the heat integration, 70% of the re-boiler duty is supplied from the circulating IL, while 66% of the cooling duty for the IL is provided by the re-boiler. Consequently, the required sizes of the heat exchanger and re-boiler reduce. We then set CO_2-eqv emission as the design target, and obtain a process flowsheet with 38 stages (Figure 3b). Note that, an increase in the number of equilibrium stages (i.e. trays) improves the separation and consumes less energy. Compared to the base design, we achieve 50% and 60% reduction in CO_2-eqv emission and energy consumption, respectively. However, the increased number of trays increases the capital cost and results in increased separation cost.

Table 1: Process Performance Comparison.

Design targets	Base design	$[C_4C_1im][PF_6]$	Improvement (%)	$[C_4C_1im][BF_4]$	Improvement (%)
Equivalent energy (kJ/kg R-410A)	2078	824	**60**	852	59
CO_2 emission (kg/kg R-410A)	0.073	0.036	**50**	0.038	48
Separation cost ($/kg R-410A)	0.081	0.071	12	0.063	**21**

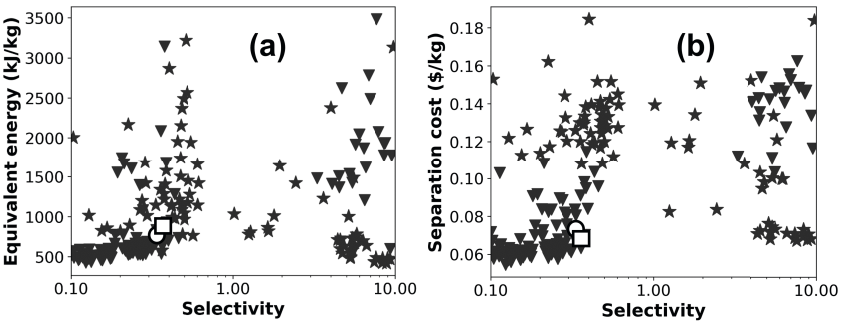


Figure 4: Selectivity mapping of ILs for (a) minimum equivalent energy, and (b) minimum separation cost. The triangle (▼) represent the designs with 99.5 mol% minimum purity and the star (★) denote the designs with 97 mol% minimum purity. $[C_4C_1im][PF_6]$ is represented by the circle (○) and $[C_4C_1im][BF_4]$ by the square (□).

Next, we consider $[C_4C_1im][BF_4]$ as solvent, and perform similar optimization. As shown in Table 1, when the design target is the minimization of separation cost, we achieve 21% reduction compared to the base design. The corresponding ED column requires 20 stages. Unlike the previous designs, here the R-410A enters in two different trays (trays no. 16 and 17). Notably, we achieve 48% reduction in CO_2-eqv emission compared to the base design under the design target of CO_2-eqv emission minimization. The corresponding process configuration requires ED column with 37 trays.

3.3. Mapping of IL Selectivity, Separation Energy and Cost

As the cations and the anions forming an IL can be arranged in many different combinations, it is possible to consider millions of hypothetical ILs. However, exhaustive synthesis and the corresponding process-scale analysis of all the ILs is impractical. To address this issue, we aim to reduce the search space of optimal ILs for R-410A separation. The motivation arises from the need to construct hypothetical ILs and predict the corresponding process performance to direct the experimental efforts. It is well known that the binary interaction parameters of a particular IL/HFC system dictates its solubility. Therefore, we focus on generating feasible binary interaction parameters via latin hypercube sampling, and construct nearly 3000 hypothetical ILs. When each of the ILs are used as solvents in the SPICE_ED optimization framework, not all are able to achieve the required R-32 and R-125 purity. For example, out of 3000 ILs, only 200 satisfy 97 mol% purity while only 117 ILs satisfy 99.5 mol% purity constraint. The equivalent energy consumption and separation cost of all the ILs (200 hypothetical, $[C_4C_1im][PF_6]$, $[C_4C_1im][BF_4]$) are shown in Figure 4. Note that the selectivity is

defined as the ratio of Henry's constant of R-32 over R-125 at 298.15 K. By definition, when the selectivity < 1, R-32 is more soluble in the IL and vice versa. We deduce three key points: a) when the selectivity is between 0.5 and 2.0, IL may not be a suitable solvent candidate for HFC separation, since the required 99.5 mol% purity is not achieved, (b) R-32 favourable (more selective) IL is desired, as the corresponding process consumes less energy and requires less separation cost, (c) The Henry's constant value of R-32 is of particular importance, since for the same selectivity, ILs can have different equivalent energy consumption. Interestingly, when the target is less than 1000 kJ/kg R-410A equivalent energy with 99.5 mol% purity, more than 86% ILs have Henry's constant value of R-32 between 0.4 and 1.6 MPa. It is worth noting that, prediction of important IL properties (e.g., viscosity, heat capacity, and heat of absorption) which might impact the overall process performance, is beyond scope of this study.

4. Conclusions

Due to the azeotropic nature of HFC mixtures, conventional separation technologies are not suitable. In an effort to address this issue, we employed extractive distillation as a means for HFC separation process intensification. Many favourable properties of ionic liquids make them an attractive candidates as solvent for the absorption of selective refrigerants in HFC mixtures in ED columns. We developed SPICE_ED framework to achieve improved process configurations for IL-based extractive distillation. Our analysis show that use of $[C_4C_1im][BF_4]$ would result in lower separation cost, while $[C_4C_1im][PF_6]$ would result in lower energy consumption and emission. We extended our analysis to hypothetical ILs by utilizing the temperature dependent binary interaction parameters and performed a multiscale mapping of IL selectivity and energy/cost. We conclude that ILs offering more solubility towards R-32 would lead to better process performance in terms of energy consumption. We also conclude that the optimal IL candidates would most likely have exhibit Henry's constants between 0.4 to 1.6 MPa for R-32. The overall energy consumption can be also reduced to as low as 500 kJ/kg and separation cost as low as \$0.06/kg of HFC.

5. Acknowledgment

The authors gratefully acknowledge the support from NSF EFRI DChem Award 2029354.

References

S. E. Demirel, J. Li, M. M. F. Hasan, 2017. Comput. & Chem. Eng. 105, 2–38.

S. E. Demirel, J. Li, M. M. F. Hasan, 2019. Current Opinion in Chemical Engineering 25, 108–113.

C. A. Faúndez, L. A. Barrientos, J. O. Valderrama, 2013. Int. J. of Refrigeration 36 (8), 2242–2250.

E. A. Finberg, M. B. Shiflett, 2021. Ind. Eng. Chem. Res.

M. S. Monjur, S. E. Demirel, J. Li, M. M. F. Hasan, 2021a. Ind. Eng. Chem. Res.

M. S. Monjur, S. E. Demirel, J. Li, M. M. F. Hasan, 2021b. Comput. Aided Chem. Eng 50, 287–293.

F. Pardo, S. V. Gutiérrez-Hernández, G. Zarca, A. Urtiaga, 2021. ACS Sust. Chem. Eng 9 (20), 7012–7021.

P. Purohit, L. Höglund-Isaksson, 2017. Atmos. Chem. Phy. 17 (4), 2795–2816.

M. B. Shiflett, A. Yokozeki, 2006. AIChE J. 52 (3), 1205–1219.

Y. Tian, S. E. Demirel, M. M. F. Hasan, E. N. Pistikopoulos, 2018. Chem. Eng. Proc. Proc. Int 133, 160–210.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50111-1

A systematic methodology for the optimisation, control and consideration of uncertainty of reactive distillation

A. Tsatse[a], S. R. G. Oudenhoven[b], A. J. B. ten Kate[b], E. Sorensen[a]*

[a]*Department of Chemical Engineering, University College London, Torrington Place, WC1E 7JE London, UK;*
[b]*Nouryon, Zutphenseweg 10, 7418 AJ Deventer, the Netherlands*
e.sorensen@ucl.ac.uk

Abstract

The main goal of this work is the development of a structured optimisation strategy for reactive distillation systems in order to prevent production failures due to operational disturbances and/or uncertainties in the design input. The framework developed is demonstrated using a case study of industrial interest, based on a systematic evaluation of optimal design and control alternatives, offering the possibility of revising the design and/or operation of the process in order to minimise the risk of production failures. This framework can be used in an early design stage to quantify the impact of specific input parameters such as kinetics on process performance and costs, and it can therefore be used to focus the experimental effort on determining the most critical parameters for the performance of a reactive distillation process.

Keywords: reactive distillation, design, operation, control, uncertainty

1. Introduction

Reactive distillation is an intensified process where reaction and separation are integrated into a single unit. However, the large difference in time scales of the two processes, where reaction typically requires minutes for significant conversion whilst VLE is almost instantaneous, means that the design and control strategies cannot be easily extended from conventional distillation. This challenge is even further amplified when disturbances and design uncertainties are considered within the process.

Various optimisation methods have been applied in the open literature for the design of reactive distillation processes (e.g. Ciric and Gu 1994, Tian et al. 2020 etc.), some of which considering process control as part of the optimisation problem (e.g. Contreras-Zarazúa et al. 2017, Bernal et al. 2018 etc.). A few contributions have discussed the impact of process design on controllability (e.g. Mansouri et al. 2015, Georgiadis et al. 2002 etc.) whilst others have focused on the design of reactive distillation processes capable to successfully operate under uncertainty (e.g. Paramasivan and Kienle 2012, Tian et al. 2020 etc.).

Although reactive distillation design, controllability and uncertainty have all been previously studied, no systematic methodology has combined all those elements to investigate potential process modifications to improve performance under process disturbances and/or market changes. This is therefore the aim of this work, to provide a

framework for the design of reactive distillation processes which are capable of dealing with production failures due to design and/or operational deficiencies. The framework and associated benefits will be demonstrated here by considering input uncertainty (particularly in reaction kinetics) on an optimal reactive distillation process.

2. Methodology

The methodology demonstrated in this work is presented in Figure 1. The methodology starts with determining the optimal design and operational parameters of the process, for instance using a methodology based on superstructure optimisation (Tsatse et al. 2021), then different process control configurations are evaluated. If, given the control scheme and the process disturbances considered, the system is not able to meet specifications and tolerate the (short-term) disturbances introduced, then the design must be revised in order to improve its control performance. If, however, specifications are met and disturbances are tolerated, the design is provisionally acceptable. In the final step, which is the focus of this work, the performance of the process is evaluated under input uncertainty. Process disturbances can be considered simultaneously with input uncertainty, although only the latter is the focus of this work. If the system cannot tolerate these (i.e. is failing to meet specifications) then the design and/or its control configuration must be revised to increase its flexibility (i.e. tolerance to uncertainties). If the system is not sensitive to the uncertainties, then the engineer can be confident that the process designed is not only economically attractive, but also capable of mitigating production failure issues due to design and/or operational deficiencies and model parameter uncertainties. The software used in this work for process simulation, optimisation and control is gPROMS ProcessBuilder v1.3.1 (Process Systems Enterprise 2021). For the uncertainty simulations in particular, the Global System Analysis (GSA) tool within ProcessBuilder was additionally used. The GSA tool is used to perform multiple model evaluations (simulations) with selected (i.e. uncertain) model input, as will be shown in section 4.

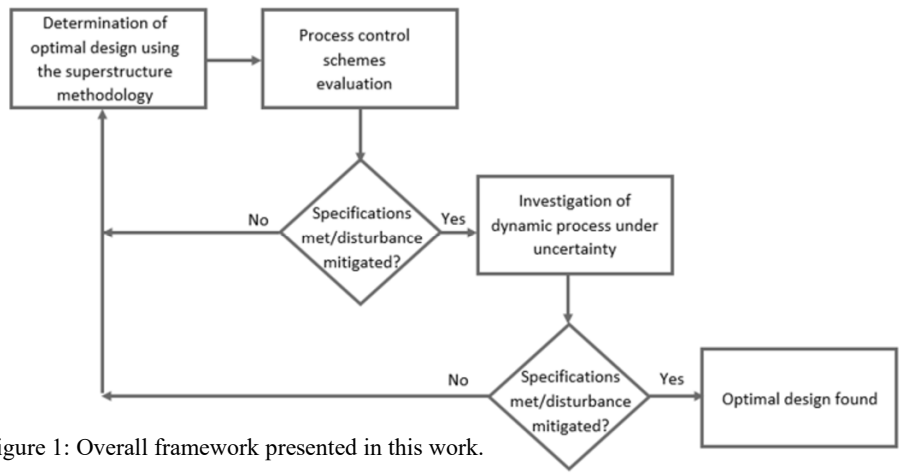


Figure 1: Overall framework presented in this work.

3. Case study

In this section, the case study used for the demonstration of the methodology is presented. This case study will only focus on the investigation of input uncertainty as the optimal design and control of the process have been previously considered in separate investigations.

A quaternary system is considered, in which the following auto-catalysed reversible reaction occurs in the liquid phase with component D as the desired product:

$$A + B \leftrightarrow C + D$$

The kinetic expressions for the forward (f) and backward (b) reaction rates are the following:

$$r_f = k_{f0} e^{-Ea_f/RT} C_A C_B$$
$$r_b = k_{b0} e^{-Ea_b/RT} C_C C_D$$

where reaction rate, r, is expressed in kmol/(m^3·s), pre-exponential kinetic factors, k_{f0} and k_{b0}, are expressed in m^3/(kmol·s), activation energy, E_a, is expressed in kJ/mol (assumed to be 80 kJ/mol for both directions), and component concentration C_i is expressed in kmol/m^3. Heat of reaction was assumed to be negligible, thus the activation energy is the same for both reaction directions and K_{eq} is independent of temperature, based on the previous assumptions. More details for the system considered and the underlying assumptions can be found in Tsatse et al. (2021) for Case study 15 in particular.

The feed streams to the overall system were one stream of reactant B of flow rate 12.6 kmol/hr, and one stream of reactant A below of the same flow rate (1:1 feed molar ratio). This corresponds to approximately 5 ktn/year of product D for full reactant conversion, which is met for production rates larger than 12.55 kmol/hr of component D. The feeds were assumed to be at their boiling points at 1 atm. Liquid hold-up of the reactive distillation column was assumed equal for all reactive stages, fixed at 0.1 m^3/reactive tray.

V-only control was applied to the system, where pressure at the top of the column (stage 2) is controlled by the condenser duty (PI control, $K_c = 20$ and $\tau = 12$ min) and the liquid levels of the reflux drum (P-only, $K_c = 2$) and the sump (P-only, $K_c = 2$) are controlled by the distillate flow rate and bottoms flow rate, respectively. Reboiler duty is manipulated in order to control bottom product purity (PI control, $K_c = 3$ and $\tau = 25$ min).

In this work, process disturbances such as feed composition change etc. (second step of the framework presented in Figure 1) will not be considered, in order to focus on input uncertainties only. For reactive distillation, critical input is mainly reaction kinetics and VLE and in early design stages, these are typically known with limited accuracy. Proper understanding of their impact on process performance and cost can therefore contribute to focused experimental effort, leading to a carefully designed and rigorous process.

Uncertainty in reaction kinetics is considered for the controlled process, using two cases (case a: uncertainty in k_{f0} so varying k_{b0} in order to keep K_{eq} at the base case value, case b: uncertainty in K_{eq} so fixing k_{f0} and varying k_{b0} to form the desired K_{eq} uncertainty values as $K_{eq}=k_{f0}/k_{b0}$)). For both cases, the uncertainty range considered was ±50% based on industrial experience and the two pre-exponential factors were grouped and varied as a multivariate enumerated set (i.e. 100 pairs or samples of the predefined values of the two pre-exponential factors to uniformly cover the uncertainty range). Based on the base-case values presented in Table 1, the value ranges considered are therefore:

a) $15.138 < k_{f0} = 30.276 < 45.414$ (10^9 m^3/(kmol·hr)) so
$6.728 < k_{b0} = 13.456 < 20.184$ (10^9 m^3/(kmol·hr)) to keep K_{eq} constant at 2.25

b) $1.125 < K_{eq} = 2.25 < 3.375$, and since $k_{f0} = 30.276 \cdot 10^9$ m^3/(kmol·hr)
$8.971 < k_{b0} = 13.456 < 26.912$ (10^9 m^3/(kmol·hr))

The Key Performance Indicators (KPIs) for the case study considered were the bottom product purity ($x_{B,D}$), which should be maintained at 0.99 mol/mol; bottom production rate (B), which should be higher than 12.55 kmol/hr; condenser (Q_C) and reboiler (Q_R) duties; as well as the total cost of the process (Production TAC) .

4. Results and Discussion

In this work, the uncertainty simulations using GSA in gPROMS ProcessBuilder v1.3.1 (Process Systems Enterprise 2021) needed approximately 0.2-5 min CPU time. The short times are due to the fact that the number of samples (100) and the number of factors (k_{f0} and k_{b0}) and responses (controlled and manipulated variables, objective function value) was moderate, as was the complexity of the flowsheet (a single reactive distillation column in this case). The initial optimal design of the case study considered was found based on the superstructure methodology and a cost objective function, subject to product quality constraints, whilst the main product of interest was component D (see Tsatse et al. 2021 for more details). The base-case input as well as the initial optimal results are presented in Table 1.

Figure 2 shows the results of the simulations when considering uncertainty in reaction kinetics (left: case a, right: case b) for the initial (lines indicated as *initial*) as well as the flexible (lines indicated as *flexible*) system. Note that *flexible system* refers to the revised process, as presented in Table 1.

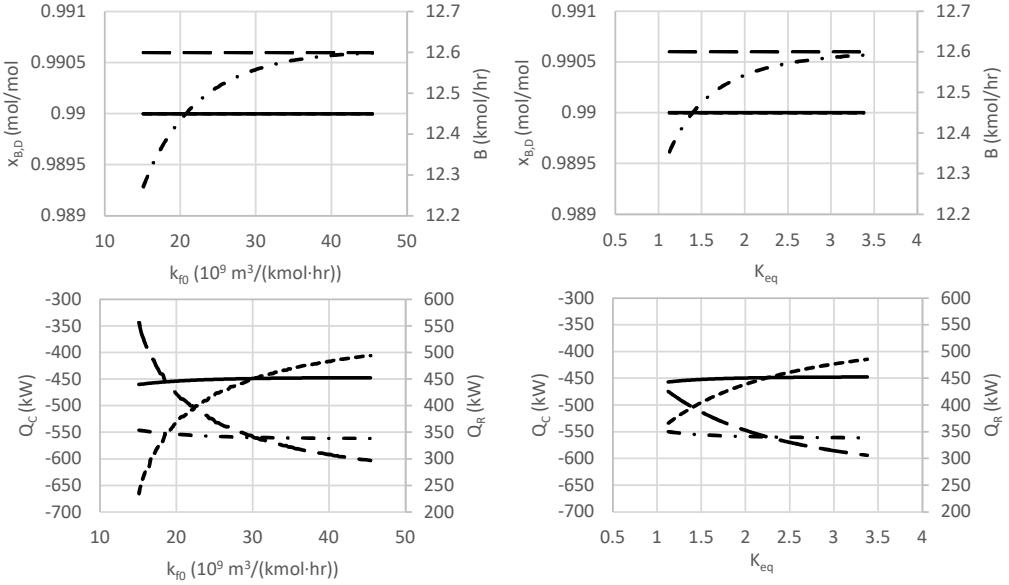


Figure 2: Uncertainty simulations. Product purity ($x_{B,D}$) together with bottom production rate (B), condenser (Q_C) and reboiler (Q_R) duties and flexible optimal dynamic controlled (V-only) design. Uncertainty in kinetics (case a-left, case b-right) is considered.

Initial design is indicated as: —— for left y-axis, and as: — · — for right y-axis. Flexible design is indicated as: – – for left y-axis, and as: — — for right y-axis.

Figure 2 (left: case a, right: case b) shows that under the uncertainty considered, the product purity ($x_{B,D}$) is maintained (top of Figure 2) by the V-only controlled system, but this is only possible by reducing the bottom production rate (i.e. B < 12.55 kmol/hr) when k_{f0} drops below $28.6 \cdot 10^9$ $m^3/(kmol \cdot hr)$). This means that with the initial optimal design and control configuration, slower kinetics (case a) down to $28.6 \cdot 10^9$ $m^3/(kmol \cdot hr)$) can be mitigated by control action alone. Similarly, lower chemical equilibrium (down to 2.08) can be mitigated by control action whilst lower values lead to violation of the desired production flow rate (right in Figure 2). Figure 2 shows that for this design and parameter set and the uncertainties considered, slower kinetics (k_{f0}) have a more significant impact on the performance than lower K_{eq}. The 50% uncertainty range considered for K_{eq} corresponds to a range of 51.5% to 64.8% conversion (the base case conversion, i.e. for K_{eq}=2.25, is 60%) which is not a broad range and can easily occur.

As the initial optimal design is not able to tolerate the entire range of uncertainty considered, a mitigation strategy must be applied which in this case is revision of column design. Re-optimising the reactive distillation column using the methodology described in Tsatse et al. (2021) leads to the optimal parameters of the new, flexible process as shown in Table 1. The uncertainty range is now re-considered, this time in the new flexible V-only controlled system and the results are shown in Figure 2. The system is now able to tolerate the entire range of uncertainty in kinetics considered, not only in terms of product purity (note that $x_{B,D}$ lines of initial and flexible design overlap) but also maintaining production rate at the desired level. Changes in condenser and reboiler duties are more significant for slow kinetics than for lower chemical equilibrium.

Table 1: Optimal results for the initial and flexible design for the case study considered (stages numbered from the top, condenser = 1, reboiler = N_T).

	Initial optimal design	**Flexible design**
k_{f0} ($m^3/(kmol \cdot hr)$)	$30.276 \cdot 10^9$	
K_{eq}	2.25	
α_{CA} - α_{AB} - α_{BD}	1.2 - 2.5 - 2	
	Values in optimal design	
Heavy feed (B) stage (N_{T1})	9	12
Light feed (A) stage (N_{T2})	23	26
Number of stages (N_T)	27	31
Reflux ratio (RR,-)	3.7	3.5
Bottoms flow rate (B, kmol/hr)	12.6	12.6
Reactive stages	2-26	2-30
Column diameter (D_C, m)	0.71	0.70
Bottom purity ($x_{B,D}$)	0.99	0.99
Operating cost (OPEX, M€/yr)	10.49	10.50
Capital cost (CAPEX, M€/yr)	0.31	0.41
Production TAC (€/kg)	2.150	2.167

This particular case study demonstrates revision of process design as the mitigation strategy is employed. However, alternative mitigation strategies, i.e. revision of control configuration and revision of the entire process including addition of ancillary units, are also supported within the framework. Since the entire range of uncertainty in reaction kinetics is now tolerated, the engineer can have confidence that the new flexible V-only controlled design can mitigate the uncertainty considered, here up to ±50% uncertainty in kinetics. Also, compared to the initial column design, the new flexible design includes four additional stages for reaction and separation to tolerate the uncertainty in kinetics considered, with less than a 1% increase (which mainly stems from the increase in capital cost) in the objective function as shown in Table 1 indicating a cost-effective alternative.

5. Conclusions

The findings indicate that, as uncertainty is expected in a reactive distillation process, a structured methodology to quantify its impact is essential as an economically optimal steady state design might nevertheless not operate successfully under process and input uncertainties leading to an unsuccessful project evaluation. Revision of its design and control strategy may therefore be required to improve its robustness. As this revision is associated with increased cost, the framework thus provides a basis to assess the relative benefits (process robustness vs cost) helping to make a more profound business decision.

References

D.E. Bernal, C. Carrillo-Diaz, J.M. Gómez and L.A. Ricardez-Sandoval, 2018, "Simultaneous Design and Control of Catalytic Distillation Columns Using Comprehensive Rigorous Dynamic Models." Industrial & Engineering Chemistry Research 57 (7), 2587– 2608.

A.R. Ciric and D. Gu, 1994, Synthesis of nonequilibrium reactivedistillation processes by MINLP optimization. AIChE J. 40 (9),1479–1487.

G. Contreras-Zarazúa, J.A. Vázquez-Castillo, C. Ramírez-Márquez, J.G. Segovia-Hernández and J.R. Alcántara-Ávila, 2017, "Multi-objective optimization involving cost and control properties in reactive distillation processes to produce diphenyl carbonate." Computers & Chemical Engineering 105: 185-196.

M.C. Georgiadis, M. Schenk, E.N. Pistikopoulos and R. Gani, 2002, "The interactions of design control and operability in reactive distillation systems." Computers & Chemical Engineering 26(4): 735-746.

S.S. Mansouri, M. Sales-Cruz, J.K Huusom, J.M. Woodley and R. Gani, 2015, "Integrated Process Design and Control of Reactive Distillation Processes." IFAC-PapersOnLine 48(8): 1120-1125.

G. Paramasivan and A. Kienle, 2012, "Decentralized Control System Design under Uncertainty Using Mixed-Integer Optimisation." Chemical Engineering & Technology 35(2): 261-271.

Process Systems Enterprise, 2021, gPROMS, https://www.psenterprise.com/products/gproms/process, 1997-2021.

Y. Tian, I. Pappas, B. Burnak, J. Katz, E.N. Pistikopoulos, 2020, ASystematic Framework for the synthesis of operable processintensification systems — reactive separation systems. Comput. Chem. Eng. 134, 106675.

A. Tsatse, S.R.G. Oudenhoven, A.J.B. ten Kate and E. Sorensen, 2021, "Optimal design and operation of reactive distillation systems based on a superstructure methodology." Chemical Engineering Research and Design, 170, 107-133.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50112-3

Equation Oriented Optimization of Multi Stream Heat Exchanger Design and Operation in Natural Gas Liquefaction Process

Saif R. Kazi[a], Lorenz T. Biegler[a*], Rahul Gandhi[b]

[a]*Department of Chemical Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15232, USA*
[b]*Shell Technology Centre, Houston, TX 77032, USA*
lb01@andrew.cmu.edu

Abstract

The natural gas liquefaction process is a cryogenic energy intensive process which requires a complex designed multi-stream heat exchanger (MHEX). In this study, we present a discrete model for the detailed design of a spiral wound heat exchanger (SWHX) used in the natural gas liquefaction process. The design model is derived by discretizing first principles heat equation inside the heat exchanger for multiple refrigerant streams and natural gas. The phase change (liquefaction or vaporization) process is modeled using complementarity constraints which are reformulated and solved as a NLP.

The SWHX model is embedded and solved inside a flowsheet model with process constraints for feasible design operations. The optimization results show that the inclusion of detailed MHEX design inside process flowsheet models is imperative to obtain optimal solutions which can be achieved in actual process performance.

Keywords: Multi-Stream Heat Exchanger, Natural Gas, Optimization

1. Introduction

Natural gas has become the largest source of energy production in US over the last years. The increase in production has led to increase in US natural gas exports to other countries (Source: U.S. EIA, 2021). Natural gas is transported overseas in a liquefied state inside huge storage tanks as liquefied natural gas (LNG). The natural gas liquefaction (NGL) process is known to be a very energy intensive cryogenic process which can account upto 52% of the cost of LNG (Petrowiki, 2018).

There are mainly three types of natural gas liquefaction (NGL) processes i.e. a) cascade liquefaction process, b) mixed refrigerant liquefaction process and c) expander based liquefaction process. Cascade liquefaction processes consist of multiple independent pure refrigeration cycles where the natural gas is cooled using propane, ethylene and methane as refrigerants sequentially. The cascade liquefaction process has high energy efficiency compared to other type of liquefaction process but its capital costs are high because of its complex design and additional individual units required.

Mixed refrigerant (MR) liquefaction processes use a nitrogen and hydrocarbon mixture (methane, ethane, propane, i-butane and n-butane). They require fewer units than cascade refrigeration process and the energy consumption is significantly lower. MR

liquefaction process can have single (SMR) (eg. PRICO process) or dual (DMR) refrigeration loop cycles. Another type of MR liquefaction process (C3MR) has a pre-cooling stage where liquid propane (C3) is used as a refrigerant to cool the natural gas before liquefying using the MR. This increases the energy efficiency of the process but requires more heat transfer and pressure control units. The expander based liquefaction process uses pure nitrogen or methane as the refrigerant in a single stage refrigeration cycle with multiple compressor stages. Unfortunately the method has lower energy efficiency than MR processes and is not suitable for high LNG yield.

Review papers (Austbø et al. (2014); Qyyum et al. (2108)) on optimization of natural gas liquefaction processes provide an extensive analysis of different methods and studies done on the topic. Most studies used process simulation software like Aspen Plus or Hysys combined with either stochastic or deterministic optimization solvers in MATLAB. Some studies have also used equation-oriented (EO) modeling in GAMS, AMPL or gPROMS with continuous and mixed-integer optimization tools like IPOPT, CONOPT and DICOPT. Previous studies (Watson et al., 2018) have used simple enthalpy balances to model the MHEX in the NGL process and neglect the effect of the exchanger design on the thermodynamic performance of the refrigerant cycles. This could severely affect the optimization results as the MHEX design has highly nonlinear correlations with the process variables and the energy efficiency of the process. To date, only Tsay et al. (2017) have presented an equation oriented model of natural gas liquefaction process with detailed multi-stream heat exchanger design. They used heat transfer coefficient and pressure drop correlations to model temperature and pressure variation inside the MHEX in the PRICO process flowsheet. Their MHEX design model was embedded inside the PRICO flowsheet and solved using a pseudo-transient EO approach which converts system of nonlinear equations into a nonlinear differential equation system. They used shooting methods in gPROMS to integrate the differential equations and the SQP algorithm as the optimization solver to solve the problem.

In this study, a new DAE-based design model is developed for multi-stream spiral wound heat exchanger (SWHX), the type of MHEX used in LNG processes. The DAE model is derived using the heat equation representing the heat transfer between the streams. The phase change inside the exchanger is modeled using complementarity constraints. Finally the DAE model is solved within a NG liquefaction flowsheet model as a simultaneous design optimization of the complete process.

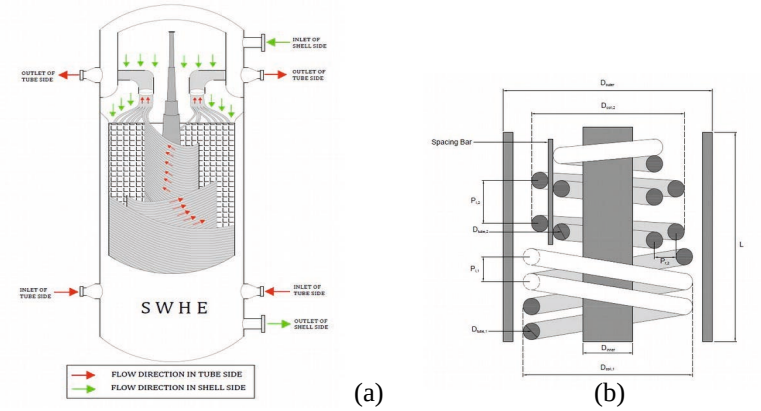


Figure 1: Inside and cross-section view of spiral wound heat exchanger

2. Model

2.1 MHEX DAE model

MHEX(s) have more complex design and geometry than single stream heat exchangers, which makes it difficult to build an optimization model for MHEX design. The SWHX consists of a central rod-like structure around which multiple tubes are coiled in a circular helix shape as shown in Figure 1a. The concentric coils have different radii and appear to form a spiral shape around the central rod. The tubeside and shellside streams enter the exchanger from opposite directions and exchange heat over the tube curved surface. Multiple streams in the tubeside (usually hot) and single stream on the shellside (usually cold) flow in a counter-current cross-flow pattern inside the MHEX as shown in Figure 2a.

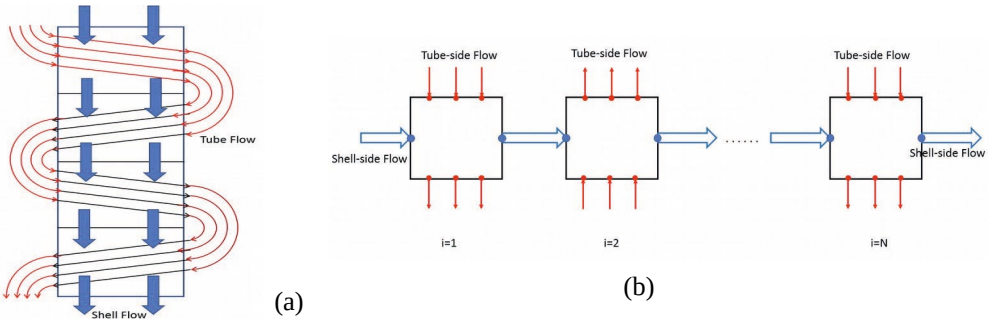


Figure 2: Discretization of SWHX into discrete finite elements

We propose a discrete model which discretizes the exchanger into discrete elements and solves energy balance using first principles heat equation inside each element as shown in Figure 2b and Figure 3.

2.2 Heat Equation

The heat equation for the streams inside SWHX is derived by discretizing the exchanger into multiple discrete finite elements. The steady-state heat equation for tubeside and shellside streams can be written as:

$$\rho_t C_t^p u_x \frac{dT_t}{dx} = -q_v \qquad \text{and} \qquad \rho_s C_s^p u_y \frac{dT_s}{dy} = q_v \tag{1}$$

where ρ is molar density, Cp is specific heat capacity, u is velocity, k is thermal conductivity, Tt , Ts are tube and shell temperature, and qv is the volumetric heat. The following assumptions are made: 1) Zero or negligible heat transfer by conduction and 2) Shell and coil side fluid flow are unidirectional.

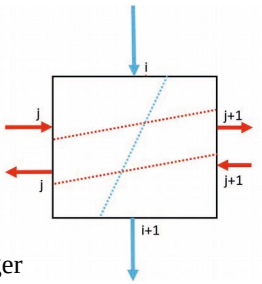


Figure 3: Discrete Element inside the exchanger

Since the streams are mixtures and are undergoing phase change, it is more meaningful to use enthalpy as the differential variable. The coupled ODE system Eq.(1) is discretized using the smaller discrete elements (shown in Figure 3) and FEM with hat functions as the basis functions resulting in the following set of equations:

For i=1...N

$$F_j\frac{(H_j^{i,2}-H_j^{i,1})}{2}+\frac{U_{j,s}\Delta A_{j,s}}{3}(T_j^{i,2}-T_s^{i+1})+\frac{U_{j,s}\Delta A_{j,s}}{6}(T_j^{i,1}-T_s^{i})=0 \qquad \forall j\in S_c$$
$$F_s\frac{(H_s^{i+1}-H_s^{i})}{2}-\sum_k^{k\in S_c}\left[\frac{U_{k,s}\Delta A_{k,s}}{3}(T_k^{i,2}-T_s^{i+1})+\frac{U_{k,s}\Delta A_{k,s}}{6}(T_k^{i,1}-T_s^{i})\right]=0 \tag{2}$$

2.3 Phase Change

The condensation (liquefaction) and evaporation process for mixed component streams can be modeled using complementarity constraints. The pure vapor, liquid and two-phase (mix of liquid and vapor) streams are in a state of vapor-liquid equilibrium (VLE). As the streams exchange heat between them, they can be imagined as a PQ flash process. The flash equations for VLE for a mixed component stream are written as:

Flash VLE Equations

$$F_j = L_j + V_j$$
$$z_{ij}F_j = x_{ij}L_j + y_{ij}V_j$$
$$y_{ij} = \beta_j K_{ij}(T_j, P_j)x_{ij}$$
$$\sum_i (y_{ij} - x_{ij}) = 0$$

and

$$\beta_j - 1 \le sV_j$$
$$\beta_j - 1 \ge -sL_j$$
$$0 \le sV_j \perp V_j \ge 0$$
$$0 \le sL_j \perp L_j \ge 0$$

(3)

The stream enthalpy (H) is written as function of its temperature (T), pressure (P) and composition (*x*,*y*) using equation of state (EOS). Additional constraints are stated to connect the state variables (H, T and P) at the element boundary as given in Kazi(2021). Also inlet and outlet temperatures are imposed as the boundary conditions.

2.4 Flowsheet

The natural gas liquefaction flowsheet is shown in Figure 4. The flowsheet model consists of mass and energy balance along with process constraints and flash equations to model the adiabatic valves. The objective function minimizes

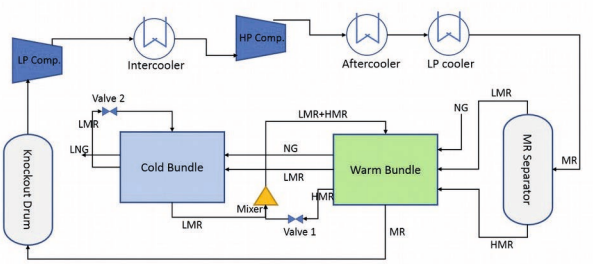


Figure 4: Natural Gas Liquefaction Flowsheet Diagram

the compressor power required to recycle the mixed refrigerant (MR) and provide enough refrigeration to liquefy natural gas. The degrees of freedom (DOF) in the optimization model are: 1) MR inlet temperature, pressure and composition, 2) Pressure drop across LMR and HMR valves and 3) Bundle break temperature (temperature between warm and cold bundles).

3. Methodology

The optimization of natural gas liquefaction process with the described MHEX DAE design model for SWHX and the NGL flowsheet model is a large scale optimization problem with nonlinearities and complementarity constraints. Therefore, it is imperative to provide a good initial point for convergence of the NLP solver. For this reason, the problem is solved in a step-by-step procedure as described:

- **Solve natural gas liquefaction flowsheet model:** The first step is solve the flowsheet model without the heat exchanger model. The model consists of only enthalpy and mass balances along with the process constraints such as EMAT, super-heated stream at compressor inlet etc.

- **Solve MHEX DAE design model:** The stream flowrates and temperatures from the flowsheet model solution are taken as inputs to the MHEX DAE design model. The DAE model is discretized and the complementarity constraints are relaxed and solved as a NLP model as in Kazi et al. (2019) .

- **Initialize and solve the combined flowsheet with detailed DAE model:** The solution from both the flowsheet and the MHEX models are used as the initial guess for the combined overall natural gas liquefaction model with detailed exchanger design. The combined optimization model is solved using standard NLP solvers such as CONOPT or IPOPT.

4. Optimization Results

The models are written using Pyomo – a Python based modeling language and the MHEX DAE model is discretized into N=20 elements each for warm and cold bundle respectively. The flowsheet model consists of 2,453 variables and 2,465 constraints, whereas the DAE model consists of 23,297 variables and 23,834 constraints. The solution time to solve the complete set of models comes to about 5 CPU min on a 16 GB RAM Intel Core i7 system with 2.70 GHz processor.

Variable	Value		MR Composition	Value
Pressure Energy Loss (Objective)	33,854.4 kW		Nitrogen	0.046
Pressure Drops (Valve 1 and 2)	3700 & 3840 kPa		Methane	0.404
Heat Duty (WB)	2,01,095 kW		Ethane	0.5
Heat Duty (CB)	97,376		Propane	0.05

Table 1: Optimal values for Flowsheet values and MR composition

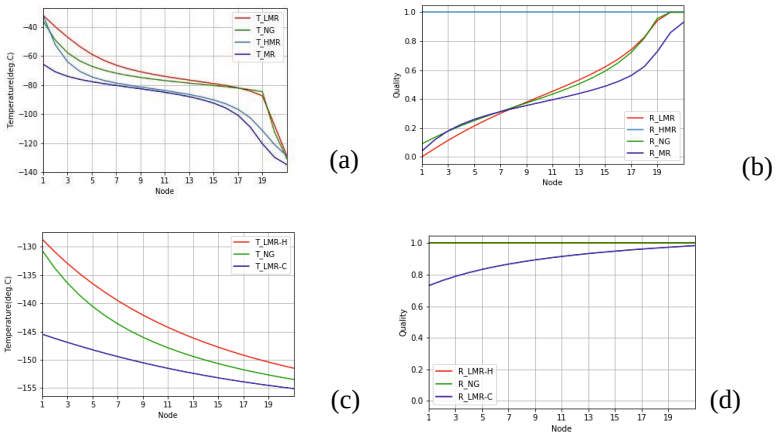


Figure 5: Temperature and liquid fraction profiles inside the exchanger ((a,b) - WB and (c,d) - CB)

5. Conclusions

A DAE model for SWHX-type MHEX is developed based on first principles heat equation. An initialization/solution strategy is presented to robustly and simultaneously solve the combined flowsheet with detailed MHEX design model. The results show that the inclusion of detailed exchanger design inside flowsheet models is able to reduce the objective value by 8% (36,840kW to 33,854kW of compressor power) as compared to the smaller flowsheet model without detailed DAE heat exchanger models. Future work will extend the DAE model to account for equipment design variables including shell & tube diameter, number of coils and geometric factors.

References

Source: U.S. Energy Information Administration, Sep 2021, Monthly Energy Review

PetroWiki, 2018, "Liquified Natural Gas" https://petrowiki.spe.org/Liquified_natural_gas_(LNG)

Bjørn Austbø, Sigurd Weidemann Løvseth, Truls Gundersen, 2014, "Annotated bibliography - Use of optimization in LNG process design and operation", *Computers & Chemical Engineering*, 71, 391-414,

Muhammad Abdul Qyyum, Kinza Qadeer, and Moonyong Lee, 2018, "Comprehensive Review of the Design Optimization of Natural Gas Liquefaction Processes: Current Status and Perspectives", *Industrial & Engineering Chemistry Research*, *57* (17), 5819-5844

H. A. Watson, M. Vikse, T. Gundersen, and P. I. Barton, 2018, "Optimization of single mixed-refrigerant natural gas liquefaction processes described by nondifferentiable models," Energy, 150, 860–876

C. Tsay, R. C. Pattison, and M. Baldea, 2017, "Equation-oriented simulation and optimization of process flowsheets incorporating detailed spiral-wound multistream heat exchanger models," AIChE Journal, 63, 9, 3778–3789

S. R. Kazi, L. T. Biegler, 2019, "Nonlinear Optimization of Detailed Heat Exchanger Models with Phase Change," Computer Aided Chemical Engineering, 47, 151–156

S. R. Kazi, "Mathematical Modeling and Optimization of Heat Exchanger Design in Process Flowsheets", PhD Thesis, Carnegie Mellon University, Pittsburgh, PA (2021)

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50113-5

Optimal Design of Extractive Dividing-Wall Column Using an Improved Sequential Least Squares Programming Algorithm

Yingjie Ma[a], Nan Zhang[a], Jie Li[a,*]

[a]*Centre for Process Integration, Department of Chemical Engineering, The University of Manchester, Manchester M13 9PL, UK*
jie.li-2@ manchester.ac.uk

Abstract

Extractive dividing wall column (EDWC) is an efficient and economic technique for separating azeotropic or close boiling-point mixtures. However, optimal design of EDWC using rigorous models is still challenging. In this work, we develop an improved sequential least squares programming (SLSQP)-based feasible path algorithm for such optimal design. The computational results show that the proposed algorithm can generate better solutions and reduce the computational time by around one order of magnitude.

Keywords: EDWC, SLSQP, feasible path algorithm, homotopy continuation

1. Introduction

Extractive dividing wall column (EDWC) is a promising intensification technique for separating azeotropic or close boiling-point mixtures, which can reduce both energy and capital costs significantly compared to the conventional extractive distillation. However, optimal design of EDWC is still challenging due to the complexity of its model with at least ten design and operating variables to be determined.

The sequential sensitivity analysis-based method is frequently used, but the method is hard to consider the interactions among variables, possibly getting a suboptimal solution. Although stochastic algorithms such as genetic algorithm (GA) have also been applied to optimize EDWC (Bravo-Bravo et al., 2010), these algorithms need a large number of simulations, leading to long computational time. The sequential quadratic programming (SQP)-based feasible path algorithm has also been reported for optimisation of EDWC using the sequential modular simulator (Yang et al., 2018). However, the convergence of the SQP algorithm in such environment is largely degraded due to inaccurate gradients used (Pattison and Baldea, 2014). To resolve such problem, Ma et al. (2020) proposed a hybrid steady-state and time-relaxation-based optimization algorithm for optimal design of EDWC in the equation-oriented environment, demonstrating a good convergence. However, the hybrid algorithm requires many pseudo-transient continuation (PTC) simulations (Pattison and Baldea, 2014). As the PTC simulation is usually much more time-consuming than the steady-state simulation, the hybrid algorithm is inefficient.

In this work, we develop an improved sequential least square programming (SLSQP)-based feasible path algorithm for optimal design of EDWC. We integrate the homotopy continuation (HC) technique with the line search to achieve effective and robust line search. The PTC simulation is activated to guarantee the convergence only when many

HC steps are needed. It is shown that the proposed algorithm can generate better solutions with the reduced computational effort by approximately one order of magnitude.

2. Problem Statement

Fig. 1a illustrates a typical EDWC, where the mixture AB is feed stream and S is the solvent. The EDWC can be modelled by the thermodynamically equivalent model with five column-sections shown in Fig. 1b, which is adapted from the model with six column-sections for DWC proposed by Pattison et al. (2016). The problem is stated below:

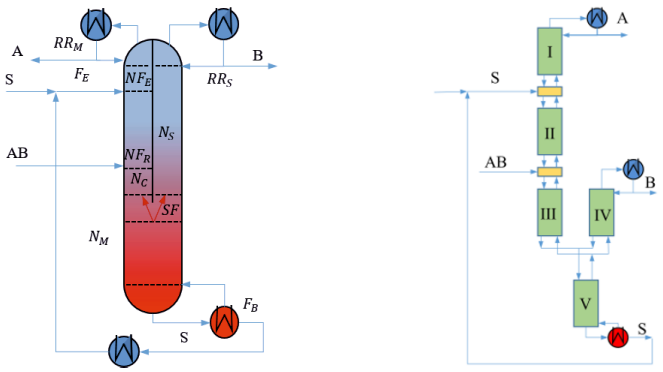


Fig. 1 a) Schematic of an extractive dividing wall column; b) thermodynamically equivalent model with five column-sections

Given: an azeotropic or close boiling-point mixture to be separated, the solvent adopted, the purity requirements of the products, and perhaps some other production requirements (such as the temperatures of some streams). Determine: solvent make-up flow rate (F_S), the number of stages in the five column-sections, reflux ratios of the main and side sections (RR_M, RR_S), split fraction of the vapour stream to the side column section (SF) and the flow rate (F_B) at the bottom. The objective is to minimise total annualised cost.

3. Mathematical Model

The EDWC is described by the rigorous equilibrium stage model with material balance, phase equilibrium, summation and enthalpy balance (MESH) equations applied to each stage. The bypass efficiency method (Dowling and Biegler, 2014) is adopted to determine the number of stages so that the current problem is a more tractable nonlinear programming (NLP) problem. Since we will use PTC simulations as safeguards of the line searches, the PTC model from Ma et al. (2017) is used for such simulations.

4. Solution method

4.1. SLSQP-based feasible path algorithm

Feasible path algorithms separate variables x in the optimisation problem as independent variables x_I and dependent variables x_D. For each given x_I, x_D (and possibly the sensitivity $\partial x_D/\partial x_I$) can be got through solving an nonlinear equation system,

$$F(x_I, x_D) = 0, \tag{1}$$

which is usually called simulation in engineering. Then, various optimisation algorithms can be applied to solve the problem in x_I space instead of the original high dimensional x space. SQP algorithm is often used in such setting as it usually requires less function evaluations (corresponding to simulations) (Schittkowski, 1982) that accounts of nearly all the computational time. As an variant of SQP, SLSQP (Schittkowski, 1982) has the same merit as SQP and is possibly more stable (Schittkowski, 1982), which is to solve a linear constrained linear least squares problem (LSQ) shown below to generate descent direction d^k at each major iteration.

$$\min \frac{1}{2}\|A^k d^k - b^k\|^2 \qquad \text{(LSQ)}$$
$$s.t. E^k d^k - f^k = 0$$
$$G^k d^k - h^k \geq 0.$$

A typical SLSQP-based feasible path algorithm can be briefly stated as follows:

Step 1: set $k \leftarrow 0$. Given x_I^0 and corresponding b^0, E^0, f^0, G^0, h^0, which can be got according to the solution of the simulation problem $F(x_I^0, x_D^0) = 0$.

Step 2: solve LSQ subproblem for d^k and check convergence condition of the original NLP problem. If solution is found, go to step 5.

Step 3: conduct line searches with some simulations to get $x_I^{k+1}$and x_D^{k+1} satisfying the Armijio condition Eq. (2).

Step 4: update A^{k+1}, b^{k+1}, E^{k+1}, f^{k+1}, G^{k+1}, h^{k+1} based on simulation results at x_I^{k+1} and BFGS formula. Set $k \leftarrow k+1$, then go back to Step 2.

Step 5: return x_I^k, x_D^k, f^k.

In step 3, the Armijio condition is as follows

$$P(x_I^{k+1}) \leq P(x_I^k) + \alpha\, \rho\, DP(0), \qquad (2)$$

where $\rho \in (0, 0.5)$ is a constant, $P(x_I^k)$ is the L1 merit function considering the values of both objective function and constraint violations, and $DP(0)$ is the directional derivative of $P(x_I)$ along direction d_k at $\alpha = 0$.

One of the main difficulties in the feasible path algorithm is to conduct the line search reliably in step 3. In Section 4.2, the HC method will be used to resolve the problem.

4.2. Homotopy continuation (HC) enhanced line search

A typical backtrack line search procedure in a major iteration k of above feasible path algorithm is as follows, where we drop the superscript k to avoid the abused use of indices. Hence, x_I^0, x_D^0 in the current Section 4.2 is actually x_I^k, x_D^k in Section 4.1.

Step 3.1: set $i \leftarrow 0$. Given x_I^0, x_D^0, and an initial step length $\alpha^0 \in (0, 1]$.

Step 3.2: set $x_I^{i+1} \leftarrow x_I^0 + \alpha^i d$, conduct simulation $F\left(x_I^{i+1}, x_D^{i+1}\right) = 0$, and evaluate the merit function $P(x_I^{k+1})$ based on the simulation results.

Step 3.3: if Eq. (2) is satisfied, go to Step 3.5.

Step 3.4: generate an α^{i+1} smaller than α^i, set $i \leftarrow i+1$, then go back to Step 3.2.

Step 3.5: returnx_I^{i+1}, x_D^{i+1}.

The line search may fail to generate an α satisfying Eq. (2) or may generate a tiny step length if the simulations continue to fail and $P(x_I^{k+1})$ cannot be evaluated in step 3.2. The former causes the premature termination of the optimisation, while the latter leads to slow progress. The HC method can be used to resolve the problem, which is to start from the solution of an equation system that can be solved more easily and then approach the solution of the original problem. In that spirit, if a simulation problem $F\left(x_I^0 + \alpha^i d, x_D^{i+1}\right) = 0$ cannot be solved in Step 3.2, we define an intermediate step length

$$\bar{\alpha}^j = t^j \alpha^i, t^j \in (0,1], t^0 = 0. \tag{3}$$

We then gradually increase $\bar{\alpha}^j$ from 0 and solve a series of simulation problems $F\left(x_I^0 + \bar{\alpha}^j d, \bar{x}_D^j\right) = 0$ with the solution $(\bar{x}_D^{j'})$ at $\bar{\alpha}^{j'} < \bar{\alpha}^j$ $(j \geq 1)$ as the initial point until $\bar{\alpha}^j = \alpha^i$ $(t^j = 1)$, reaching to the solution of the original problem. In this way, when $\bar{\alpha}^j$ and $\bar{\alpha}^{j'}$ are close enough and the implicit function $x_D(x_I)$ is continuous with nonsingular Jacobian matrix, $\bar{x}_D^{j'}$ will be within the convergence basin of the Newton method for solving the simulation problem, leading to a converged simulation. Note that $\bar{\alpha}^0 = 0$, $\bar{x}_I^0 = x_I^0$ and $\bar{x}_D^0 = x_D^0$. Such process leads to a homotopy path as shown in Fig. 2a.

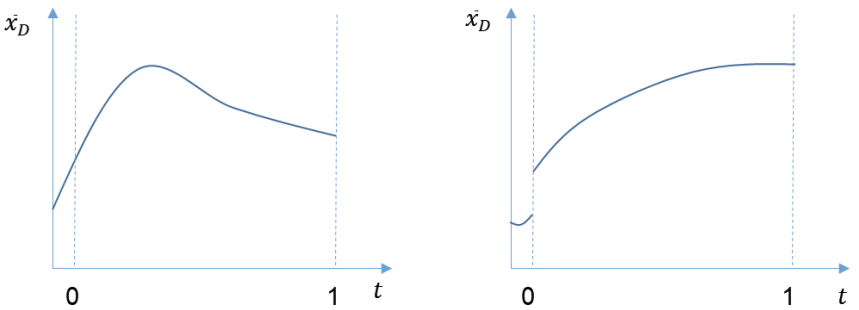


Fig 2. a) homotopy path without singular point, b) homotopy path with discontinuous point.

A merit of applying HC method in line search algorithm is that we have the chance to terminate the HC calculation before t^j reaches to 1 if the merit function $P(x_I^0 + \bar{\alpha}^j d)$ starts to increase. This is because a larger α will lead to a higher merit function under the assumption that $P(x_I^0 + \alpha d)$ has a single minimum along the direction d for $\alpha \in [0,1]$, while the aim of the whole line search algorithm is to find its minimum. The chance to terminate early is an advantage of the proposed HC method to the PTC method, which doesn't provide meaningful intermediate information before the PTC simulation finishes.

However, the HC method cannot guarantee a converged simulation if the assumptions of continuity and nonsingularity are not satisfied as shown in Fig. 2b, which frequently happens during the line search for the optimisation of EDWC. If x_I^0 is at or rather close to the discontinuous point, the HC method will be useless for promoting the convergence of the simulation for the first line search step. In such case, we have to use the PTC method to guarantee the convergence. When the number of HC steps (denoted as n_{HC}) is greater than a certain number (e.g. 10), the PTC method will be then activated. The proposed

simultaneous homotopy continuation and line search method is given below,

Step 3.2.1: set $j \leftarrow 0$. Given x_I^0, x_D^0, $\alpha^i \in (0,1]$, n_{HC}; set $t^j \leftarrow 0.5$, $flag^{LS} \leftarrow False$, $\alpha^{LS} \leftarrow None$, $x_I^{LS} \leftarrow None$, $x_D^{LS} \leftarrow None$, $\alpha^{\min} \leftarrow 0$, $x_I^{\min} \leftarrow$, $P^{\min} \leftarrow \infty, t_c \leftarrow 0$,

Step 3.2.2: generate $\bar{\alpha}^j$ according to Eq. (3), $\bar{x}_I^j \leftarrow x_I^0 + \bar{\alpha}^j d$, conduct simulation $F(\bar{x}_I^j, \bar{x}_D^j) = 0$, and evaluate $P(\bar{x}_I^j)$ based on the simulation results.

Step 3.2.3 if the simulation converges, $t_c \leftarrow t^j$, go to Step 3.2.4; otherwise, go to Step 3.2.6.

Step 3.2.4: if Eq. (2) is satisfied, set $flag^{LS} \leftarrow True$, $\alpha^{LS} \leftarrow \bar{\alpha}^j$, $x_I^{LS} \leftarrow \bar{x}_I^j$, $x_D^{LS} \leftarrow \bar{x}_D^j$.

Step 3.2.5 if $P(\bar{x}_I^j) < P^{\min}$ and $t^j < 1$, set $P^{\min} \leftarrow P(\bar{x}_I^j)$, $\alpha^{\min} \leftarrow \bar{\alpha}^j$, $x_I^{\min} \leftarrow \bar{x}_I^j$, $x_D^{\min} \leftarrow \bar{x}_D^j$ and set $t^{j+1} \in (t_c, 1]$; otherwise, go to Step 3.2.7.

Step 3.2.6: if $j < n_{HC}$, set $t^{j+1} \in (t_c, t^j)$; otherwise, set $t^j \leftarrow 1$, apply the PTC method for simulation, evaluate $P(\bar{x}_I^j)$ and then go to Step 3.2.7;

Step 3.2.7: if $flag^{LS} = True$, return $flag^{LS}$, α^{LS}, x_I^{LS}, x_D^{LS}; otherwise, return $flag^{LS}$, $\alpha^{\min}$, $x_I^{\min}$, $x_D^{\min}$. Go to Step 3.3 of previous algorithm.

Here, $flag^{LS}$ indicates whether a step length satisfying Armijio condition has been found ($True$ if found; otherwise, $False$), and α^{LS}, x_I^{LS} and x_D^{LS} are the corresponding step length, independent and dependent variables. P^{min} denotes the minimum penalty function value obtained until the current iteration, and $\alpha^{\min}$, $x_I^{\min}$ and $x_D^{\min}$ are the corresponding step length, independent and dependent variables. t_c is the largest homotopy parameter leading to a converged simulation until the current iteration.

5. Case study

The separation of acetone and chloroform mixture using EDWC is used to illustrate the capability of the proposed algorithm. The optimal design of EDWC for the same separation has been reported in Ma et al. (2020). The example is solved on a desktop with 3.20 GHz Intel core i7 processor, 16 GB RAM and Windows 64-bit operating system. The optimisation problem involves 9399 equations, 16 inequality constraints and 10570 variables including 90 decision variables. The objective is to minimize total annualized cost (TAC). We conduct optimisations from six different initial points as with Ma et al. (2020). The results are shown in Table 1.

Table 1. Comparative results from the algorithm of Ma et al. (2020) and the proposed algorithm

Initial point		1	2	3	4	5	6
Time/s	Ma2020	1259	7722	2263	3698	1650	1714
	Ours	1498	587	903	380	1197	1519
Number of Simulations	Ma2020	239	302	409	509	380	262
	Ours	478	268	528	229	466	796
TAC/	Ma2020	6081	6116	6106	6124	6137	6097
(k$ year^{-1})	Ours	6076	6076	6076	6078	6094	6077

As seen from Table 1, both algorithms can find optimal solutions from all six initial points. The proposed algorithm needs much less computational time than the hybrid algorithm of Ma et al. (2020) from all initial points except from the initial point 1. For instance, starting from the second initial point, the proposed algorithm needs 587 CPU s, whilst the hybrid algorithm requires 7722 CPU s. More importantly, the proposed algorithm always obtains a bit better solution with TAC reduced by 0.08% ~ 0.75%, compared to the hybrid algorithm. In addition, the proposed algorithm often requires more simulations than the hybrid algorithm, which do not lead to higher computational time for the proposed algorithm because the time-consuming PTC simulations are used much less compared with the hybrid algorithm.

6. Conclusions

In this work we proposed an improved SLSQP-based feasible path optimisation algorithm for optimal design of EDWC using rigorous models. We integrate HC simulations with line searches to get step lengths satisfying Armijio condition reliably. To further improve the efficiency, we proposed a criteria to terminate the HC calculation early. To further guarantee the convergence of simulations, we use PTC simulations as the last safeguard. Finally, one case study from literature indicates the proposed algorithm can be 10 times faster than the state-of-art feasible path algorithm and also generates better solutions.

Acknowledgements

The authors acknowledge financial support from China Scholarship Council – The University of Manchester Joint Scholarship (201809120005)

References

Bravo-Bravo, C., Segovia-Hernández, J.G., Gutiérrez-Antonio, C., Durán, A.L., Bonilla-Petriciolet, A., Briones-Ramírez, A., 2010. Extractive Dividing Wall Column: Design and Optimization. Ind. Eng. Chem. Res. 49, 3672-3688.

Dowling, A.W., Biegler, L.T., 2014. Rigorous Optimization-based Synthesis of Distillation Cascades without Integer Variables, in: Jiří Jaromír Klemeš, P.S.V., Peng Yen, L. (Eds.), Comput. Aided Chem. Eng. Elsevier, pp. 55-60.

Ma, Y., Luo, Y., Yuan, X., 2017. Simultaneous Optimization of Complex Distillation Systems with a New Pseudo-transient Continuation Model. Ind. Eng. Chem. Res. 56, 6266-6274.

Ma, Y., Zhang, N., Li, J., Cao, C., 2020. Optimal design of extractive dividing-wall column using an efficient equation-oriented approach. Front. Chem. Sci. Eng. 15, 72-89.

Pattison, R.C., Baldea, M., 2014. Equation-Oriented Flowsheet Simulation and Optimization Using Pseudo-Transient Models. AIChE J. 60, 4104-4123.

Pattison, R.C., Gupta, A.M., Baldea, M., 2016. Equation-oriented optimization of process flowsheets with dividing-wall columns. AIChE J. 62, 704-716.

Schittkowski, K., 1982. The nonlinear programming method of Wilson, Han, and Powell with an augmented Lagrangian type line search function. Numer. Math. 38, 115-127.

Yang, A., Wei, R., Sun, S., Wei, S.a., Shen, W., Chien, I.L., 2018. Energy-Saving Optimal Design and Effective Control of Heat Integration-Extractive Dividing Wall Column for Separating Heterogeneous Mixture Methanol/Toluene/Water with Multiazeotropes. Ind. Eng. Chem. Res. 57, 8036-8056.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50114-7

Biphasic Dehydration of Sugars to 5-Hydroxymethylfurfural and Furfural—Multiscale Modeling for Easier Optimization and More Accurate Solvent Selection

Abhimanyu Pudi[a,b], Martin P. Andersson[a], Seyed Soheil Mansouri[a*]

[a]*Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søltofts Plads 228A, Kgs. Lyngby, DK-2800, Denmark*
[b] *Sino-Danish College, University of Chinese Academy of Sciences, 19A Yuquan Road, Shijingshan District, Beijing, 100049, China*
seso@kt.dtu.dk

Abstract

An integrated and multiscale modeling framework is introduced to accurately model the biphasic dehydration of a mixed carbohydrate feed from typical lignocellulosic waste biomass to form 5-hydroxymethylfurfural and furfural. This modeling framework integrates computational chemistry into a process model, allowing for a greater exploration space for process design. Moreover, this multiscale model is used to demonstrate more accurate solvent selection that is in line experimental data, unlike the existing solvent screening methods. For this purpose, a pool of five commonly used organic solvents for this system are considered, which are 1-butanol, 2-butanol, methyl isobutyl ketone, 2-methyltetrahydrofuran, and tetrahydrofuran.

Keywords: green chemistry, process intensification, computational chemistry, mathematical modeling, lignocellulosic biomass valorization.

1. Introduction

The current global scenario for energy and chemicals consumption features the impending exhaustion of fossil resources and the undeniable threat of global warming as major challenges to be tackled by mankind. Due to the escalating demands for energy, bulk chemicals and materials, alternative sources of feedstock are restlessly sought after. In the recent decades, the biorefinery concept has emerged as an alternative for the generation of these goods via the sustainable processing of biomass of diverse nature following chemical, thermochemical, enzymatic or fermentative pathways (Corma et al., 2007). Conversion of biomass into fuels and chemicals bids a potential opportunity to fulfill the energy needs of the upcoming decades. Owing to their relevance in synthesis, 5-hydroxymethylfurfural (HMF) and furfural have been considered as outstanding building blocks for chemicals and fuels in the US Department of Energy's list of top value-added chemicals from biomass (Bozell and Petersen, 2010). Both these compounds offer great possibilities considering their chemical functionality and allow the production of a wide array of chemicals with very diverse applications. Putten et al. (2013) reviewed thoroughly the synthetic pathways starting from HMF leading to products with applications as: monomers for subsequent polymerization, highlighting diols from HMF, 2,5-diformylfuran, 2,5-furandicarboxylic acid or 5-hydroxymethyl-2-furan carboxylic acid; fine chemicals, including products of

interest as pharmaceuticals, agrochemicals, flavors and fragrances; and fuel components, such as dimethylfuran, levulinic acid or methyl tetrahydrofuran.

Much of the earlier work on the synthesis of these building block chemicals is reported in monophasic systems which suffer from thermodynamic limitations, low selectivity, low yield, or undesired side-products. A smart strategy to overcome some, if not all, of these hurdles is the use of a multiphase reaction-extraction system. However, the design and optimization of such systems is hardly straightforward. For instance, the choice of the organic solvent in the aqueous biphasic system to produce HMF and furfural from sugars is a crucial factor for both process economics and sustainability. Mathematical modeling can be valuable for efficiently analyzing and designing these complex systems. For example, solution and reaction properties of the many chemicals involved (reactants, solvents, products, coproducts, and catalysts) need to be described; the extent of miscibility (totally, partially, or effectively immiscible) must be established; the phases where reactions occur need to be identified; and the reaction and mass transfer mechanisms must be established. Also, the effects of chemically inert species on partitioning and of mixture composition on reaction rates must be characterized. However, the commonly used thermodynamic models lack the necessary thermodynamic parameters for every case and are inherently limited to the portion of the chemical design space for which every binary interaction parameter is available.

In this respect, COSMO-based models, such as COSMO-RS (Klamt, 1995; Klamt et al., 1998), are valuable alternatives for describing liquid-phase thermodynamics since they do not require any binary interaction parameters. Furthermore, such models allows for easy integration of quantum chemical calculations into a process modeling framework, greatly expanding the envelope of chemical species that can be modeled at a high level of accuracy. Here, an integrated and multiscale modeling architecture (Pudi et al., 2020) is employed to design and simulate the biphasic process based on COSMO-based thermodynamic models that do not require any binary interaction parameters. This modeling approach allows for not only easier optimization of reaction conditions but also more accurate solvent selection compared to the existing screening methods.

2. Multiscale modeling approach

This multiscale approach was first presented in an earlier work (Pudi et al., 2020). With python as the high-level interface, three different tools are employed at three different scales: density functional theory for the description of individual molecules, COSMO-RS for the description of individual physicochemical phenomena, and mathematical modeling for the description of all the interlinked phenomena in a biphasic reactive system.

3. Biphasic system

In an effort to achieve carbon neutrality and circular economy, lignocellulosic waste biomass has been identified as a promising carbon source and widely studied for the production of HMF and furfural. Therefore, a typical composition of rice straw (20% xylose, 35% fructose, and 45% lignin and other inert materials) is considered as feed (Amiri et al., 2010). Three reactions are considered in the system: xylose dehydration to furfural, fructose dehydration to HMF, and HMF rehydration to levulinic acid (LA) and formic acid (FA).

$$\text{xylose} \rightarrow \text{furfural} + 3H_2O \quad (1)$$

$$\text{fructose} \rightarrow \text{HMF} + 3H_2O \quad (2)$$

$$\text{HMF} + 2H_2O \rightarrow \text{LA} + \text{FA} \quad (3)$$

Other degradation reactions, including humin formation, have not been considered in this work. The calculations are carried out for a typical reaction temperature of 150 °C and 1:1 volume ratio of aqueous and organic phases. Five commonly used organic solvents for this process are considered to evaluate the impact of this choice on system performance and compare the accuracy of the multiscale model with experimental data and other solvent screening works. The candidates are 1-butanol, 2-butanol, methyl isobutyl ketone (MIBK), 2-methyltetrahydrofuran (2-MeTHF), and tetrahydrofuran (THF).

4. Results and discussion

4.1. Steady state properties of the system

Results from the multiscale process model are presented in Table 1. At equilibrium, the conversions of xylose and fructose are essentially 100 %. Since only reaction is considered for xylose, all of the converted reactant is converted to furfural. In the case of fructose, it appears that it is completely converted to LA and FA, leaving no presence of HMF. These results are unsurprising once the standard Gibbs' free energies of these reactions in the aqueous phase are considered: -164 kJ/mol for xylose dehydration to furfural, -182 kJ/mol for fructose dehydration to HMF, and -121 kJ/mol for HMF rehydration to LA and FA. These values remain do not decrease by more than 15–30 kJ/mol as the reaction proceeds from the feed composition to equilibrium. This explains the complete conversion of the feed reactants and HMF. Feed sugar conversions of >99% at equilibrium are also observed in experiments (Yang et al., 2012).

Table 1 Equilibrium results (in percentages) at 150 °C calculated using the multiscale model

Solvent	Furfural Yield	Furfural Extracted	HMF Yield	LA Yield	LA Extracted	FA Extracted
1-butanol	100	87	0	100	77	60
2-butanol	100	86	0	100	74	56
MIBK	100	81	0	100	58	40
2-MeTHF	100	83	0	100	62	50
THF	100	88	0	100	77	68

The biphasic system has minimal effect on the fundamental kinetics in the aqueous phase (Weingarten et al., 2010), but it is crucial to maximize product yield by extracting the desired products into the organic phase. Although the choice of the solvent also depends on other process objectives, this work only compares their extractive abilities. All the considered solvents extract more than 80% of the produced furfural. However, contrary to the results of other solvent screening works (Esteban et al., 2020), our

results show that 1-butanol and 2-butanol extract higher percentage of furfural than MIBK and 2-MeTHF. Moreover, our results agree with experimental studies (Amiri et al., 2010). The strength of our multiscale model in comparison to other solvent screening works lies in the estimation of a solvent's extractive ability.

4.2. Extractive ability of the solvent candidates

Extractive ability of the organic solvent is crucial for a well-designed biphasic system that can capture as much of the desired products as possible. Table 2 presents a comparison between the partition coefficients of all the products in five different solvents at feed conditions. These coefficients (P_{OA}) are calculated as the ratio of mole fraction of a component in the organic phase to that in the aqueous phase. Earlier studies in solvent selection chose to calculate partition or distribution coefficients at room temperature and/or in a purely ternary solvent-water-solute system (Blumenthal et al., 2016; Esteban et al., 2020). However, partition coefficients could vary by significant margins at higher temperatures and in the presence of other compounds. In general, high partition coefficients are desired both at higher temperatures during the reaction (to maximize yield and selectivity) and at lower temperatures after the reaction (to maximize the amount of product in the extraction phase).

Although the values reported in this work cannot be quantitatively compared to the concentration-based partition coefficients usually reported in experimental studies, the general trends can be compared. Most of the commonly used organic solvents for this application provide a higher P_{OA} for furfural than HMF (Esteban et al., 2020), and this trend is also seen in the values reported in this work. However, comparing the partition coefficients in Table 2 to the amounts of products extracted in Table 1, there is no direct correlation. For example, MIBK has the highest P_{OA} for furfural but recovers the least amount of furfural.

Herein lies the reason for our multiscale model's better prediction of the real behavior observed in experiments. While all the solvent screening work published in the literature so far have based their results on partition coefficients at one particular state of the biphasic system (or worse, in pure ternary systems), our multiscale model takes into account the changing compositions. If necessary, it can also consider the change in temperatures if the reaction and settling are performed at different temperatures.

Table 2 Partition coefficients of all the products at feed composition and 150 °C

Solvent	HMF	Furfural	Levulinic Acid	Formic Acid
1-butanol	8.1	10.9	5.5	1.6
2-butanol	8.4	11.5	5.7	1.6
MIBK	11.2	19.4	6.7	1.8
2-MeTHF	10.2	15.6	6.0	1.9
THF	7.7	9.5	5.0	1.9

In addition to the partition coefficients, it is important to consider the degree of miscibility between the solvent and water. Most studies neglect the amount of water dissolved in the organic phase in their solvent selection process. However, the mole fraction of water partitioning into the organic phase is around 30–50% for all the

commonly used organic solvents for this process. The solubility of one solvent in the other creates downstream separation and recycling challenges that should be considered in the overall process economics and sustainability analyses. A recent study has identified several alternatives in this aspect (Wang et al., 2021).

5. Current limitations

As noted earlier, only three reactions are considered in this work and only one of them is a degradation reaction. For example, no degradation reactions were considered for furfural, which caused 100% furfural yield. However, soluble and insoluble humins are known to form in this system, which reduces the yield of furfural and HMF. In addition, this process in practice is rarely taken to equilibrium in order to maximize the yield of desired products (Weingarten et al., 2010). Therefore, a more comprehensive study must include kinetics and assess the dynamic behavior of the system.

6. Conclusions

A multiscale modeling approach based on quantum chemistry and COSMO-RS is successfully employed in this work to model the steady state behavior of biphasic dehydration of sugars to valuable products such as HMF and furfural. This approach conforms to prior knowledge and experimental data. In addition, solvent selection based on this multiscale model presents more accurate results than the existing screening approaches that sometimes contradict experimental behavior. The presented approach paves the way for a more comprehensive work that can be used for systematic process optimization.

References

Amiri, H., Karimi, K., Roodpeyma, S., 2010. Production of furans from rice straw by single-phase and biphasic systems. Carbohydr. Res. 345, 2133–2138.

Blumenthal, L.C., Jens, C.M., Ulbrich, J., Schwering, F., Langrehr, V., Turek, T., Kunz, U., Leonhard, K., Palkovits, R., 2016. Systematic Identification of Solvents Optimal for the Extraction of 5-Hydroxymethylfurfural from Aqueous Reactive Solutions. ACS Sustain. Chem. Eng. 4, 228–235.

Bozell, J.J., Petersen, G.R., 2010. Technology development for the production of biobased products from biorefinery carbohydrates—the US Department of Energy's "Top 10" revisited. Green Chem. 12, 539–554.

Corma, A., Iborra, S., Velty, A., 2007. Chemical Routes for the Transformation of Biomass into Chemicals. Chem. Rev. 107, 2411–2502.

Esteban, J., Vorholt, A.J., Leitner, W., 2020. An overview of the biphasic dehydration of sugars to 5-hydroxymethylfurfural and furfural: A rational selection of solvents using COSMO-RS and selection guides. Green Chem. 22, 2097–2128.

Klamt, A., 1995. Conductor-like screening model for real solvents: A new approach to the quantitative calculation of solvation phenomena. J. Phys. Chem. 99, 2224–2235.

Klamt, A., Jonas, V., Bürger, T., Lohrenz, J.C.W., 1998. Refinement and parametrization of COSMO-RS. J. Phys. Chem. A 102, 5074–5085.

Pudi, A., Karcz, A.P., Shadravan, V., Andersson, M.P., Mansouri, S.S., 2020. Modeling of Liquid-Liquid Phase Transfer Catalysis: Process Intensification via Integration of Process Systems Engineering and Computational Chemistry, in: Pierucci, S., Manenti, F., Bozzano, G.L., Manca, D. (Eds.), Proceedings of the 30th European Symposium on Computer Aided Process Engineering (ESCAPE30). Elsevier B.V., pp. 43–48.

Putten, R.-J. van, Waal, J.C. van der, Jong, E. de, Rasrendra, C.B., Heeres, H.J., Vries, J.G. de, 2013. Hydroxymethylfurfural, A Versatile Platform Chemical Made from Renewable Resources. Chem. Rev. 113, 1499–1597.

Wang, Z., Bhattacharyya, S., Vlachos, D.G., 2021. Extraction of Furfural and Furfural/5-Hydroxymethylfurfural from Mixed Lignocellulosic Biomass-Derived Feedstocks. ACS Sustain. Chem. Eng. 9, 7489–7498.

Weingarten, R., Cho, J., Conner, W.C., Huber, G.W., 2010. Kinetics of furfural production by dehydration of xylose in a biphasic reactor with microwave heating. Green Chem. 12, 1423–1429.

Yang, Y., Hu, C.W., Abu-Omar, M.M., 2012. Synthesis of furfural from xylose, xylan, and biomass using AlCl 3·6H2O in biphasic media via xylose isomerization to xylulose. ChemSusChem 5, 405–410.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50115-9

Study of Mass Transfer Coefficient of CO_2 Capture in different Solvents using Microchannel: A Comparative Study

Bushra Khatoon, Shabih-Ul-Hasan, and M. Siraj Alam*

Department of Chemical Engineering, MNNIT Allahabad, Prayagraj-21104, India
**msalam@mnnit.ac.in*

Abstract

Process intensification with micro channels is known for the lowest production cost, high efficiency, safest-clean production rate, and energy-saving equipment. It is well known that by replacing conventional channels with micro-level channels, CO_2 emission can be controlled efficiently with the comparatively high CO_2 removal rate. In the present study, we compare the CO_2 absorption rate in presence of different solvents like water, amino acids, mixture of amines, ammonia, NaOH, and KOH in terms of mass transfer coefficient that is based on different absorption phenomena and come-up with some guidelines. Also, an attempt is made to develop a correlation for the Sherwood number and the results obtained are compared with the predicted available co-relations in literature. This study concludes that the use of microchannels can enhance the mass transfer coefficient as well as CO_2 absorption rate several times in comparison to conventional channels and amines are proved to be a better solvent in comparison to other solvents for CO_2 removal in microchannels.

Keywords: Absorption, Mass transfer coefficient, Amines, CO_2, Microchannels.

1. Introduction

In this era, industrial growth is on boom due to rapid advancement and implementation of new techniques at the same time its growth is hampered significantly due to CO_2 emissions. Absorption of CO_2 in solvents like water, amino acids (MEA, DEA, and MDEA), mixture of amines (Mackowiak et al., 2018), ammonia, NaOH, and KOH is promising both by physically and chemically in micro-channels. Akkarawatkhoosith et al., (2020) shows that the chemisorption rate is lowest in conventional channels, and its value increases as the size of channel diameter decreases. It is well known that physisorption of CO_2 by water at low temperature and elevated pressure is comparatively low w.r.t chemisorption. However, the absorption rate can be enhanced by replacing the solvents, using some additive in existing solvents, varying the operating conditions such as flow rates (Q), concentration (C), etc., replacing physical absorption phenomena with chemical absorption, and by reducing the channel diameter to mini and micro level. As a result, we discuss the need of microchannels over conventional and mini contactors/reactors on the basis of mass transfer coefficients values in the next section.

2. Selection of system

2.1 Conventional, Mini, and Micro-channels

On the basis of mass transfer coefficients values we justify the need of microchannels over conventional channels/reactors. Figure 1 (a & b) shows a comparison of the values of mass transfer coefficient for conventional channels (e.g. packed columns, tray columns, trickle bed reactors), and microchannels (range: 200 μm ≧ D_h ＞10 μm). Chemical absorption involves mass transfer rate and mass transfer coefficient can be enhanced by the parameters such as flow rates (Q), concentration (C), temperature (T), pressure (P) etc. (Al-Hindi et al., 2018). It can be easily seen in Figure 1 that the microchannels and a special type of microchannel gives a higher value of mass transfer coefficient than the conventional and mini-channels for the absorption process.

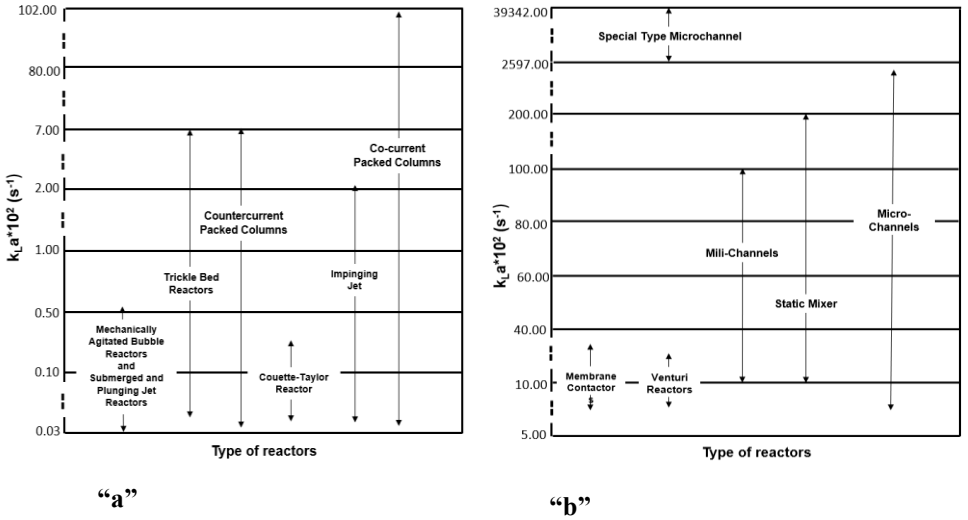


Figure 1: Schematic representation of mass transfer coefficient for different contactors and reactors.

2.2 Physical/Chemical Absorption

Dong et al., (2020) shows that the chemisorption of CO_2 has an absorption rate around 3 to 10 times greater than the physical absorption. As the diameter of the channels decreases (micro-channels) chemisorption always gives higher values of mass transfer coefficient over physical absorption (Akkarawatkhoosith et al., 2020). For microreactors, the range of mass transfer coefficient for physical absorption comes under the 0.3 - 7 s-1 and for chemisorption, the range extended from 0.7 - 100 s-1 for CO_2 absorption in amines (Yao et al., 2017).

2.3 Solvents

The solvents shown in Table 1 gives good absorption rate for conventional channels. These solvents can also give good absorption rate in mini and micro-channels as well provided channels should be clean before vaporization of solvents. Because absorption of CO_2 in water, NaOH, KOH, and ammonia forms soluble salts (Carbonates and Bicarbonates) may block the channels of small diameters. Thus, the amines and its

blended solutions can be proved to be good solvents for the CO_2 absorption process in microchannels and a special type of microchannel (Figure 1).

Table 1: CO_2 absorption in different type of solvents reported in literature.

Authors	Absorption	Absorption mechanism
Elhajj et al., (2020)	Water	$CO_2 + H_2O \rightleftharpoons H_2CO_3; H_2CO_3 \rightleftharpoons HCO_3^- + H^+;$ $HCO_3^- \rightleftharpoons CO_3^{2-} + H^+$
Kraub et al., (2017)	NaOH	$CO_2 + 2OH^- \rightleftharpoons CO_3^{2-} + H_2O$
Liu et al., (2009)	Ammonia	$NH_3 + CO_2 + H_2O \rightleftharpoons NH_4HCO_3$
Kim et al., (2014)	MEA	$CO_2 + 2MEA \rightleftharpoons MEAH^+ + MEACOO^-$
Rinker et al., (1996)	DEA	$CO_2 + R_1R_2NH \rightleftharpoons R_1R_2NH^+COO^-$ $R_1R_2NH^+COO^- + B \rightleftharpoons R_1R_2NCOO^- + BH^+$
Donaldson et al. (1980)	MDEA	$(R)_3N + CO_2 + H_2O \rightleftharpoons (R)_3NH^+ + HCO_3^-$ Where $(R)_3N$ is any tertiary amine e.g., MDEA
Conway et al. (2015)	Mixture of Amines	$Amine_1 + Amine_2 + CO_2 \rightleftharpoons Amine_1CO_3^- + Amine_2H^+$

2.4 Concentration

When CO_2 absorbs in any solvent, several parameters affect the absorption rate like concentration, flow rates, temperature, pressure, channel diameter, type of absorption, and solvents. Section 2.1, 2.2, and 2.3 helps to decide the channel diameter range, type of absorption (physical/chemical absorption), and solvents. Table 2 shows low temperature and high pressure is a favorable condition for CO_2 loading (means mole of CO_2 load per mole of solvent). Table 3 indicates that higher absorption efficiency can be achieved with higher solvent rate and low value of CO_2 flow rate.

Table 2: CO_2 absorption in amines at different temperature and pressure.

Amines	Temperature (K)	Pressure (kPa)	CO_2 loading (mole/mole)	Authors
MEA	313	15.70	0.56	Prachi Singh (2011)
MEA	373	30.40	0.238	Guevara et al., (1993)
DEA	298	6.89	0.57	Lee et al., (1972)
DEA	313	10.70	0.59	Benamor et al., (2005)
MDEA	313	101.325	0.46-0.58	Chowdhary et al., (2013)

Table 3: CO_2 absorption in amines at different flow rates.

Amines	Liquid load (m^3/m^2-h)	Gas flow rate (L/h)	Removal Efficiency (%)	Authors
MEA, DEA, MDEA	10	-	90, 54, 04	Aroonwilas et al., (2004)
MEA, DEA, MDEA	4.8	-	43, 35, 03	Aroonwilas et al., (2004)
MDEA	-	100, 200, 300, 400	< 40, ≈ 20, < 10, < 05	Pan et al., (2014)

2.5 Sherwood number

From section (2.1-2.4) that chemisorption in microchannels enhance mass transfer coefficient/absorption rate by using high flow rate of solvent (amines), low flow rate of

CO_2, low temperature, and high pressure. We represent mass transfer coefficient in terms of Sherwood number (Sh) for gas liquid-phase and use the method of Buckingham Pi theorem in which Sherwood number is a function of gaseous Reynolds number (Re_G), Reynolds number of liquid (Re_L), Schmidt number for gas (Sc_G), and the ratio of liquid to gas velocity (V_L/V_G) as given by Eq. (1 & 2).

$$Sh = f.(Re_G.Re_L.Sc_G.\frac{V_L}{V_G}) \quad (1)$$

$$Sh = n_1.Re_G^{n_2}.Re_L^{n_3}.Sc_G^{n_4}.\frac{V_L}{V_G}^{n_5} \quad (2)$$

In order to obtain a more accurate Sherwood number, the Schmidt number of gas is introduced in Eq. (2) in which n_1, n_2, n_3, n_4, and n_5 are the fitting parameters and their values taken are 0.084, 0.12, 0.385, 0.3, and 1 respectively. These values are obtained with the help of experimental data, performed in the laboratory and the parameters used in the experiment are given in Table 4. The final modified form of the equation is given by Eq. (3).

$$Sh = 0.084\,.Re_G^{0.12}Re_L^{0.385}Sc_G^{0.3}\frac{V_L}{V_G} \quad (3)$$

Table 4: Parameters used for the present study of experimental and theoretical work.

Parameters	**Values**	**Parameters**	**Values**	**Parameters**	**Values**
Gas	CO_2	Temperature	298, 313, and 338 K	Schmidt number	5891.684, 4359.327, and 3969.112
Amine	Aq. MEA	Model	Homogeneous mixture model	Reynolds number	$13 < Re_G < 193$ $90 < Re_L < 540$
Pressure	1 atm.	V_L/V_G	10	System	Microchannel

3. Results and discussion

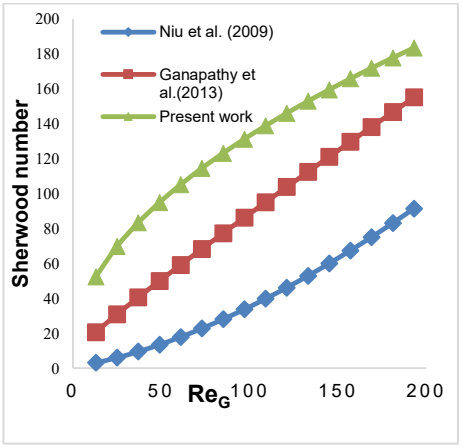


Figure 2: Comparison of present work with the other empirical co-relations.

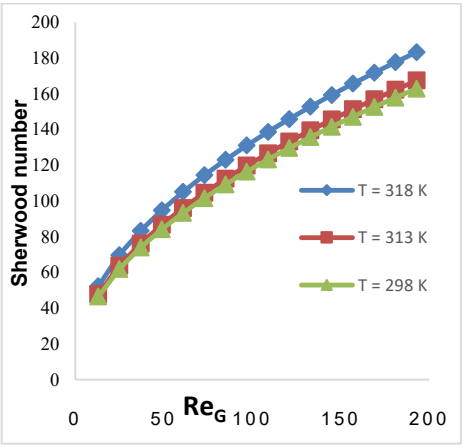


Figure 3: Study of Sherwood number at a different temperature, for CO_2 absorption in aq. MEA.

Figure 2 and Figure 3 shows the comparison of modified correlation given by Eq. (3) with that of Niu et al., (2009) and Ganapathy et al., (2013) which is mentioned in Eq. (4) and (5) respectively.

$$Sh.a.d_h = 0.11 Re_G^{0.39} Re_L^{0.7} Sc_L^{0.5} \quad (4)$$

$$Sh.a.d_h = 10.201 Re_G^{0.206} Re_L^{0.218} Sc_L^{0.5} \quad (5)$$

It can be clearly seen that the modified correlation outperforms in comparison to the others and gives higher values of Sherwood number throughout the range of Re_G (0-200) (Figure 2). It is also observed that replacing Sc_L with Sc_G gives the higher value of Sherwood number and thereby higher value of mass transfer coefficient as well. Figure 3 indicates that with increase in temperature, the value of the Sherwood number increases as a result the mass transfer coefficient also increases at higher temperature values.

4. Conclusions

The presented work investigates the conditions to enhance the higher rate of CO_2 removal and solvent selection, and equipment's based on literature review and also suggest a modified correlation of Sherwood number that gives higher values of mass transfer coefficient. The main findings are as follows:

- In conventional, mini and microchannel, it is found that microchannels/special type of microchannels provides a high value of mass transfer coefficient.
- It is found that chemisorption with amine solution gives high rate of absorption with low flow rate of CO_2, high flow rate of solvents, low temperature and high pressure. Amines are proved to be good solvent for microchannels because it doesn't form carbonates and bi-carbonates and thereby prevent choking problems in channels.
- On that basis of developed empirical co-relation, it can say that that the selection of Sc_G over Sc_L provides much better results than the other empirical co-relations which are also verified with the help of experimental work.
- At constant pressure, as the temperature increases the value of Sherwood number increases which leads to increase in the value of mass transfer coefficient. Thus, it can be concluded that the 318 K favours the chemisorption process of CO_2 in MEA.

References

Akkarawatkhoosith N., Nopcharoenkul W., Kaewchada A., Jaree A., 2020, Mass transfer correlation and optimization of carbon dioxide capture in a microchannel contactor: A case of CO_2 rich gas. Energies, 13, 5465.

Al-Hindi M., Azizi F., 2018, Absorption and desorption of carbon dioxide in several water types. The Canadian Journal of Chemical Engineering, 96, 274-284. Industrial and Engineering Chemistry Research, 43, 2228-2237.

Aroonwilas A., Veawab A., 2004, Characterization and comparison of the CO_2 absorption performance into single and blended alkanolamines in a packed column.

Benamor A., Aroua M. K., 2005, Modeling of CO_2 solubility and carbamate concentration in DEA, MDEA and their mixtures using the Deshmukh–Mather model. Fluid Phase Equilibria, 231, 150-162.

Chowdhary F. A., Yamada H., Higashi T., Matsuzaki Y., Kazama S., 2013, Synthesis and characterization of new absorbents for CO_2 capture. Energy Procedia, 37, 265-272.

Conway W., Bruggink S., Beyad Y., Luo W., Cabrera M. I., Puxty G., Feron P., (2015), CO_2 absorption into aqueous amine blended solutions containing monoethanolamine (MEA), N, N-dimethylethanolamine (DMEA), N, N-diethanolamine (DEEA) and 2-amino-2-methyl-1-propanol (AMP) for post-combustion capture process, Chemical Engineering Science, 126, 446-454.

Dong R., Chu Di., Sun Q., Jin Z., 2020, Numerical simulation of the mass transfer process of CO2 absorption by different solutions in a microchannel. The Canadian Journal of Chemical Engineering, 98, 2648-2664.

Donaldson T. L., Nguyen Y. N., 1980, Carbon dioxide reaction kinetics and transport in aqueous amine membranes. Ind. Engineering Chemical Fundamental, 19, 260-266.

Elhajj J., Al-Hindi M., Azizi M., 2020, A review of the absorption and desorption process of carbon dioxide in water systems. Industrial and Engineering Chemistry Research, 53, 2-22.

Guevara F. M., Libreros E. R., Trejo A., 1993, Gas solubility of carbon dioxide and hydrogen sulfide in mixtures of sulfolane with monoethanolamine. Fluid Phase Equilibria, 86, 225-231.

Ganapathy H., Al-Hajri E., Ohadi M., 2013, Mass transfer characteristics of gas-liquid absorption during Taylor flow in mini/microchannel reactors. Chemical Engineering Science, 101, 69–80.

Kim I., Hoff K. A., Mejdell T., Heat of absorption of CO_2 with aqueous solutions of MEA: new experimental data. Energy Procedia, 63, 1446-1455.

Kraub M., Rzehak R., 2017, Reactive absorption of CO_2 in NaOH: Detailed study of enhancement factor models. Chemical Engineering Sciences, 166, 193-209.

Lee, J. I., Otto, F. D., Mather, A. E., 1972, Solubility of carbon dioxide in aqueous diethanolamine solutions at high pressures. Journal of Chemical and Engineering Data, 17, 465-468.

Liu J., Wang S., Zhao B., Tong H., Chen C., 2009, Absorption of carbon dioxide in aqueous ammonia, Energy Procedia, 1, 933-940.

Mackowiak J. F., Syring K., Thomas A., Leimbrink M., Skiborowski M., Gorak A., Mackowiak J., 2018, Absorption of carbon dioxide using enzyme activated amine solution in columns with random packings. Chemical Engineering Transactions, 69, 2283-9216.

Niu H., Pan L., Wang S., 2009, Flow pattern and gas-liquid interfacial area in a microchannel contactor. Modern Chemical Industry, 29, 60-64.

Pan M. Y., Qian Z., Shao L., Arowo M., Chen J. F., Wang J. X., 2014, Absorption of carbon dioxide into N-methyldiethanolamine in a high-throughput microchannel reactor. Separation and Purification Technology, 125, 52-58.

Rinker E. B., Ashour S. S., Sandall O. C., 1996, Kinetics and Modeling of Carbon Dioxide Absorption into Aqueous Solutions of Diethanolamine. Industrial & Engineering Chemistry Research, 35, 1107-1114.

Singh P., Amine based solvent for CO_2 absorption form molecular structure process, Thesis. University of Twente, The Netherlands, 2011.

Yao C., Zhu K., Liu Y., Liu H., Jiao F., Chen G., 2017, Intensified CO_2 absorption in a microchannel reactor under elevated pressures. Chemical Engineering Journal, 319, 179-190.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50116-0

Techno-Economic Study of Intensified Ethylene Oxide Production Using High Thermal Conductivity Microfibrous Entrapped Catalyst

Chinmoy B. Mukta[a], Nikhil R. Rayaprolu[a], Selen Cremaschi[a], Mario R. Eden[a*] and Bruce J. Tatarchuk[a,b]

[a]*Department of Chemical Engineering, Auburn University, Auburn, AL 36849, USA*
[b]*IntraMicron Inc.. 368 Industry Drive, Auburn AL 36832, USA*
edenmar@auburn.edu

Abstract

Ethylene oxide (EO), a high-volume chemical intermediate, is produced through highly exothermic partial oxidation reactions. It is one of the most energy-intensive and inefficient processes in the chemical process industry. Even a small increase in efficiency through process intensification can significantly reduce the harmful impact on the environment and improve economic performance. The exothermic partial oxidation reaction necessitates the use of a large number of long, small-diameter tubes inside the EO reactor to achieve fast convective heat removal. Moreover, localized hotspots can originate from flow channelling inside the reactor, causing some tubes to operate under runaway conditions, producing unwanted, complete oxidation of ethylene in the ignited tubes, resulting in much lower overall selectivity. The newly developed MicroFibrous Entrapped Catalysts (MFECs) is a non-woven microfibrous metal mesh made of either nickel, steel, or copper. MFECs provide better thermal management of the exothermic reactions through enhanced conductive instead of convective heat transfer inside the reactors. This work assesses the impact of using MFECs inside the EO reactor on the overall process. We have evaluated the effect of the number of ignited tubes (10%, 4%, 2%, 0%) inside the reactor for the conventional process and compared it to the MFEC configuration, which inherently avoids ignition. Avoiding ignited tubes results in higher EO production and reduced formation of CO_2, leading to reduced separation cost. The MFEC configuration leads to an overall product cost reduction of 18.4%.

Keywords: Intensification; Process Simulation; Derivative-free Optimization.

1. Introduction

The ever-growing use of commodity products is causing great demand for high-volume raw material intermediates such as ethylene oxides, which are extensively used in plastic bottles, anti-freeze, sports gear, detergents, and paints. In 2019 alone, the United States produced 2.8 million metric tons of ethylene oxides (EO) and is projected to increase 3-4% per year over the next decade. Ethylene oxidation is considered to have great potential for reducing of carbon emissions (Brueske *et al*., 2015).

The newly developed MicroFibrous Entrapped Catalysts (MFECs) is a non-woven microfibrous metal mesh made of either Nickel, Steel, or Copper. MFECs provide better thermal management of the exothermic reactions through enhanced conductive instead of convective heat transfer inside the reactors. It also has a better ignition prevention capacity because of its ability to mitigate hotspots. The technology has been

experimentally proven for a variety of exothermic reaction systems, including Fisher-Tropsch Synthesis (Choudhury *et al*., 2020). Moreover, MFECs having a large surface area and high void space can help in even flow distribution along the reactor. All these properties of MFECs are advantageous in EO production, especially in terms of increased per pass ethylene conversion, prevention of hotspots, enhanced process safety, stable performance, and extended catalyst activity (Sheng *et al*., 2013).

This work assesses the impact of using MFECs inside the EO reactor on the overall process. First, a plant-scale baseline simulation model of the conventional EO process has been developed using Aspen Plus^{TM}. The optimum design parameters are found using derivative-free optimization (DFO). Next, an economic analysis tool, ECON (Kalakul *et al*. 2014), is used to calculate capital and operating costs. We evaluated the effect of the number of ignited tubes (10%, 4%, 2%, 0%) inside the reactor for the conventional process and compared it to the MFEC configuration, which inherently avoids ignition.

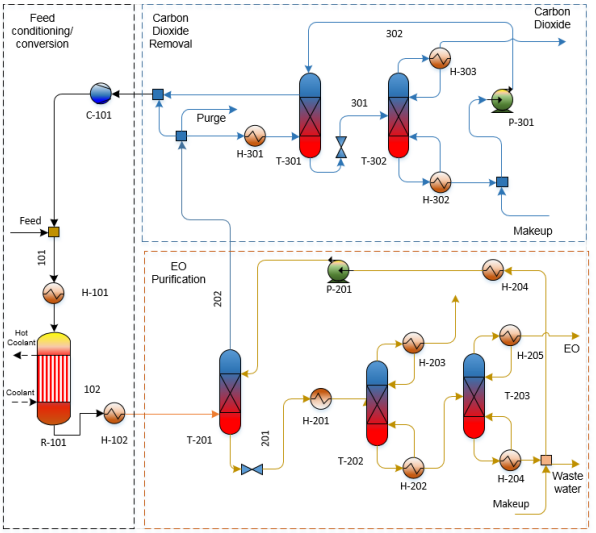


Figure 1. Ethylene oxide process PFD

2. Process Description and Design

2.1 Process description

The conventional oxygen-based process, which is reported to offer higher efficiency towards EO formation than the air-based process variant, was selected as the base-case design for later process intensification (Barecka *et al*., 2017; Peschel *et al*., 2012). Based on reported EO capacities, the typical product flowrate was selected as 100 kilotons per year with a purity of 99.1%. The reactor and EO feed conditioning and conversion sections are modeled using the UNIFAC group contribution method. Moreover, the model applicability was verified by comparing simulation results with process plant data. For the EO absorption and purification sections, SR-POLAR and experimentally regressed CPA models are used, respectively (Barecka *et al*., 2017). The carbon dioxide separation sections are modeled by the Electrolyte Non-Random-Two-Liquid (ENRTL) model with the Redlich-Kwong (RK) equation for the vapor phase. Henry coefficients are used for modeling the solubility of gases.

2.2. Feed conditioning and conversion

The EO reactor (R-101) is modeled as a multitubular, packed-plug flow reactor. A silver-based catalyst was chosen based on its reported high selectivity and work rate (Kobe *et al*., 2002). Typical reaction conditions for industrial EO production uses excess ethylene with 8% conversion of ethylene and 2% EO in the reactor outlet.

2.3. EO Purification

EO is absorbed in water under high pressure (20 bars) in a counter-current column (T-201). The water flow rate is adjusted to achieve total recovery of the diluted EO. The gas stream leaving the absorber, which is depleted in EO, is split into three streams: a fraction of the stream is sent to the CO_2 removal section, whereas the rest is directly recycled to the reactor to limit separation costs. A small fraction (<1%) of the gas stream is purged to avoid build-up of impurities. In column T-202, the dissolved gases are vented. Finally, in column T-203, EO is desorbed under lower pressure, reaching a purity of 99%.

2.4. Carbon dioxide removal

Carbon dioxide is scrubbed by physical and chemical absorption in hot potassium carbonate (K_2CO_3) solution (30% in water); MEAs are usually not used as they might form stable EO complexes. The reaction kinetics reported by Kothandaraman (2010) were used for modeling the absorption. The absorption column consists of several separation trays and operates at 20 bar, while the desorption column is operated at 1 bar. The purity of the CO_2 stream recovered as the top product was fixed at 95% to limit solvent loss.

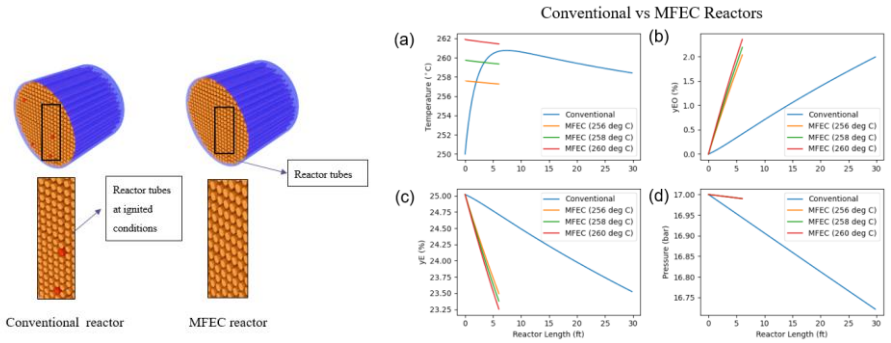


Figure 2. Ignited tubes inside conventional reactor vs. MFEC reactor (left) and its effect on reactor temperature, pressure and ethylene conversion (right)

2.5. Tube burnout inside EO reactor

The exothermic partial oxidation reaction necessitates the use of a large number of long, small-diameter tubes inside the EO reactor to achieve fast convective heat removal (Partopour & Dixon, 2016). Moreover, localized hotspots can originate from flow channeling inside the reactor, causing some tubes to operate under runaway conditions, producing unwanted, complete oxidation of ethylene in the ignited tubes, resulting in much lower overall selectivity (Geitenbeek *et al*., 2018; Kimmerle *et al*., 2009). As indicated in Figure 2 (left), EO reactors are continuously run even when a number of tubes are operating under ignited conditions which considerably impact reactor temperature profiles for conventional reactors. Ignited tubes inside the reactor result in waste of valuable raw materials and loss of product. Considerable amounts of excess waste CO_2 is produced inside the reactor. MFEC reactors which have better heat

dissipation capabilities have the potential to prevent most of these problems. Figure 2(a) shows the reactor operating conditions for both MFEC and conventional reactor (Fogler, 2011). Figure 2(b-c) show the ethylene oxide (yEO) and ethylene (yE) mole fractions along the reactor, respectively. The figure shows six feet of tubes might be sufficient compared to thirty feet of conventional reactor with same number (24800) of tubes. Figure 2(d) shows the pressure drop along the reactor for both reactor configurations.

3. Optimization Algorithm

Figure 3 illustrates the DFO algorithm, a type of optimization algorithm which does not use derivatives to reach optimality. The objective function is treated as a black-box model or analytical equations, which provides the objective function values for a set of decision variables. Next, both a set of decision variables and the objective function combination is passed back to the DFO algorithm, where a new set of decision variables are calculated based on optimality direction and subsequently further evaluated. This process is carried until the termination criterion (number of iterations) is met.

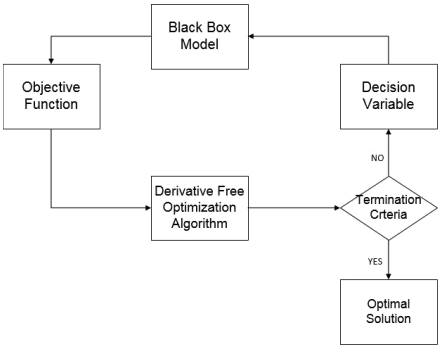


Figure 3. DFO algorithm

In this work, a DFO based algorithm named RBFOpt (Radial Basis Function) is used where a surrogate-model-based search method is employed to select the optimal design and operating conditions to achieve minimized objective function (Costa & Nannicini, 2018; Rios & Sahinidis, 2013). We used annualized cost as our objective function, which is calculated using Equations (1) - (8). The cost of each piece of equipment was calculated using standard cost coefficients from literature (Turton *et al*., 2008). The raw material (EO and ethylene) cost has been calculated using a commodity chemical database (ICIS Chemical Business, October 2020).

$$min:\ TAC = \frac{i(i+1)^n}{(i+1)^n - 1} \times IC + AUC \tag{1}$$

$$st.\ IC = \sum_j Cost_j(q_{s,j}) \tag{2}$$

$$AUC = 24 \times 300 \times \sum_j Utility_j(q_{o,j}) \tag{3}$$

$$q = \Theta\ (x,\ z,\ p(x)) \tag{4}$$

$$\sum_j Utility_j \leq Utility_0 \tag{5}$$

$$x_{product} \geq purity \quad (6)$$

$$x \in X \quad (7)$$

$$z \in Z \quad (8)$$

4. Results and Discussion

We used the DFO algorithm to find the optimum configuration for the overall EO process. The optimized process configuration parameters are given in Table 1 for varying degrees of tube ignition (note that BU indicates the bounds using during optimization). As CO_2 formation due to ignition increases inside the reactor, the requirements for the separation equipment increase. It can also be seen that the EO purification column (T-203) contributes most to the separation cost.

Table 1: Optimal process configuration for the EO process

	0%	2%	4%	10%	BU
No. Stages in EO Absorption Column, T-201	15	15	15	15	8-28
Solvent Flowrate in Column T-201 (ton/hr)	188	191	195	209	1-1000
No. Stages in EO Purifier Column, T-202	14	14	15	18	5-20
Feed Stage in EO Purifier Column, T-202	8	9	5	7	3-20
Pressure in EO Purifier Column, T-202 (bar)	2	2	2	2	1-15
No. Stages in EO Purifier Column, T-203	65	68	70	84	35-95
Feed Stage in EO Purifier Column, T-203	50	32	25	44	3-95
Reboiler Duty, T-203 (MW)	29	30	30	31	1-100
Pressure in EO Purifier Column-2, T-203 (bar)	1	1	1	1	1-15
No. Stages in CO_2 Absorption Column, T-301	6	5	5	5	4-7
Solvent Flowrate, T-301 (ton/hr)	21	26	31	48	1-100
No. Stages in CO_2 Desorption Column, T-302	9	11	13	20	3-22
Feed Stage in CO_2 Desorption Column, T-302	4	4	3	3	3-22

The capital and annualized cost for different varying levels of tube ignition in the EO process are shown in Table 2 (note that CV indicates the conventional reactor configuration). As the ignition percentage increases, the annualized cost also increases due to increased solvent requirements and an increase in the associated reboiler duties of the purification columns in Table 1. A larger overall cost reduction is found for the MFEC process mainly related to catalyst cost, as the MFEC system requires much less catalyst inside the reactor than the conventional process.

Table 2: Optimal process configuration for the EO Process

	0% MFEC	0% CV	2% CV	4% CV	10% CV
Capital Cost (10^6\$)	170.08	179.08	180.79	181.46	183.89
Annualized Cost (10^6\$)	167.45	179.05	180.96	181.78	184.84
EO Production (kton/year)	111.12	111.12	109.27	107.04	100.00
Cost of EO Production (\$/tons)	1506.93	1611.32	1656.08	1698.24	1848.40

5. Conclusions

In this paper, we have investigated various reactor ignition cases of the EO process and examined how the conditions inside the reactor and the use of MFEC can significantly impact the overall economics of the process. We have used DFO optimization to find the optimum process configuration in each case. If tube ignition can be avoided inside the reactor, the separation cost can be reduced for the intensified process by using MFEC, while also achieving optimum design and operating conditions. Moreover, our economic analysis clearly shows that the use of MFECs can significantly reduce the annualized cost of production from $1,848/ton to $1,506/ton, corresponding to a reduction of 18.4%.

References

Barecka, M. H., Skiborowski, M., & Górak, A. (2017). A novel approach for process retrofitting through process intensification: Ethylene oxide case study. Chemical Engineering Research and Design, 123, 295-316.

Brueske, S., Kramer, C., & Fisher, A. (2015). Bandwidth Study on Energy Use and Potential Energy Saving Opportunities in US Chemical Manufacturing (No. DOE/EE-1229). Energetics.

Choudhury, H. A., Cheng, X., Afzal, S., Prakash, A. V., Tatarchuk, B. J., & Elbashir, N. O. (2020). Understanding the deactivation process of a microfibrous entrapped cobalt catalyst in supercritical fluid Fischer-Tropsch synthesis. Catalysis Today, 343, 112-124.

Costa, A., & Nannicini, G. (2018). RBFOpt: an open-source library for black-box optimization with costly function evaluations. Mathematical Programming Computation, 10(4), 597-629.

Fogler, S. H. (2011). Essentials of chemical reaction engineering.

Geitenbeek, R. G., Nieuwelink, A. E., Jacobs, T. S., Salzmann, B. B., Goetze, J., Meijerink, A., & Weckhuysen, B. M. (2018). In situ luminescence thermometry to locally measure temperature gradients during catalytic reactions. ACS catalysis, 8(3), 2397-2401.

ICIS Chemical Business (2020): October 2020 Ethylene Oxide.

Kalakul, S., Malakul, P., Siemanond, K., & Gani, R. (2014). Integration of life cycle assessment software with tools for economic and sustainability analyses and process simulation for sustainable process design. Journal of cleaner production, 71, 98-109.

Kimmerle, B., Grunwaldt, J. D., Baiker, A., Glatzel, P., Boye, P., Stephan, S., & Schroer, C. G. (2009). Visualizing a catalyst at work during the ignition of the catalytic partial oxidation of methane. The Journal of Physical Chemistry C, 113(8), 3037-3040.

Kobe, J. M., Evans, W. E., June, R. L., & Lemanski, M. F. (2002). Epoxidation Industrial. Encyclopedia of catalysis.

Kothandaraman, A. (2010). Carbon dioxide capture by chemical absorption: a solvent comparison study (Vol. 72, No. 01).

Partopour, B., & Dixon, A. G. (2016). Reduced microkinetics model for computational fluid dynamics (CFD) simulation of the fixed-bed partial oxidation of ethylene. Industrial & Engineering Chemistry Research, 55(27), 7296-7306.

Peschel, A., Jörke, A., Sundmacher, K., & Freund, H. (2012). Optimal reaction concept and plant wide optimization of the ethylene oxide process. Chemical engineering journal, 207, 656-674.

Rios, L. M., & Sahinidis, N. V. (2013). Derivative-free optimization: a review of algorithms and comparison of software implementations. Journal of Global Optimization, 56(3), 1247-1293.

Turton, R., Bailie, R. C., Whiting, W. B., & Shaeiwitz, J. A. (2008). Analysis, synthesis and design of chemical processes. Pearson Education.

Sheng, M., Gonzalez, C. F., Yantz Jr, W. R., Cahela, D. R., Yang, H., Harris, D. R., & Tatarchuk, B. J. (2013). Micro scale heat transfer comparison between packed beds and microfibrous entrapped catalysts. Engineering Applications of Computational Fluid Mechanics, 7(4), 471-485.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50117-2

Power Systems Infrastructure Planning with High Renewables Penetration

Can Li[a*], Antonio J. Conejo[b], Peng Liu[c], Benjamin P. Omell[c], John D. Siirola[d], Ignacio E. Grossmann[a]

[a]*Department of Chemical Engineering, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, USA*
[b]*Department of Integrated Systems Engineering, The Ohio State University, 1971 Neil Avenue, Columbus, OH 43210,USA*
[c]*National Energy Technology Laboratory, Pittsburgh, PA 15236, United States*
[d]*Center for Computing Research, Sandia National Laboratories, P.O. 5800, Albuquerque, NM, 87185, USA*

canli.pse@gmail.com

Abstract

With the increasing penetration of renewable generating units, especially in remote areas not well connected with load demand, there is growing interest to co-optimize generation and transmission expansion planning (GTEP) in power systems. Due to the volatility in renewable generation, a planner needs to include the operating decisions into the planning model to guarantee feasibility. Three different formulations, i.e., a big-M formulation, a hull formulation, and an alternative big-M formulation, are reported for transmission expansion. To address the computational challenge, we propose a nested Benders decomposition algorithm and a tailored Benders decomposition algorithm that exploit the structure of the GTEP problem. Using a case study from Electric Reliability Council of Texas (ERCOT), we are able to show that the proposed tailored Benders decomposition outperforms the nested Benders decomposition. The coordination in the optimal generation and transmission expansion decisions from the ERCOT study implies that there is an additional value in solving GEP and TEP simultaneously. The detailed results of this paper has been published in Li et al. (2021).

Keywords: Power Systems, Generation Transmission Expansion, Mixed-integer Programming, Decomposition Algorithm

1. Introduction

Generation expansion planning (GEP) of power systems involves determining the optimal size, location, and construction time of new power generation plants, while minimizing the total cost over a long-term planning horizon (Conejo et al., 2016). There is a growing interest to use mathematical programming models to solve generation expansion planning problems (Lara et al., 2018; Tso et al., 2020). Conventional power units are dispatchable thermal power plants that can provide stable power output. However, with the increased penetration of renewable generation technologies, such as solar and wind, power systems nowadays need to be more flexible so as to adjust to the volatile power generation from

renewables. In this case, operations decisions, such as unit commitment, ramping decisions, become important to assess system feasibility. Transmission expansion planning (TEP) refers to installing new transmission lines or expanding the capacities of existing transmission lines in a power system. Bahiense et al. (2001) propose a mixed integer disjunctive model for transmission network expansion. GEP and TEP are generally solved as two independent optimization problems. However, the significant penetration of renewables into power systems may lead to their concentration in remote areas not well connected to load demand. Therefore, installing renewables in those remote areas could compromise transmission expansion. The recognition of transmission's interaction with generation expansion has motivated the development of co-optimization methods to consider the tradeoffs between generation and transmission expansion (Krishnan et al., 2016). This paper is an extension of the GEP model reported in Lara et al. (2018) to a GTEP model. The long version of this paper has been published in Li et al. (2021).

2. Problem Statement

We are given different types of existing and known generating units and the generating units' nameplate (maximum) capacity; expected lifetime; fixed and variable operating costs; fixed and variable start-up cost; cost for extending their lifetimes; CO_2 emission factor and carbon tax, if applicable; fuel price, if applicable; and operating characteristics such as ramp-up/ramp-down rates, operating limits, contribution to spinning and quick start fraction for thermal generators, and capacity factor for renewable generators. Also given are existing and candidate transmission lines between any of the two neighboring buses. The susceptance, distance, and capacity of each transmission line are known. We use DC power flow equations to calculate the power flow in each transmission line. These equations are built based on Kirchhoff's voltage and current laws which differ from the network flow model used in the work of Lara et al. (2018). In the network flow model, the transmission network is represented similarly to pipelines where the flows only observe energy balance at each node while ignoring Kirchhoff's laws.

With the above input data, the spatial and temporal representations in Li et al. (2021), the proposed GTEP model is to decide: a) when and where to install new generators, storage units and transmission lines; b) when to retire generators and storage units; c) whether or not to extend the life of the generators that reached their expected lifetime; d) unit commitment of the thermal generators during the representative days; e) power generation of the generator clusters and power flows through the transmission lines. The objective is to minimize the overall cost including operating, investment, and environmental costs (e.g., carbon tax and renewable generation quota).

3. Transmission Expansion Formulation

One of the major constributions of this paper is to compare different formulations for transmission expansion. For the candidate transmission lines, we can write the following disjunction, where $NTE_{l,t}$ is a logic variable whose value can be True or False indicating whether or not transmission line l is installed in year t. If line l already exists in year t, the corresponding power flow has to satisfy DC power flow equation and upper and lower bounds. Otherwise, the corresponding power flow is zero. We assume that all the

candidate transmission lines are standard. In other words, the susceptance of the candidate transmission lines B_l are parameters in the model.

$$\begin{bmatrix} NTE_{l,t} \\ p^{\text{flow}}_{l,t,d,s} = B_l(\theta_{s(l),t,d,s} - \theta_{r(l),t,d,s}) \\ -F^{\max}_l \le p^{\text{flow}}_{l,t,d,s} \le F^{\max}_l \end{bmatrix} \vee \begin{bmatrix} \neg NTE_{l,t} \\ p^{\text{flow}}_{l,t,d,s} = 0 \end{bmatrix} \quad \forall l \in \mathcal{L}^{\text{new}}, t, d, s \tag{1}$$

Standard approaches, i.e., big-M reformulation and hull reformulation (Grossmann and Trespalacios, 2013), are available to reformulate disjunctions (1) into mixed integer constraints.

The big-M formulation of the disjunction is,

$$-M_l(1-nte_{l,t}) \le p^{\text{flow}}_{l,t,d,s} - B_l(\theta_{s(l),t,d,s} - \theta_{r(l),t,d,s}) \le M_l(1-nte_{l,t}) \quad \forall l \in \mathcal{L}^{\text{new}}, t, d, s \tag{2}$$

$$-F^{\max}_l nte_{l,t} \le p^{\text{flow}}_{l,t,d,s} \le F^{\max}_l nte_{l,t} \quad \forall l \in \mathcal{L}^{\text{new}}, t, d, s \tag{3}$$

This big-M formulation is most commonly used in the literature (Conejo et al., 2016) for TEP.

The hull formulation is,

$$p^{\text{flow}}_{l,t,d,s} = B_l \Delta\theta^1_{l,t,d,s} \quad \forall l \in \mathcal{L}^{\text{new}}, t, d, s \tag{4}$$

$$\theta_{s(l),t,d,s} - \theta_{r(l),t,d,s} = \Delta\theta^1_{l,t,d,s} + \Delta\theta^2_{l,t,d,s} \quad \forall l \in \mathcal{L}^{\text{new}}, t, d, s \tag{5}$$

$$-\pi \cdot nte_{l,t} \le \Delta\theta^1_{l,t,d,s} \le \pi \cdot nte_{l,t} \quad \forall l \in \mathcal{L}^{\text{new}}, t, d, s \tag{6}$$

$$-\pi(1 - nte_{l,t}) \le \Delta\theta^2_{l,t,d,s} \le \pi(1 - nte_{l,t}) \quad \forall l \in \mathcal{L}^{\text{new}}, t, d, s \tag{7}$$

where $\Delta\theta^1_{l,t,d,s}$ and $\Delta\theta^2_{l,t,d,s}$ are disaggregated variables for the angle difference of transmission line l. Variable $\Delta\theta^1_{l,t,d,s}$ is equal to the angle difference if transmission line l has been installed in year t. Otherwise, $\Delta\theta^2_{l,t,d,s}$ equals to the angle difference. In addition to equations (4)-(7), equation (3) needs to be included in the hull formulation. The hull formulation has more continuous variables than the big-M formulation but it avoids using the big-M parameters of equations (2).

Alternative big-M formulation: Besides the big-M and hull formulations, an alternative big-M formulation is proposed by Bahiense et al. (2001). In this formulation, additional continuous variables $p^{\text{flow}+}_{l,t,d,s}$, $p^{\text{flow}-}_{l,t,d,s}$, $\Delta\theta^+_{l,t,d,s}$, $\Delta\theta^-_{l,t,d,s}$, are introduced, where the superscript '+' means that the flow is in the same direction as the nominal direction of transmission line l, i.e., from the sending-end node $s(l)$ to the receiving-end node $r(l)$; superscript '-' means the opposite direction. By defining these new continuous variables, equation (2) is replaced by equations (8a) to (8d) and equation (3) is replaced by equations (8g) and (8h).

$$p^{\text{flow}+}_{l,t,d,s} - B_l \Delta\theta^+_{l,t,d,s} \le 0 \quad \forall l \in \mathcal{L}^{\text{new}}, t, d, s \tag{8a}$$

$$p^{\text{flow}-}_{l,t,d,s} - B_l \Delta\theta^-_{l,t,d,s} \le 0 \quad \forall l \in \mathcal{L}^{\text{new}}, t, d, s \tag{8b}$$

$$p^{\text{flow}+}_{l,t,d,s} - B_l \Delta\theta^{+}_{l,t,d,s} \geq -M_l(1 - nte_{l,t}) \quad \forall l \in \mathcal{L}^{\text{new}}, t, d, s \tag{8c}$$

$$p^{\text{flow}-}_{l,t,d,s} - B_l \Delta\theta^{-}_{l,t,d,s} \geq -M_l(1 - nte_{l,t}) \quad \forall l \in \mathcal{L}^{\text{new}}, t, d, s \tag{8d}$$

$$p^{\text{flow}}_{l,t,d,s} = p^{\text{flow}+}_{l,t,d,s} - p^{\text{flow}-}_{l,t,d,s} \quad \forall l \in \mathcal{L}^{\text{new}}, t, d, s \tag{8e}$$

$$\theta_{s(l),t,d,s} - \theta_{r(l),t,d,s} = \Delta\theta^{+}_{l,t,d,s} - \Delta\theta^{-}_{l,t,d,s} \quad \forall l \in \mathcal{L}^{\text{new}}, t, d, s \tag{8f}$$

$$p^{\text{flow}+}_{l,t,d,s} \leq F^{\max}_{l} nte_{l,t} \quad \forall l \in \mathcal{L}^{\text{new}}, t, d, s \tag{8g}$$

$$p^{\text{flow}-}_{l,t,d,s} \leq F^{\max}_{l} nte_{l,t} \quad \forall l \in \mathcal{L}^{\text{new}}, t, d, s \tag{8h}$$

$$p^{\text{flow}+}_{l,t,d,s}, p^{\text{flow}-}_{l,t,d,s}, \Delta\theta^{+}_{l,t,d,s}, \Delta\theta^{-}_{l,t,d,s} \geq 0 \quad \forall l \in \mathcal{L}^{\text{new}}, t, d, s \tag{8i}$$

4. Proposed Algorithms

The proposed MILP GTEP model typically involves millions or tens of millions of variables, which makes the model not directly solvable by the commercial solvers. We propose two algorithms to efficiently solve this problem.

4.1. Nested Benders decomposition

Lara et al. (2018) apply a nested Benders decomposition algorithm to solve their GEP model. Like in the GEP model, the nested Benders decomposition algorithm decomposes the fullspace of the GTEP problem by year.

4.2. Benders decomposition

Instead of solving the GTEP problem sequentially by year as in the nested Benders decomposition, we treat all the investment-related variables as complicating variables and include all these variables in a single Benders master problem. The Benders algorithm iterates between the Benders master problem and the Benders subproblems.

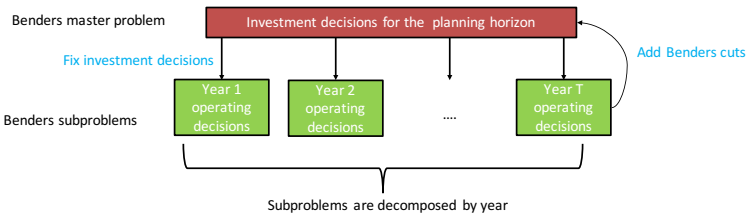


Figure 1: Tailored Benders decomposition algorithm applied to the GTEP problem

5. Results

We carry out a GTEP case study for ERCOT (Texas region in the US). It is divided into five geographical regions: Northeast, West, Coast, South, and Panhandle. We also test the two decomposition algorithms described in section 4. The nested Benders decomposition is implemented in Pyomo/Python (Hart et al., 2011). The tailored Benders decomposition implementation is from CPLEX. The computational results of the two proposed decomposition algorithms are shown in Table 1. The tailored Benders decomposition algorithm is able to solve all the three formulations to within 1% optimality gap within 10,000 seconds wall time (real wall clock time).

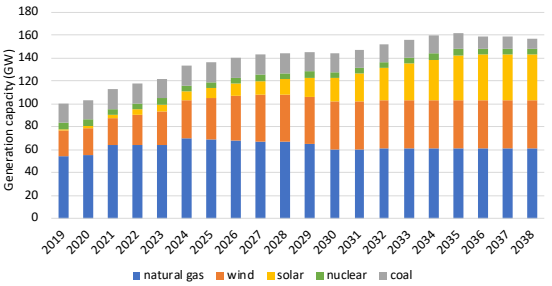


Figure 2: Aggregated generation expansion results

Table 1: Computational results of the two proposed decomposition algorithms using different formulations

Algorithm	Formulation	UB ($\$10^9$)	LB ($\$10^9$)	Gap	Wall time (secs)
tailored Benders	big-M	283.7	282.6	0.38%	5,115
tailored Benders	alternative big-M	283.9	281.6	0.82%	3,693
tailored Benders	hull	282.6	280.6	0.71%	8,418
nested Benders	big-M	295.7	268.9	9.98%	53,682
nested Benders	alternative big-M	294.2	265.5	10.81%	43,389
nested Benders	hull	288.0	269.3	6.97%	37,577

The capacities of different generation technologies from 2019 to 2038 are shown in Figure 2. The results include high capacities of solar and wind. The aggregated natural gas capacity of the five regions increases in the first few years, reaches its peak in 2024 and gradually decreases afterwards due to the retirement of old generators and the increase in carbon tax, which makes the natural gas generators less competitive compared with solar and wind generators. The nuclear capacities are unchanged throughout the planning horizon. The coal capacities are unchanged in the first few years

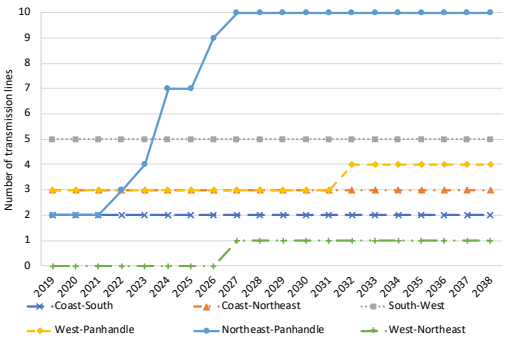


Figure 3: Transmission expansion results

and start decreasing in 2029 because of reaching their nominal lifetimes. No storage unit is installed. Therefore, the renewable generation when the net load is negative has to be curtailed. The total discounted renewable curtailment cost is $1.64 billion in 20 years. The number of transmission lines built over the planning horizon are shown in Figure 3. Most of the transmission lines are built for Northeast-Panhandle and South-West in order to transfer the power generated by the renewable sources in West and Panhandle to other regions.

6. Conclusions

We have developed models and algorithms for capacity expansion of power systems with high penetration of renewables. For PSE researchers, the capability to analyze power systems enables to study hybrid energy systems that have both electricity generators and electricity/heat consumers, such as chemical plants.

References

L. Bahiense, G. C. Oliveira, M. Pereira, and S. Granville. A mixed integer disjunctive model for transmission network expansion. *IEEE Transactions on Power Systems*, 16 (3):560–565, 2001.

A. J. Conejo, L. Baringo, S. J. Kazempour, and A. S. Sissiqui. *Investment in Electricity Generation and Transmission - Decision Making under Uncertainty*. Springer International Publishing, 2016. ISBN 978-3-319-29501-5. doi: 10.1007/978-3-319-29501-5.

I. E. Grossmann and F. Trespalacios. Systematic modeling of discrete-continuous optimization models through generalized disjunctive programming. *AIChE Journal*, 59(9): 3276–3295, 2013.

W. E. Hart, J.-P. Watson, and D. L. Woodruff. Pyomo: modeling and solving mathematical programs in python. *Mathematical Programming Computation*, 3(3):219, 2011.

V. Krishnan, J. Ho, B. F. Hobbs, A. L. Liu, J. D. McCalley, M. Shahidehpour, and Q. P. Zheng. Co-optimization of electricity transmission and generation resources for planning and policy analysis: review of concepts and modeling approaches. *Energy Systems*, 7(2):297–332, 2016. ISSN 1868-3975.

C. L. Lara, D. S. Mallapragada, D. J. Papageorgiou, A. Venkatesh, and I. E. Grossmann. Deterministic electric power infrastructure planning: Mixed-integer programming model and nested decomposition algorithm. *European Journal of Operational Research*, 271(3):1037–1054, 2018.

C. Li, A. J. Conejo, P. Liu, B. P. Omell, J. D. Siirola, and I. E. Grossmann. Mixed-integer linear programming models and algorithms for generation and transmission expansion planning of power systems. *European Journal of Operational Research*, 2021.

W. W. Tso, C. D. Demirhan, C. F. Heuberger, J. B. Powell, and E. N. Pistikopoulos. A hierarchical clustering decomposition algorithm for optimizing renewable power systems with storage. *Applied Energy*, 270:115190, 2020.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50118-4

An Optimization Model for the Design and Operation of Reliable Power Generation Systems

Seolhee Cho[a] and Ignacio E. Grossmann[a*]

[a] *Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA*
** Corresponding Author: grossmann@cmu.edu*

Abstract

This paper aims to develop a new optimization model for the design and operation of reliable power generation systems. This work optimizes the selection of redundant or backup units and operating units to maximize the reliability and to minimize the cost. In particular, every possible failure state that the power generation systems can have is investigated to evaluate the system reliability. To achieve this goal, we develop an optimization model that minimizes the total cost using Generalized Disjunctive Programming (GDP). The GDP model includes two decision variables: the first is a selection of redundant units with different sizes to increase the reliability of systems, and the second is a selection of operating units to satisfy the power demand. Specifically, the model determines the system reliability and corresponding expected power production by considering the number of redundant and operating units, and possible failure states under each design and operation mode. The model imposes a penalty when the demand is not satisfied, and the system has a low reliability. We have applied the proposed model in a small power plant (one stage with up to three generators) and verified through a sensitivity analysis that the model installs larger and more units to improve the system reliability as penalty rates increase.

Keywords: Redundancy, Reliability, Design, Operation, Optimization

1. Introduction

As evidenced by the Texas power crisis in 2021[1], the failure of power generation systems can lead to extreme events. Therefore, power generation systems should be designed to have high reliability to withstand failures of one or multiple components, and supply near uninterrupted electric power to industries and households. Reliability indicates the probability that a system will perform its required function properly even if one or multiple units fail (Sherwin et al., 2020). Since the goal of power generation systems is to consistently provide electric power, securing high reliability in their design and operation is a highly desirable objective. Numerous studies on the design/planning of power generation systems and reliability evaluation have been reported. However, previous works have dealt with these problems separately. Lara et al. (2018) have focused on optimizing the generation capacity to satisfy the power demand, whereas Amusat et al. (2016) have evaluated the reliability after designing the power generation systems. Studies that integrate design and reliability have also been reported, but they only consider a couple of generators or transmission lines failures (Moreira et al., 2016). Given the

[1] In February 2021, massive electricity generation systems in Texas were failed due to severe winter storms. Due to this, more than 4.5 million homes and businesses were left without power for several days, and at least 210 people were killed.

recent crisis, there is a strong motivation for a more comprehensive method to consider reliability in the design and operation phases of the power generation systems.

One method to improve reliability of the power generation systems at the design phase is to add redundant or backup units, which allows the systems to operate even if one or multiple generators fail (Kim et al., 2016). This approach is known as 'reliability-based design optimization,' and various studies on this topic have been conducted. Ye et al. (2018) develop a mixed integer nonlinear programming model for the optimal design of chemical process. Ortiz-Espinoza et al. (2021) present a multi-objective reliability-based design optimization model by combining economics, reliability, and safety. Chen et al. (2021) propose a two-stage stochastic generalized disjunctive programming (GDP) model by considering reliability and endogenous/exogenous uncertainties. Since these works assume that their target systems operate at a steady state, the authors mainly focus on optimizing the number of redundant units for the reliable design. However, since power systems operate in unsteady state due to time-varying power demand, reliability is also influenced by the operational strategies that the systems use to satisfy the load demand. Specifically, backup units can have a dual role in power generation systems. They can remain as backup units in case of low power demand or change to operating units when the power demand is high. Such dual purpose of redundant units depending on the load demand should be considered in the design and operation of reliable power generation systems. To our knowledge, this issue has not been addressed before in the literature for the design optimization of power generation systems with considerations of reliability.

This paper aims to develop a new optimization model for the design and operation of reliable power generation systems. This work optimizes the number of redundant units and operating units to maximize the reliability and to minimize the cost. In the remainder of this paper, we develop an optimization model using Generalized Disjunctive Programming (GDP), which is a high-level model representation that involves equations, disjunctions, and logic propositions. We then verify the effectiveness of the proposed model by solving an illustrative example.

2. Problem statement

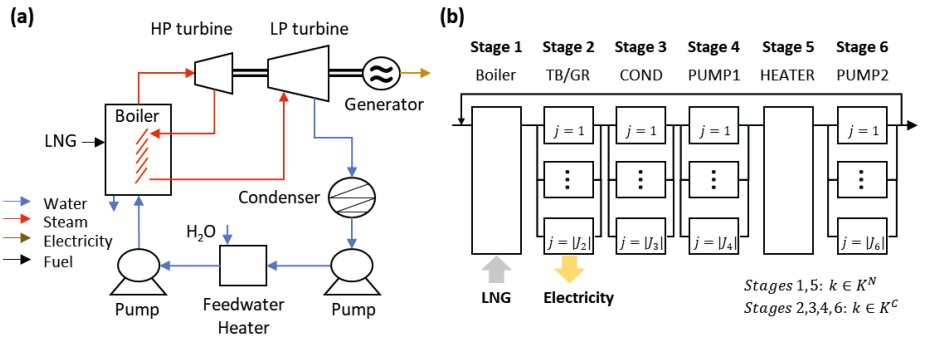


Figure 1. (a) Flow diagram of thermal power plant, (b) circular-parallel systems structure

Given is a natural gas power plant with multiple stages $k \in K$ including turbine, boiler, and pump, parallel identical units $j \in J_k$ for each stage k, and set of discrete capacities $c \in C_k$ of the units j in stages k. The stages can be classified into two groups: noncritical stage $k \in K^N$ that do not consider reliability, and critical stages $k \in K^C$ that do consider

reliability. A set of stage designs $h \in H_k$ and corresponding operation modes $m \in M_{k,h}$, and time periods $t \in T$ are also given. Specifically, $h = 1$ means one unit installation and $h = H$ means all potential units are installed. Likewise, $m = 1$ represents one unit operation mode, $m = M$ refers to the mode in which all units are simultaneously operated. Each stage k has different failure states $s \in S_{m,k,h}$ depending on design h and operation mode m, which can be classified into (i) successful operation states ($S^F_{m,k,h}$) and (ii) partial operation states $\left(S^P_{m,k,h}\right)$. There are also corresponding operating reliabilities: successful operation reliability ($R^F_{k,t}$) and partial operation reliability ($R^P_{k,t}$). While 'successful operation states' indicate the operation states in which the power generation capacity is sufficient to satisfy the load demand, 'partial operation states' refer to the operation states in which the power generation capacity is insufficient to meet the load demand, but still can produce electric power at a limited level. The $R^F_{k,t}$ and $R^P_{k,t}$ are probabilities that such successful or partial operation states exist in stage k in time t, respectively. The major assumptions used in the model are: (i) There is one unit in noncritical stages, which has sufficiently large capacity; (ii) A redundant unit can be a backup of any operating unit; (iii) Repair and maintenance processes are not considered.

3. Model formulation

The model is developed using Generalized Disjunctive Programming (GDP) (Grossmann and Trespalacios, 2013), which can be expressed in terms of Boolean and continuous variables, algebraic constraints, disjunctions, and logic propositions.

$$
\begin{bmatrix}
Z_{k,h} \\
\sum_{c \in C_k} y_{k,j,c} = i, \qquad i = 1, \dots, h
\end{bmatrix}
\quad \textbf{(a) Investment disjunction}
$$

$$
\textbf{(1) Number of installed unit}
$$

$$
\bigvee_{h \in H_k}
\begin{bmatrix}
W_{m,k,h,t} \\
\sum_{j \in J_k} \sum_{c \in C_k} x_{k,j,c,t} = i, \qquad i = 1, \dots, m \\
\sum_{c \in C_k} \varepsilon_j \rho_{k,c} x_{k,j,c,t} \le AUC_{k,j,t} \quad \forall j \in J_k \\
AUC_{k,j,t} \le \sum_{c \in C_k} \rho_{k,c} x_{k,j,c,t} \quad \forall j \in J_k \\
R^F_{k,t} = \sum_{s \in S^F_{m,k,h}} \left\{ \prod_{j \in J^O_{s,m,h}} \lambda_{j,k} \prod_{j \in J^N_{s,m,h}} \sigma_{j,k} \right\} \\
R^P_{k,t} = \sum_{s \in S^P_{m,k,h}} \left\{ \prod_{j \in J^O_{s,m,h}} \lambda_{j,k} \prod_{j \in J^N_{s,m,h}} \sigma_{j,k} \right\} \\
EP^s_{k,t} \le \omega_t \sum_{j \in J^O_{s,m,h}} \sum_{c \in C_k} AUC_{k,j,t} R^s_{k,t} \quad s = 1 \\
EP^s_{k,t} = \omega_t \sum_{j \in J^N_{s,m,h}} IC_{k,j} R^s_{k,t} \quad \forall s \ne 1, s \in S_{m,k,h}
\end{bmatrix}
\quad \textbf{(b) Operation disjunction}
$$

$$
\textbf{(2) Number of operating unit}
$$

$$
\left.\begin{array}{l} \\ \\ \\ \end{array}\right\} \textbf{(3) Available operating capacity of unit in time}
$$

$$
\left.\begin{array}{l} \\ \\ \\ \\ \\ \end{array}\right\} \textbf{(4) Successful and partial operation reliability}
$$

$$
\textbf{(5) Expected production by successful operation}
$$

$$
\textbf{(6) Expected production by partial operation}
$$

$$
t \in T \qquad k \in K^C
$$

$$
\begin{cases}
\sigma_{j,k} = 1 & \forall j \in J^B_{s,m,h}, k \in K \qquad J^B_{s,m,h} \cup J^F_{s,m,h} = J^N_{s,m,h} \quad \forall j \in J_k \\
\sigma_{j,k} = \left(1 - \lambda_{j,k}\right) & \forall j \in J^F_{s,m,h}, k \in K \qquad J^O_{s,m,h} \cup J^N_{s,m,h} = \emptyset \quad \forall j \in J_k
\end{cases}
\qquad \textbf{(7)}
$$

There are two Boolean variables related to investment and operation decisions. $Z_{k,h}$ is true if design h is selected for stage k; false otherwise (Equation (a)). $W_{m,k,h,t}$ is true if stage k is in operation mode m in time t for design h; false otherwise (Equation (b)). The

binary variable $y_{k,j,c}$ indicates the installation of unit j with specified capacity c in stage k and $x_{k,j,c,t}$ indicates the operation of unit j with specificed capacity c in stage k and time t. $\lambda_{j,k}$ is a reliability of unit j in stage k and $\sigma_{j,k}$ defined in Eqn (7), states that when the unit j belongs to set of backup units in failure state s under design h and operation mode m ($J^B_{s,m,h}$), the unit reliability ($\lambda_{j,k}$) will be 1. If the unit j belongs to set of failed unit ($J^F_{s,m,h}$), the unit unreliability ($1-\lambda_{j,k}$) will be used to calculate system reliability.

$$\sum_{c\in C_k} y_{k,j,c} \le 1 \quad \forall k \in K^C, j \in J_k \tag{8}$$

$$\sum_{c\in C_k} x_{k,j,c,t} \le 1 \quad \forall k \in K^C, j \in J_k, t \in T \tag{9}$$

$$IC_{k,j} = \sum_{c\in C_k} \rho_{k,c} y_{k,j,c} \quad \forall k \in K^C, j \in J_k \tag{10}$$

$$TEP_{k,t} = \sum_{s\in S} EP^s_{k,t} \quad \forall k \in K^C, t \in T \tag{11}$$

$$ASC_{k,t} = \sum_{j\in J_k} AUC_{k,j,t} \quad \forall k \in K^C, t \in T \tag{12}$$

$$DC_{k,t} = \xi R^P_{k,t} \quad \forall k \in K^C, t \in T \tag{13}$$

$$\left.\begin{array}{l} \veebar_{h\in H_k} Z_{k,h} \quad \forall k \in K^C \\ Z_{k,h} \Leftrightarrow \veebar_{m\in M_{k,h,t}} W_{m,k,h,t} \quad \forall h \in H_k, k \in K^C, t \in T \end{array}\right\} \tag{14}$$

$$\left.\begin{array}{l} y_{k,j+1,c} \le y_{k,j,c} \quad \forall k \in K^C, j \in J_k, c \in C_k \\ x_{k,j+1,c,t} \le x_{k,j,c,t} \quad \forall k \in K^C, j \in J_k, c \in C_k, t \in T \end{array}\right\} \tag{15}$$

$$y_{k,j,c} \ge x_{k,j,c,t} \quad \forall k \in K^C, j \in J_k, c \in C_k, t \in T \tag{16}$$

$$\left.\begin{array}{l} ASC_{k,t} \le \psi_k \quad \forall k \in K^N, t \in T \\ TEP_{k,t} \le ASC_{k,t} \quad \forall k \in K^N, t \in T \end{array}\right\} \tag{17}$$

$$TEP_{k+1,t} = \eta_k TEP_{k,t} \quad \forall k \in K, t \in T \tag{18}$$

$$Z_{k,h}, W_{m,k,h,t} \in \{True, False\};\ x_{k,j,c,t}, y_{k,j,c} \in \{0,1\};\ IC_{k,j}, AUC_{k,j,t}, ASC_{k,t}, R^s_{k,t}, EP^s_{k,t}, TEP_{k,t}, DC_{k,t}, F_t \ge 0$$

Eqns. (8) and (9) state that only one capacity can be installed and operated. Eqns. (10) – (12) indicate capacity of unit installed in stage k, total expected power production, and available capacity of stage k, respectively. Eqn. (13) is a downtime penalty and Eqn. (14) is a logic constraint for disjunction. Eqn. (15) are symmetry breaking constraints, meaning that a unit can only be selected if the one with higher priority is selected. Eqn. (16) indicates that installed units only can be used. Eqn. (17) constrains the expected production and operating capacity of noncritical stage $k \in K^N$. Eqn. (18) states that the expected production of stage $k+1$ is estimated by using expected production of stage k and conversion rate of stage k.

The objective function in (19) is to minimize the total cost, which includes the investment cost ($\delta_{k,c} y_{k,j,c}$), start-up cost ($\theta_{k,c} x_{k,j,c,t}$), expected fuel cost to purchase natural gas ($\pi_t F_t$), expected operating cost ($\gamma_k TEP_{k,t}$), downtime penalty ($DC_{k,t}$), and unmet demand penalty (PN_t). The system is charged penalties for the unmet demand, as shown by Equation (20). Rather than considering a bi-criterion optimization problem, we assume that shortfalls in power demand and low reliability are penalized so as to formulate the optimization problem as a single objective problem for cost minimization.

$$\min Z = \sum_{k\in K}\sum_{j\in J_k}\sum_{c\in C_k} \delta_{k,c} y_{k,j,c} + \sum_{k\in K}\sum_{j\in J_k}\sum_{c\in C_k}\sum_{t\in T} \theta_{k,c} x_{k,j,c,t} + \sum_{t\in T} \pi_t F_t + \sum_{k\in K}\sum_{t\in T} \gamma_k TEP_{k,t} + \sum_{k\in K}\sum_{t\in T} DC_{k,t} + \sum_{t\in T} PN_t \tag{19}$$

$$\left.\begin{array}{l} \begin{bmatrix} V_{1,t} \\ TEP_{k,t} < \omega_t D_t \\ PN_t = (\omega_t D_t - TEP_{k,t})\alpha \end{bmatrix} \vee \begin{bmatrix} V_{2,t} \\ TEP_{k,t} \ge \omega_t D_t \\ PN_t = 0 \end{bmatrix} \quad \forall k = GN, t \in T \\ V_{1,t} \veebar V_{2,t} \quad \forall t \in T \end{array}\right\} \tag{20}$$

The GDP given by (1)−(20) can be transformed into a Mixed-Integer Nonlinear Programming (MINLP) using Big-M (BM) and/or Hull Reformulation (HR) (Grossmann and Trespalacios, 2013). This paper uses both methods, and Eqns. (5) and (6) are transformed into MILP constraints by using an exact linearization (Avraamidou and Pistikopoulos, 2019, and Garcia-Herreros et al., 2015).

4. Illustrative example

To verify the proposed model, the power system that has one stage ($k = 1$) involving up to three generators ($|J_k| = 3$) and three different sizes ($|C_k| = 3$) are analysed. Here the one stage stands for the generator stage. The total time horizon is 10 months, which is divided into 10 periods (i.e., 1 month). Table 1 shows the parameter values for the example.

Table 1. Parameter for illustrative example

Parameter	Symbol	Value	Parameters	Symbol	Value
Nameplate capacity (MW)	$\rho_{k,c}$	50,80, 100	Purchase cost of natural gas ($/MMBtu)	π_t	5
Minimum operating capacity (ratio of $\rho_{k,c}$, %)	ε_j	10	Production cost ($/MWh)	γ_k	5
Unit reliability	$\lambda_{j,k}$	0.97	Downtime penalty rate ($/hr)	ξ	1000
Conversion rate	η_k	0.4278	Installation cost (k$/unit)	$\delta_{k,c}$	10, 13, 15
Unmet demand penalty rate ($/MWh)	α	100	Start-up cost ($/unit)	$\theta_{k,c}$	100, 160, 200

Table 2. Numerical results of illustrative example

Solution method	MINLP		MILP	
Solver	BARON		CPLEX	
Cost (k$)	164.1		164.1	
Average reliability	**0.9758**		**0.9758**	
Reformulation	BM	HR	BM	HR
Equations	1,110	1,212	2,190	4,452
Cont. variables	267	1,083	627	2,883
Binary variables	108	360	108	360
CPU (sec.)	4.040	11.110	0.360	0.687

Figure 2. Optimal design and operation

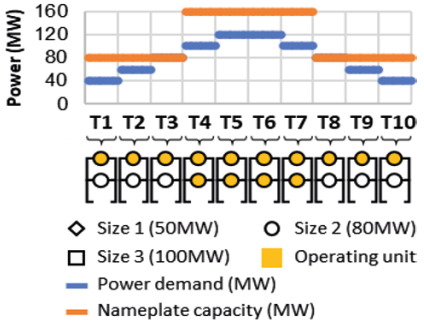


Table 2 shows the numerical results obtained with BARON and CPLEX in GAMS 32.1.0 on an Intel Core i7-10510U CPU, 1.80GHz. Although the sizes of the MILP reformulations are larger, their CPU times are significantly shorter than the MINLP. The proposed model predicts the total cost of $164,192 including unmet demand penalty of $144 and downtime penalty of $2,328. As shown in Figure 2, the model installs two medium size generators of 80 MW each yielding a total of 160MW. While the second generator remains as a backup when the demand is relatively low (from T1 – T3 and T8 – T10), both generators are used to meet the demand during T4 – T7.

5. Sensitivity analysis

To analyze the impact of unmet demand and downtime penalty rates on design and operation of reliable power systems, two alternative cases that have different penalty rates are suggested (Case 1: α = $500/MWh, ξ = $5,000/hr, Case 2: α = $1,000/MWh, ξ = $10,000/hr). As shown in Figure 3(a), the system with higher penalty rates than base case (c.f., Base case: α = $100/MWh, ξ = $1,000/hr) tends to install two larger units (each 100 MW) so as to improve reliability, and the cost is also increased to 171.9 k$. When the unmet and downtime penalties are significantly higher than other two cases (Base case and Case 1) (Figure 3(b)), the system decides to install three medium size generators (each 80 MW) and have one unit as a backup during all the periods, which results in the highest reliability (0.9989) and cost (175.1 k$).

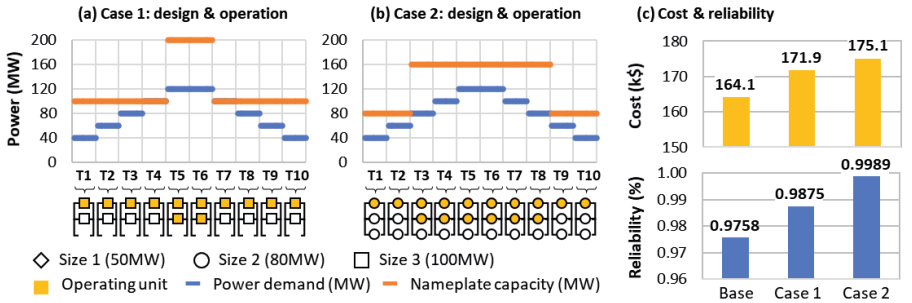


Figure 3. (a) and (b): Optimal design and operation of alternative cases, (c): total cost and average reliability of all cases

6. Conclusions

This paper has presented a mathematical optimization model for the design and operation of reliable power generation system. This work optimizes the number of redundant units and operating units to maximize the reliability and to minimize the cost. We propose a GDP formulation to represent the reliability and expected production, which are essential factors to determine the design and operation of power generation systems. Through a small example and sensitivity analysis, we found that the optimal system involves more and larger units to improve the system reliability as the penalty rates increase. Future work will involve other operation problems in power systems such as economic dispatch and unit commitment to evaluate the reliability more precisely by using a more rigorous reliability model such as Markov chain theory.

Acknowledgement

This work was conducted as part of the Institute for the Design of Advanced Energy Systems (IDAES) with support through the Simulation-Based Engineering, Crosscutting Research Program within the U.S. Department of Energy's Office of Fossil Energy and Carbon Management.

References

Amusat OO, Shearing PR, Fraga ES. Optimal integrated energy systems design incorporating variable renewable energy sources. Computers & Chemical Engineering, 2016

Avraamidou S, Pistikopoulos EN. A Bi-Level Formulation And Solution Method For The Integration of Process Design And Scheduling. Computer Aided Chemical Engineering, 2019

Chen Y, Ye Y, Grossmann IE, Chen B. Integrating Reliability and Uncertainty in Process Synthesis. Computer Aided Chemical Engineering, 2021

Garcia-Herreros P, Zhang L, Misra P, Arslan E, Mehta S, Grossmann IE. Mixed-integer bilevel opimization for capacity planning with rational markets, Computers and Chemical Engineering, 2015

Grossmann IE, Trespalacios F. Systematic modeling of discrete-continuous optimization models through generalized disjunctive programming. AIChE Journal, 2013.

Kim H, Kim P. Reliability–redundancy allocation problem considering optimal redundancy strategy using parallel genetic algorithm. Reliability Engineering & System Safety, 2017

Lara CL, Mallapragada DS, Papageorgiou DJ, Venkatesh A, Grossmann IE. Deterministic electric power infrastructure planning: Mixed-integer programming model and nested decomposition algorithm. European Journal of Operational Research, 2018.

Moreira A, Pozo D, Street A, Sauma E. Reliable Renewable Generation and Transmission Expansion Planning: Co-Optimizing System's Resources for Meeting Renewable Targets. IEEE Transactions on Power Systems, 2017.

Ortiz-Espinoza AP, Ye Y, Grossmann IE, Jiménez-Gutiérrez A. Multi-objective optimization for the incorporation of safety and reliability considerations in process design. Computer Aided Chemical Engineering, 2021

Sherwin DJ, Bossche A. The Reliability, Availability and Productiveness of Systems. Springer, 1993.

Ye Y, Grossmann IE, Pinto JM. Mixed-integer nonlinear programming models for optimal design of reliable chemical plants. Computers and Chemical Engineering, 2018

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50119-6

Rule-based Method for Retrofitting Conventional Processes with Integrated Units

Eduardo Perez-Cisneros[a], Mario R. Eden[b], Rafiqul Gani[c*]
[a]*Departmento de IPH, UAM-Iztapalapa, Mexico City, Mexico*
[b]*Department of Chemical Engineering, Auburn University, Auburn, AL 36849, USA*
[c]*PSE for SPEED Company, Ordup Jagtvej 42D, DK-2920 Charlottenlund, Denmark*
[*]*rgani2018@gmail.com*

Abstract

Targeted process improvements can be achieved through process intensification and/or integration. Three alternative options, depending on the reference, are highlighted in the paper. The rules needed, to establish when process intensification and/or integration are feasible pathways to improvement, are given and through case studies, their successful applications are highlighted.

Keywords: Reactive distillation; Divided wall columns, Hybrid distillation schemes; Intensification; Optimization; Integration

1. Introduction

Integrated technologies can perform multiple tasks simultaneously within the same physical unit or sequentially in separate physical units with the objective to achieve a set of performance improvement targets. In the case of the former, a reactive distillation column (RDC) is a well-known example, while in the case of the latter, a divided wall column (DWC) and a hybrid distillation scheme (HDS) are well-known examples. In RDCs, both reaction and separation tasks are performed simultaneously on one or more stages of the column (Tian *et al*., 2018; Tula *et al*., 2020). In DWCs multicomponent separations that normally require multiple distillation columns are performed in a single column by placing a partial vertical partition (pre-fractionator) inside the main column (Asprion and Kaibel, 2010). In HDSs, two-unit operations, where one is distillation, are combined to perform specific separation tasks such that each operates at their highest efficiencies (O'Connell *et al*., 2019). Note that HDS is not membrane distillation.

Distillation is the common separation technique in RDC, DWC and/or HDS. However, separation by distillation is energy intensive and the majority of the separation operations found in the chemical and related industries are primarily distillation. Therefore, a large percentage of indirect emissions of carbon dioxide (CO_2) could also be attributed to distillation (Angelini *et al*., 2005). Accordingly, the challenge of the energy-environment interactions requires consideration of options to perform the same separation tasks at lower energy consumptions (Pistikopoulos *et al*., 2021). Opportunities exist for substantial reductions in CAPEX, OPEX, waste and CO_2 emission through simultaneous and/or sequential integration of operations.

Options to totally replace distillation as a separation technique by less energy intensive separation techniques have been proposed (Sholl and Lively, 2016), but their use depends on availability of, for example, materials that could serve as mass separating agents for membrane-based or for adsorption-desorption based separation techniques. Also, from a practical point of view, while it would be possible not to use distillation in future chemical processes, it would be practically impossible to replace the thousands of existing

distillation columns that are currently in operation all over the world. A more pragmatic alternative is to employ intensification and/or integration options based on the currently used separation techniques together with new separation techniques in hybrid schemes.

In this paper, feasibility of employing integrated technologies and related issues is highlighted. That is, in which chemical processes the reaction and separation operations can be intensified into an RDC, or which separation operations can be combined into DWC units, or which distillation-based separation operations can be retrofitted by an HDS? Note that in all cases the objective is to improve energy demand and related sustainability metrics without changing the product specifications and the feed mixture. The first two options indicate the replacement of an existing design with a more sustainable design, while the third refers to use of retrofit options without removing existing equipment.

Here, we present a rule-based, three-step method, which assists designers in deciding whether a more sustainable RDC can be configured to represent (and replace) a known chemical process; or, which separation operations can be combined into DWC; or, which distillation columns can be modified to HDSs; together with results from case studies highlighting different features of the three-step method. In particular, the results from the case studies highlight the potential to significantly improve the process in terms of sustainability metrics. More sustainable alternatives mean increased profit, lower CO_2 emission and sensitive environmental impacts. The objective is to find non-tradeoff solutions, if feasible, within the limits of the equipment parameters.

2. Method

Very briefly, the work-flow for the method has three main steps: 1) check using the rule-based method, if any of the three alternatives satisfy a set of feasibility criteria; 2) if the answer is yes, then apply the respective design methods for the feasible candidates; 3) apply target-based process improvement methods to obtain the final sustainable process design. A sample of the rules employed by the work-flow of step-1 is given in Table 1 for the three options. For step-2, stages 1 & 2 of the sustainable design method (Tula *et al*., 2017a) are applied, while for step-3, extended versions of stage-3 of the same method are applied. Therefore, the methods-tools for these steps are described here.

Table 1: Selected rules for process intensification and/or integration

Rules	Reactive Distillation	Divided Wall Column	Hybrid Distillation Scheme
1	Is the reaction exothermic?	Are there two or more distillation columns connected sequentially?	Are high purity products obtained as distillate?
2	Is there only one reaction?	Is it necessary to separate a multicomponent mixture?	Does the host column have 1 feed & 2 products?
3	Does the reaction and separation involve liquid and vapor phases?	Can the lightest & heaviest boiling products leave as top & bottom products in DWC?	Is the separation difficult (indicated by their driving force values)?
4	Does the separations involve azeotropes or close boiling pairs?	Is the number of high purity products less than or equal to 2 for a DWC?	Based on the cut-off equation (Eq 1), is the potential saving > 20%?
5	Does the separations involve 3 or more products?	Can the separations in the targeted columns be operated at similar pressures?	Is the feasibility of the other separation technique in HDS confirmed?
6	Are high purity products desired?	Does the targeted columns involve extraction and recovery operations?	The mixture to be separated does not involve azeotropes?
7	Are there energy intensive separations or waste issues?	Is the middle product flowrate low?	Are retrofit alternatives desired?

In step-1, based on the problem data and applying the rules, if the answer is yes to all the questions (for the RDC case), then the potential for generating a sustainable alternative employing RDC will exist. For DWC and HDS, additional rules and/or data are needed for rules 5-7. Step-2 involves the actual synthesis and design of the respective intensified-integrated alternative (Tula *et al*., 2017a). Step-3 verifies and further refines the solution from step-2, and ranks the alternatives in terms of chosen sustainability metrics based on techno-economic analysis and process simulation. See Tula et al. (2017a) for more details.

3. Case Studies

Results from three case studies involving RDC, DWC and HDS are presented. Detailed results for all case studies can be obtained from the authors.

3.1. Application of RDC (replacement and/or retrofit of base case)

Process intensification issues are highlighted through a new case study involving the production of methyl chloride (MeCl), which is produced from the reversible gas-phase reaction between methanol and hydrogen chloride, also producing water. Dehydration of methanol may occur giving di-ethyl ether and water. As in the production of methyl acetate (Agreda & Partin, 1984) or dioxalane (Castillo-Landero *et al*., 2018), this process can also be intensified to RDC because the answers to the rules in Table 1 are yes.

A base case process design exists for MeCl production (Dantus, 1999) and the goal is to produce the required amount of MeCl (90,000 metric tons/year), with a desired purity of MeCl (96-99.5 mol%). Design of the conventional process achieves these goals with higher operating and raw material costs. Figures 1 and 2 show two configurations generated by applying the established RDC synthesis and design methods (Sanchez-Daza *et al*., 2003). The main differences between the two configurations are: i) configuration-1 (Fig 1) considers a chiller to process the top product (MeCl) containing surplus reagent (HCl) for later separation using a conventional distillation column to recover the pure product and the surplus reagent; ii) configuration-2 (Fig 2) has a side product, which is sent to a membrane-based separation unit from which the product (MeCl) is separated from the unreacted HCl.

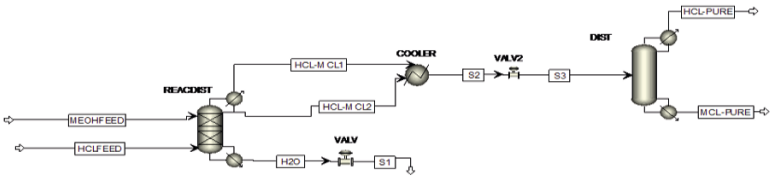


Figure 1: RDC configuration-1 for the production of methyl chloride

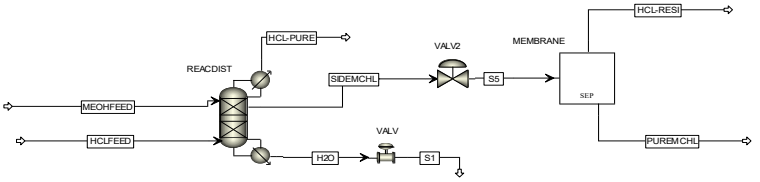


Figure 2: RDC configuration-2 for the production of methyl chloride

As the two reversible gas-phase reactions are exothermic, the heat produced by the reactions can be used to reach the boiling temperatures at each reactive stage. Therefore, the two RDC configurations require less energy than the base case process: configuration-1 requires -14.94 MW in the condenser and 14.44 MW in the reboiler, while, configuration-2 requires -15.53 MW in the condenser and 14.41 MW in the reboiler. The

furnace energy consumption used to pre-heat the reactants in the base case process is around 146.54 MW. The savings in cost of capital and operation are significant, even after taking into consideration the investiment for the extra unit operation. Note that as a retrofit option, one of the three existing distillation columns could be used as the RDC as the number of stages match. The reactive stages would be packed with catalyst. Another advantage of process intensification is that the RDC configurations achieve a higher purity product (99.9% MeCl), compared to the base case of 96% purity.

3.2 Application of DWC (replacement of base case)

The feasibility of the application of a DWC is highlighted considering a simpler version of the separation problem (Tamuzi *et al*., 2020), where a 6-compound hydrocarbon mixture consisting of C2, C3, iC4, nC4, C5+ is considered. There are 5 products, consisting of C2, C3, iC4, n-C4, C5+ and separation by distillation trains require a minimum of 4 columns corresponding to four separation tasks represented by splits of C2/C3, C3/i-C4, i-C4/n-C4 & n-C4/C5+. The answers to the rules from Table 1 are yes, and so any two of adjacent columns of the distillation train can be intensified into a DWC. Four possible configurations are feasible and the best is shown in Figure 3. All columns are designed by the reverse driving force based method in terms of sequence of the splits (separation task), the number of stages, feed location and product purities. Note that these are the minimum design variables that need to be specified to determine all other variables by simulation. Further reductions in energy could be achieved by adding separation units, as highlighted in Figure 3, in hybrid distillation schemes (described in section 3.3).

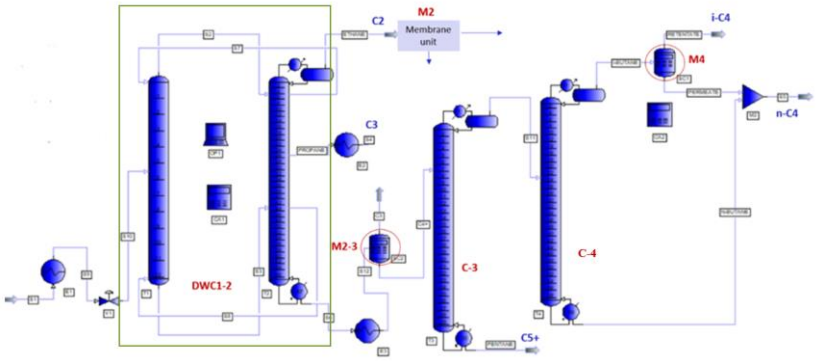


Figure 3: Flowsheet of a separation process with a DWC.

In Figure 3, the first two columns for the C2/C3 and C3/i-C4 splits are combined into one DWC. Membrane unit M2 purifies the distillate (top) product from DWC1-2 and M4 purifies the distillate top product from column C-4 by removing the high boiling compounds from their respective feed streams. Membrane unit M2-3 on the other hand, removes the low boiling compound from its feed stream. In all cases, the compound present in the smaller amount is removed, making the membrane unit feasible in terms of flux limitations and membrane area needed for the separation. It is able to reduce the energy consumption by around 45.32% compared to the optimized 4-column configuration. Other DWC combinations, such as two DWC (DWC1-2 and DWC 3-4; C1, DWC2-3, C4; C1, C2, DWC3-4, with or without membrane units, give lower reductions in energy consumptions.

3.3. Application of HDS (retrofit of base case)

In principle, any distillation column (also RDC and DWC) can be considered for HDS if the rules given in Table 1 give “yes” answers. Then, for the identified column (separation

task), the work-flow involves selection of the operating pressure to determine the driving force and its maximum location for the binary pair of compounds defined as a key compound for the separation task. Next, for a column with fixed number of stages and feed location, simulation-based data is generated to obtain plots of reboiler duty versus distillate product purity as a function of driving force (O'Connell *et al*., 2019). Next, the $Q_{cut\text{-}off}$ (the reboiler duty corresponding to a lower product purity) and $Q_{desired}$ (reboiler duty corresponding to the target high purity product) are located on the reboiler duty versus distillate purity plot and used to determine the likely energy savings, $Q_{savings}$:

$$Q_{savings} = \left(\frac{Q_{desired} - Q_{cut-off}}{Q_{desired}}\right) * 100 \quad (1)$$

If this value is > 20%, an HDS would be feasible, provided a separation technique for the further purification separation task can be found. For example, find a suitable membrane for a membrane-based separation. The synthesis-design of an HDS represents step-3 of the 3-step method. A typical HDS problem could be defined as:

Given - a mixture to be separated into two or more almost pure compounds; *Required* - design an appropriate hybrid distillation scheme that can deliver the specified products at lower energy consumption than a single distillation operation.

The work-flow for the above problem solution is given by O'Connell *et al*. (2019) and available in the ProCAFD software (Tula *et al*., 2017b), which provides options for fast and reliable separation process design, including hybrid separation schemes for azeotropic as well as non-azeotropic mixtures. In this paper, some of the applicability issues are highlighted. Table 2 lists different distillation columns that have been retrofitted (based on their reported designs) to more sustainable HDS and verified through steady state simulation with AVEVA PRO/II (AVEVA, 2021).

Table 2: Hybrid distillation schemes reported in published papers

Problem	Mixture compounds	Energy saved	Reference
1	i-butane + n-butane*	44.4%	14
2	i-pentane + n-pentane*	37.6%	14
3	Methanol + water*	34.5%	14
4	2,2 dimethyl butane + n-hexane	25-45%	4
5	Styrene + ethylbenzene	25-45%	4
6	Benzene + cyclohexane*	42.5%	15
7	Acetonitrile + water	38.5	15

* One separation task in a multicomponent separation process or a multi-operation process

As an example, consider problem 1 in Table 2 (also, column C-4 in Fig 3). Using the data reported by Tula *et al*. (2017b), the available driving force for i-butane and n-butane is computed to be 0.095. From the corresponding reboiler duty versus distillate product purity, potential energy savings is calculated to be 44% (with $Q_{desired}$ as 32.33 GJ/hr and Q_{cutoff} as 17.98 GJ/hr). The distillate cut-off composition is 85.0% and the specified distillate purity is selected as 99.5%. The HDS for this mixture is shown in Fig 3 (column C-4). A membrane with permeabilities of 0.2196 kmole/(m^2 hr) for n-butane and 0.0008 kmole/(m^2 hr) for i-butane is available in the database. This gives a required membrane area of 109.24 m^2. Using a membrane cost of 2500 Euro/m^2, the additional CAPEX is calculated to be 273,100 Euro. Savings in energy costs is calculated to 114.1 Euro/hr, giving a payback time of 0.27 years. Details of this solution can be found in Tula *et al*. (2017b). Considering driving force values ranging from 0.085 to 0.095, with the lower value indicating actual column data, a range of energy savings data can be generated.

4. Conclusion

The intensification and integration options considered in this paper confirm their potential in terms of tackling the energy-environment nexus. The large number of energy intensive chemical processes and within them, distillation operations, should be targeted for improvement in terms of energy and environment where the additional investment could be recovered from the additional profit, without negatively impacting the environment. It should be noted that while the operations of some chemical processes are energy intensive, others are sensitive to environmental impacts but not energy intensive. Therefore, targeted improvements must address both these issues. The proposed rules and the three steps method are able to quickly identify the candidates for improvement, and opportunities for new advances in methods and associated tools to tackle the challenges. In this way, they add an intelligence feature to the current design methods. Current and future work is expanding the rules-based feasibility analysis as well as creating a database of solved problems that can serve as repository of knowledge and data to share with the community. The rules also need to be updated to avoid exceptions.

References

V. H. Agreda, L. R. Partin, 1984, Reactive distillation process for the production of methyl acetate. US Patent 4435595.

P. Angelini, 2005, Materials for Separation Technologies: Energy and Emission Reduction Opportunities, Industrial Technologies Program, DE-AC05-00OR22725. US DOE.

N. Asprion, G. Kaibel, 2010, Dividing wall columns: fundamentals and recent advances, *Chemical Engineering and Processing: Process Intensification,* 49, 139-146.

AVEVA, 2021, PRO/II User Manual.

A. Castillo-Landero, A. Jiménez-Gutiérrez, R. Gani, 2018, Intensification methodology to minimize the number of pieces of equipment and its application to a process to produce dioxolane products, *Industrial & Engineering Chemistry Research*, 57, 9810-9820.

M. Dantus, 1999, PhD Dissertation. Department of Chemical Engineering. Oklahoma State University, Stillwater, Oklahoma.

J. P. O'Connell, M. R. Eden, A. K. Tula, R. Gani, 2019, Retrofitting Distillation Columns with Membranes, *Chemical Engineering Progress*, 115; 41-49

E. N. Pistikopoulos, A. Barbosa-Povoa, J. H. Lee, R. Misener, A. Mitsos, G. V. Reklaitis, V. Venkatasubramanian, F. You, R. Gani, 2021, Process Systems Engineering–The Generation Next? *Computers & Chemical Engineering*, 147, 107252

O. Sánchez-Daza, E. S. Pérez-Cisneros, E. Bek-Pedersen, R. Gani, 2003, Graphical and stage-to-Stage methods for reactive distillation column design, *AIChE J.*, 49, 2822–2841.

D. S. Sholl, R. P. Lively, 2016, Seven chemical separations to change the world, *Nature*, 532, 435.

A. Tamuzi, N. Kasiri, A. Khalili-Garakani, 2020, A. Design and optimization of distillation column sequencing for NGL fractionation processes, *Journal of Natural Gas Science & Engineering*, 76, 103180.

Y. Tian, S. E. Demirel, M. M. F. Hasan, E. N. Pistikopoulos, 2018, An overview of process systems engineering approaches for process intensification: State of the art, *Chemical Engineering and Processing-Process Intensification,* 133, 160-210.

A. K. Tula, D. K. Babi, J. Bottlaender, M. R. Eden, R. Gani, 2017a, A computer-aided tool for sustainable process synthesis-intensification, *Comput & Chem Eng*. 105, 74-95.

A. K. Tula, B. Befort, N. Garg, K. V. Camarda, R. Gani, 2017b, Sustainable process design & analysis of hybrid separations, *Computers & Chemical Engineering*, 105, 96-104.

A. K. Tula, M. R. Eden, R. Gani, 2020. Computer-aided process intensification: Challenges, trends and opportunities. *AIChE Journal*, 66, e16819

R. Vooradi, S. P. Venkata, A. K. Tula, B. A. Sarath, M. R. Eden, R. Gani, 2018, Hybrid Separation Scheme for Azeotropic Mixtures–Sustainable Design Methodology*, Chemical Engineering Transactions*, 69, 637- 42.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50120-2

Integration of Design and Operation for the CO_2-based Methanol Synthesis

Christian Hoffmann[a*], Erik Esche[a], Jens-Uwe Repke[a]

[a]*Process Dynamics and Operations Group, Technische Universität Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany*

c.hoffmann@tu-berlin.de

Abstract

Conventionally, many approaches for process design are based on steady-state assumptions. Only in the last two decades, dynamic aspects have also been considered, but frameworks for large-scale systems are still scarce. Recently, we introduced our framework of a fully discretized dynamic model in combination with an optimal economic NMPC. In this contribution, we apply this framework on the CO_2-based methanol synthesis, which represents a large-scale case study of 50,000 continuous variables for a time horizon of around 1 h. Feed variability is modeled by an amplitude-modulated pseudo-random binary sequence to gain as much information on the dynamics as possible.

Keywords: Integration of Design and Operation, Methanol Synthesis, Economic NMPC.

1. Introduction

Process design is a fundamental aspect of process systems engineering. The design must ensure that all requirements regarding product purities are met. Additional constraints for process parameters, e.g., due to safety concerns, must hold as well. Conventionally, process design and control design are performed sequentially or as an iterative process (Seider et al., 2017). However, this approach may lead to dynamic constraint violations, hinder robust performance, or simply reduce the controllability of a process during plant operation (Malcolm et al., 2007). Therefore, performing these two subsequent tasks simultaneously by integration of design and operation (IDO) is highly desirable. The most common approaches have been summarized in several reviews, for example, by Vega et al. (2014). Previously, we suggested using the objective function of an economic nonlinear model-predictive controller (eNMPC) to account for the connection of economics and variability (Hoffmann et al., 2019). We also studied how parametric uncertainty can be included by using the Unscented Transform (Hoffmann et al., 2020). In this contribution, we apply our framework on a considerably larger example: the CO2-based methanol synthesis. In the next section, we describe our IDO-eNMPC framework. Afterward, we outline the dynamic process model of the case study. Finally, we present the results obtained with our framework and analyze them regarding their merit.

2. Framework

The mathematical structure of the IDO framework is given in Equation (1): An objective function that includes both investment and operating costs is minimized subject to equal-

ity and inequality constraints. The constraints are functions of states x, controls u, input variabilities ν, parameters θ, and design variables d (for example, the volume of equipment). In addition, constraints determining the initial states (subscript 0) and terminal states (subscript t) are necessary. In the framework, all equations are fully discretized using orthogonal collocation on finite elements. The indices represent the collocation point c_p and the finite element fe. States and controls must also lie within a certain domain $\mathcal{X}$ and $\mathcal{U}$, respectively:

$$\min_{u,\,d} \quad f\left(x_{cp,fe}, u_{cp,fe}, d\right) \tag{1a}$$

$$\text{s.t.} \quad 0 = g\left(x_{cp,fe}, u_{cp,fe}, \nu_{cp,fe}, \theta, d\right) \quad \text{(Dynamic process model)}, \tag{1b}$$

$$0 \leq h\left(x_{cp,fe}, u_{cp,fe}, \nu_{cp,fe}, \theta, d\right) \quad \text{(Path constraints)}, \tag{1c}$$

$$0 \leq h_0\left(x_0, u_0, \nu_0, \theta, d\right) \quad \text{(Initial constraints)}, \tag{1d}$$

$$0 \leq h_{\mathrm{t}}\left(x_{\mathrm{t}}, u_{\mathrm{t}}, \nu_{\mathrm{t}}, \theta, d\right) \quad \text{(Initial constraints)}, \tag{1e}$$

$$x_{cp,fe} \in \mathcal{X},\ u_{cp,fe} \in \mathcal{U} \quad \text{(Domain)} \tag{1f}$$

The input variability is represented by amplitude-modulated pseudo-random binary sequences (APRBS), a signal type that allows for high information gain and that is suitable for nonlinear systems (Nelles, 2001).

3. Model

The dynamic process model is based on the flowsheet in Figure 1. The feed gas (CO_2 and H_2) is compressed, potentially heated to reactor inlet temperature, and passes the fixed-bed reactor to form methanol and water. The products are condensed whereas the gases are re-compressed and recycled to the reactor. In the tank, liquid methanol and water are stored and the product stream is withdrawn.

3.1. Balance equations

The model consists of a two-dimensional (time and axial) model of reactor with dynamic mole, energy, and momentum balances (only the mole balance is shown here due to space limitations):

$$\varepsilon_{\mathrm{cat}} \cdot \frac{\partial c_c}{\partial t} = -c_c \frac{\partial w}{\partial z} + w \frac{\partial c_c}{\partial z} + \left(1 - \varepsilon_{\mathrm{cat}}\right) \cdot \rho_{\mathrm{cat}} \cdot \sum_r \nu_{c,r} r_r \tag{2}$$

Therein, $\varepsilon_{\mathrm{cat}}$ is bed porosity, c_c is molar concentration of component c, w is velocity, ρ_{cat} is catalyst density, $\nu_{c,r}$ and r_r are stoichiometric coefficient and reaction rate, respectively. Here, the reaction kinetics by Slotboom et al. (2020) are used. The tank consists of dynamic mole balances. Heat exchangers and compressors are modeled under steady-state assumptions.

3.2. Cost functions

The cost of the equipment is determined from cost functions with degression exponent n:

$$C = C_{\mathrm{ref}} \left(\frac{cap}{cap_{\mathrm{ref}}} \right)^n. \tag{3}$$

Therein, C and C_{ref} are costs at capacity cap and reference capacity cap_{ref}, respectively. Operating costs are computed by multiplying the electricity or product costs with the integral mean of the respective material or energy stream.

3.3. Initial conditions

During the model formulation and first attempts to solve the system, it became clear that the large number of unfixed initial conditions in a spatially distributed system, i. e., the reactor, causes an additional challenge. If these are not assigned consistently by any method, the optimizer may choose huge initial concentrations of the product and thus artificially maximize the product yield. This cannot be resolved by simply bounding the initial conditions as they should also represent are realistic concentration, temperature, and velocity profiles. Also, fixing the initial conditions to an arbitrary value is impossible because they are inherently related to the design, which will change during the iterations. For this reason, a steady-state version of the model is also incorporated into the model, which is solved simultaneously with the dynamic model. The steady-state sub-model provides the initial conditions for the dynamic case, but it will also still depend on the same design variables.

3.4. Path constraints

Several path constraints are required to ensure a stable behavior: First, the temperature must always remain below 260 °C to avoid catalyst deactivation – this temperature is higher in reality; the value was chosen to achieve an active bound at the optimal solution. Second, stability must be ensured so that concentration profiles do not drift over time. This is achieved by demanding that the terminal states match the initial condition within a certain threshold. This is not required for every state of the model, but for key variables, such as the inlet pressure and temperature of the reactor, and the holdup of the tank, for example:

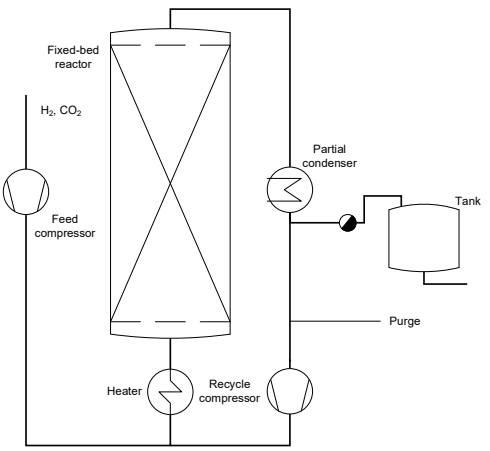


Figure 1: Flowsheet for the case study.

$$|HU_{\mathrm{tank}}(t=0) - HU_{\mathrm{tank}}(t=t_{\mathrm{t}})| \leq \varepsilon_{\mathrm{tank}} \tag{4}$$

This ensures the stability of the obtained trajectories under the respective input variability. In addition, the integral average of the produced flow should be close to the nominal product flow to achieve production goals:

$$\left| \frac{1}{t_{\mathrm{t}}} \int_{t=0}^{t=t_{\mathrm{t}}} F_{\mathrm{prod}}(t)\,\mathrm{d}t - F_{\mathrm{prod,st-st}} \right| \leq \varepsilon_{\mathrm{prod}} \tag{5}$$

The integral is evaluated with Radau quadrature using the same Radau collocation points as in the discretization scheme. In addition, the change between consecutive values of the same variable can be restricted, for example, the reactor's feed pressure:

$$|P_{\text{reactor}}(t = t_k) - P_{\text{reactor}}(t = t_{k+1})| \leq \varepsilon_{\text{pressure}} \tag{6}$$

3.5. Implementation and framework

The resulting partial differential equation system is discretized via orthogonal collocation in both the spatial coordinate of the reactor and the time coordinate. 20 spatial finite elements were chosen in preliminary trials. The number of finite elements for the time coordinate can be chosen freely. The system roughly scales to 50,000 variables per 10 finite elements in time. The model is formulated within AMPL (Fourer et al., 2003) and interfaced to Python via AMPLPy. AMPL is chosen because it allows the use of sets for model generation, includes a powerful tool for automatic differentiation to generate derivatives, and disposes of interfaces to many solvers. The external layer in Python provides the data (parameter and variable values), performs the initialization, manages loops and saves the results.

4. Results and discussion

In this contribution, both CO_2 and H_2 feed are assumed to be variable inputs $\nu_{cp,fe}$ that are represented by two APRBS, which means that their values at each collocation point are given by the APRBS. Both are assumed to have a standard deviation of the variability range of 20 mol s^{-1}. The APRBS are generated by our own sampling framework (Weigert et al., 2020).

4.1. Initialization

Even if the model is initialized at steady-state, an initialization strategy is required to solve Problem (1). For this purpose, an homotopy approach is taken so that the current input depends on an homotopy parameter φ:

$$F_{st=1,c} = \varphi \cdot F_{st=1,c}^{\text{APRBS}} + (1 - \varphi) \cdot F_{st=1,c}^{\text{nominal}} \tag{7}$$

Hence, Problem (1) is solved repeatedly with increasing φ from 0.1 to 1. In our experience, an increase of 0.1 per step suffices to converge the system. Should this not be the case, the step size can be made smaller.

4.2. Comparison of steady-state and dynamic results

The results obtained with the framework are compared to results obtained under steady-state conditions only. Exemplary numerical values for design decision are given in Table 1. In both cases, the tube diameter lies at the lower bound. The reactor inlet pressure is lower in the dynamically obtained design (lower costs for feed compression) whereas the inlet temperature is higher (faster kinetics) that are necessary in cases where the stoichiometry might no be ideal occasionally. The reactor length is significantly shorter in the dynamic case, which is compensated by higher recycle ratios and thus results in a larger energy

Table 1: Comparison of design based on steady-state and dynamic consideration. Results under dynamic conditions are average values or are given in ranges. Lower (LB) and upper bounds (UB) during the optimization are given if applicable.

Design variable / operating condition	LB	UB	Steady-state	Dynamic
Average feed compressor power, MW	0	–	12.7	11.1
Average reactor inlet pressure, bar	30	100	60	50
Reactor inlet temperature, K	273	–	320	384 – 426
Reactor length, m	0	–	5.3	2.6
Tube diameter, m	0.05	–	0.05	0.05
Number of tubes	10	–	14,400	25,600
Condenser area, m^2	0	–	1740	1200
Recycle compressor, MW	0	–	2.8	4.2
Recycle ratio	1.0	–	4.4	2.6 – 8.6

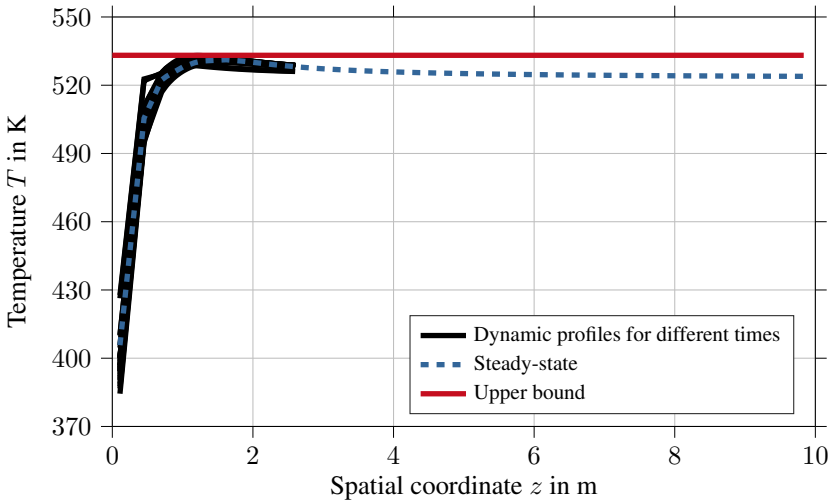


Figure 2: Comparison of axial temperature profiles at different times and the steady-state temperature profile.

consumption of the recycle compressor. Figure 2 shows a comparison between the steady-state temperature profile (long reactor) and the dynamic, axial temperatures for different times. For both profiles, the upper bound is never violated. In the case of the dynamic optimization, this is, however, only possible because the design has notably changed.

4.3. Computational times

The computations were performed on an Intel® Core™ i7-4770K CPU @ 3.50GHz running Ubuntu 18.04.6. AMPL version 20210326 in combination with IPOPT 3.12.6 was used to formulate and solve the optimization problems. Because a steady-state solution and a

dynamic solution are looked at here, it is not a fair comparison in terms of the required computational time. However, the reader may get an impression of how the computational time scales with the system's size. The steady-state solution is usually obtained within a few seconds. This may vary with the quality of the initialization. The dynamic solution, which is always initialized at the steady-state solution, may require several days of computations. However, this is still deemed acceptable for design problems, which typical do not require the rapid solution that is demanded from online applications, such as model-predictive control.

5. Conclusion and outlook

This contribution applied a methodology for the integration of design and operation on a large-scale system. The results were obtained by full discretization of the dynamic process model with subsequent solution of the resulting nonlinear programming problem. We compared the obtained results with those obtained by a steady-state optimization and found significant differences in the design, e.g., the reactor length, which would justify to consider the dynamics during process design. However, the length of the time horizon might have significant impact on the results. Therefore, this must be investigated in the future.

References

R. Fourer, D. M. Gay, B. W. Kernighan, 2003, AMPL : a modeling language for mathematical programming, Thomson/Brooks/Cole, Pacific Grove, CA.

C. Hoffmann, E. Esche, J.-U. Repke, 2019, Integration of design and control based on large-scale NLP formulations and an optimal economic NMPC, Computer Aided Chemical Engineering, 47, 125–130.

C. Hoffmann, J. Weigert, E. Esche, J.-U. Repke, 2020, Integration of Design and Operation Using Dynamic Perturbation and Chance Constraints with Unscented Transform, Computer Aided Chemical Engineering, 48, 751–756.

A. Malcolm, J. Polan, L. Zhang, B. A. Ogunnaike, A. A. Linninger, 2007, Integrating systems design and control using dynamic flexibility analysis, AIChE Journal, 53, 2048–2061.

O. Nelles, 2001, Nonlinear system identification : from classical approaches to neural networks and fuzzy models, Springer-Verlag Berlin Heidelberg, Berlin, Heidelberg.

Y. Slotboom, M. J. Bos, J. Pieper, V. Vrieswijk, B. Likozar, S. R. A. Kersten, D. W. F. Brilman, 2020, Critical assessment of steady-state kinetic models for the synthesis of methanol over an industrial Cu/ZnO/Al2O3 catalyst, Chemical Engineering Journal, 389, 124181.

P. Vega, R. L. de Rocco, S. Revollar, M. Francisco, 2014, Integrated design and control of chemical processes – Part I: Revision and classification, Computers & Chemical Engineering, 71, 602–617.

J. Weigert, C. Hoffmann, E. Esche, J.-U. Repke, 2020, Enabling Dynamic Real-Time Optimization under Uncertainty using Data-Driven Chance Constraints, Computer Aided Chemical Engineering, 48. Elsevier, 1189–1194.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50121-4

Blue Syngas Synthesis via the Integration of Gasification and Reforming Processes

Hussain A. Alibrahim[a], Siddig S. Khalafalla[a], Usama Ahmed[a,b] and Umer Zahid[a,c*]

[a] *Chemical Engineering Department, King Fahd University of Petroleum & Minerals (KFUPM), Dhahran 31261, Saudi Arabia*
[b] *Interdisciplinary Research Center for Hydrogen and Energy Storage, King Fahd University of Petroleum & Minerals (KFUPM), Dhahran 31261, Saudi Arabia*
[c] *Interdisciplinary Research Center for Membranes & Water Security, King Fahd University of Petroleum & Minerals (KFUPM), Dhahran 31261, Saudi Arabia*
uzahid@kfupm.edu.sa

Abstract

Syngas is an important intermediate feedstock to produce various downstream chemicals and clean fuels. In this study, two process models are developed by integrating the gasification and dry-reforming models in the parallel and series configuration to produce the syngas at the rate of 10,000 kmol/h with H_2/CO ratio of 2. The heat integration is also developed in a way to utilize the heat energy from the coal-derived syngas into the dry-reformer without any energy penalties. The proposed integrated designs can enable the utilization of fossil fuels in an environment friendly, technically feasible and an economical way.

Keywords: Integration; carbon capture and utilization; simulation; syngas.

1. Introduction

Synthesis gas commonly known as “syngas” is an important component for the synthesis of chemicals and fuels. It drives many industrial facilities including power generation, fertilizers, polymers, and production of bulk chemicals. Depending on the feedstock, oxidizing agent, desired syngas ratio and downstream application, multiple syngas production technologies exist. Gasification and reforming are the most commonly employed commercial technologies for the syngas production (Medrano-García et al., 2018). Generally, gasification is used for the solid feed stocks such as coal, coke, and biomass while reforming is the preferred process for the gaseous and liquid feed stocks. While coal is still the primary source of energy for many countries, coal based processes are usually associated with high carbon emissions (Alibrahim et al., 2019).

Reforming technologies may employ steam (in steam reforming), oxygen (partial oxidation reforming) or carbon dioxide (dry reforming) as the oxidant depending on the type of reforming process. Steam reforming (SR) and dry reforming (DR) are endothermic processes while partial oxidation (POX) of hydrocarbon is an exothermic process. Recently, a great attention has been given, particularly in the process and catalyst development, to the DR processes (Alenazey et al., 2021; Alibrahim et al., 2021). This is due to the fact that the DR consumes two major greenhouse gases in which methane or other higher hydrocarbons along with carbon dioxide can be

converted to the syngas (Afzal et al., 2018). However, the commercialization of DR based processes have been hindered due to certain process limitations. The main problematic aspect of the DR technology is the deactivation of active catalyst surface due to carbon deposition at high temperature, along with high energy demand for the endothermic reactive system, low syngas ratio and equilibrium limitation due to the influence of reverse water gas shift reaction (RWGS). Elbashir et al. (Elbashir et al., 2018) reported that the operational limitations of DR process can be overcome by combining the DR with other reforming process such as SR and POX. Similarly, Man et al. (Man et al., 2014) studied the integration of SR, DR and gasification processes by utilizing the coke-oven gas for the coal-to-olefins (CTO) process. Their results showed that the coke-oven gas assisted process with the coal gasification improves the process efficiency by 10% compared to the conventional CTO process. Several studies in the past have focused on the integration of various reforming and gasification technologies in order to combine the synergies and suppress the imperfection of one technology to another (Balasubramanian et al., 2018; Carapellucci and Giordano, 2020; Summa et al., 2019; Wang et al., 2020). Qian et al. (Qian et al., 2015) evaluated the utilization of CO_2 supply from WGS reactor in the tri-reforming for the coal-to-methanol (CTM) process. They reported a decrease in the carbon emissions by 44% and an improvement of carbon utilization by 25% compared to the conventional CTM process. Chen et at. (Chen et al., 2019) investigated the integration of conventional coal gasification and coke-gasification processes to produce methanol product. They showed the integrated design to have 14.3% higher carbon utilization efficiency and 34.6% reduced carbon emissions compared to the conventional CTM process.

The goal of this work is to develop the conceptual design based on the process intensification techniques that can produce the syngas of varied composition by the integration of gasification and DR processes. The combination of an endothermic and exothermic processes for energy-effective utilization offers lower exergy destruction. The proposed designs offer an improved performance in terms of energy requirement, carbon emissions and cost due to the process intensification and synergies coupling. First, standalone simulation models are developed for the coal gasification and DR respectively to assess the performance in terms of efficiency and syngas quality. Then, two integrated process designs are proposed with series and parallel integration between gasification and DR process. In order to have a fair comparison among various designs, the syngas production capacity is set as 10,000 kmol/hr with a syngas ratio (H_2:CO) of 2 which is suitable for many downstream applications. Finally, a detailed techno-economic analysis has been performed in order to ascertain the performance and economic feasibility of the proposed designs.

2. Process description

2.1. Series integration

In the case of gasification and DR series integration, coal preparation and gasification units are similar to that of the standalone gasification process. However, the gas cleaning unit only contains H_2S removal system while the CO_2 removal unit is employed in the WGS unit after the shift reactors as shown in the figure 1. An important aspect of this design is the heat integration between the gasification and the reforming reactor. All the heat duty required by the DR reactor has been provided from the gasification reactor, hence making the net heat duty of the DR reactor zero. The DR unit is placed downstream of H_2S removal unit where it receives a bulk of gas mixture

mainly containing syngas and CO_2. Given that H_2/CO ratio is below unity after coal gasification, high water-gas shift rate is expected. The CO_2 captured from the WGS section is sent back to the DR section as a feed to react with freshly fed methane. The syngas ratio at the exit of dry reforming process is 0.822 which is low compared to the target syngas ratio of 2. Therefore, a part of raw syngas stream is sent to the WGS unit where 99.6 % CO conversion is achieved. Rectisol process is used to capture the 99.7 mol. % CO_2 while the high purity H_2 is blended with the bypassed syngas to achieve the desired product syngas.

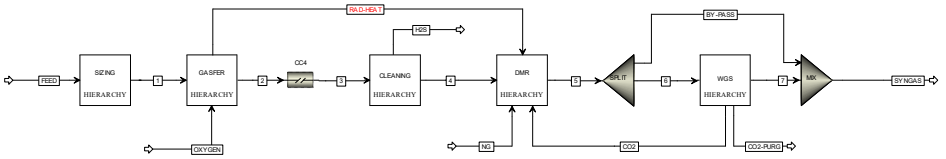


Figure 1. Series design integration of gasification and dry reforming for syngas production

2.2. Parallel integration

In the case of parallel integrated design, the coal preparation and gasification sections are the same as that of the standalone gasification process. The gas cleaning unit has H_2S removal and CO_2 removal sections employing the Rectisol process. The captured high purity (99.7 mol.%) CO_2 stream is sent to the dry reforming unit which is in parallel configuration with the gasification process. The raw syngas from the outlet of gas cleaning unit has a H_2 to CO ratio of 0.76 which is lower than the required syngas ratio of 2. Therefore, a part of the raw syngas is directed to the WGS unit in order to convert CO to H_2, while the remaining syngas bypasses the WGS unit. Design Spec function in Aspen Plus has been utilized to vary the split fraction in order to achieve the desired product syngas ratio. Since the syngas coming from the gasification and dry reforming sections is deficient in H_2 content, the Design Spec function calculates the amount of syngas that should be directed to the WGS, while the remaining syngas is bypassed. The second CO_2 capture unit is installed in the WGS unit to separate the produced CO_2 from the CO which is sent back to the DR unit. The bypassed syngas, syngas from the DR and high purity H_2 from the WGS are then mixed to produce the desired product syngas. Similar to the series integrated design, the parallel integration provides all the required DR heat duty from the gasification section as shown in the figure 2.

In this study, two cases (series and parallel) are demonstrated for the syngas production rate of 10,000 kmol/h with a H_2/CO ratio of 2. However, the proposed designs are flexible in terms of changing the quantity and/or quality of the product syngas. The feed coal and natural gas flowrates can be adjusted for a desired production rate of syngas in a way that the reformer net duty remains zero. Similarly, the quality of the syngas can be adjusted by changing the split fraction upstream of the WGS unit. However, it should be noted that any such change will affect the overall carbon utilization and emissions from the process.

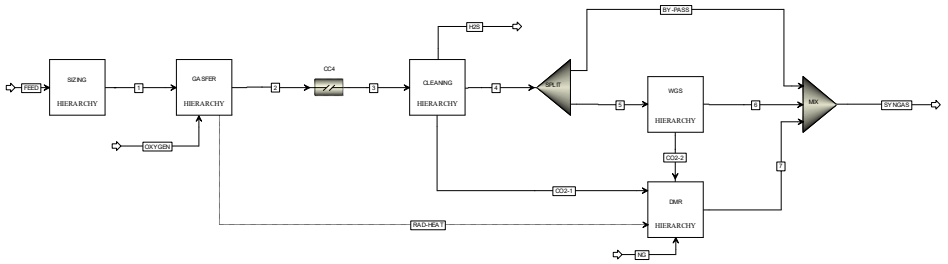


Figure 2. Parallel design integration of gasification and dry reforming for syngas production

3. Results

3.1. Energy analysis

Figure 3 shows the total energy requirement in terms of utilities for the four cases for the syngas production of 10,000 kmol/h with H_2 to CO ratio of 2. The results show that in the case of standalone DR, more than 62.8 % of the total energy requirement is in the form of heating. In the case of standalone gasification, both the heating and cooling requirements contributes significantly to the total energy demand. On the other hand, approximately 43.25 MW of net power is generated from the steam cycle. The electricity demand for the standalone DR process is the highest among all the designs. Since, the delivery pressure of the product syngas is set at 28 bar, the syngas produced from the DR process at 4 bar needs to be compressed to 28 bar requiring a three-stage compression train. The electrical energy consumed by the 3-stage compressor is 28.2 MW which is significantly higher than the pumping requirement of the gasification process. The integration of coal gasification and DR shows considerable reduction in the total energy input as shown in figure 3. The results show that the integrated designs reduce the heating requirement by 52.4% and 61.8% for the series and parallel designs respectively compared to the standalone DR. Figure 4 shows the breakdown of the total energy requirement for each section of the four cases. The reformer is the largest energy consumer in the case of standalone DR process requiring around 46.5% of the total energy. By performing an integration between the gasification and DR, the reformer duty is completely removed by supplying the required heat from the gasification. The net heat supplied from the gasification reactor in the series and parallel integration is equivalent to 82.16 MW and 81.92 MW respectively. Since, all the cases are designed for the production of an equimolar flowrate, the amount of feeds in the integrated design is considerably reduced compared to the standalone gasification case. The result reveals that an integrated design requires 45.2% and 54% less coal and natural gas feed rates respectively, compared to the standalone gasification and DR process.

3.2. Economic analysis

The analyses performed in the previous section clearly shows the technical feasibility for the design integration between the gasification and DR process. The results revealed a positive impact on the overall process performance in terms of carbon emissions and energy requirements for the series and parallel integrations. This section explores the economic viability of the integrated designs compared with the standalone gasification and DR process. A detailed economic analysis has been performed to calculate the fixed

capital and operating costs. Aspen Economic Analyzer has been used to perform the sizing and capital cost estimation of the equipment.

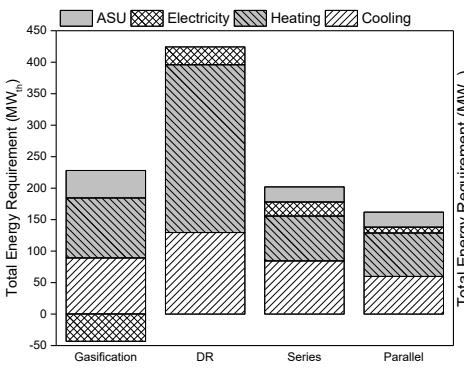


Figure 3. Utility-wise energy consumption for the standalone and integrated designs

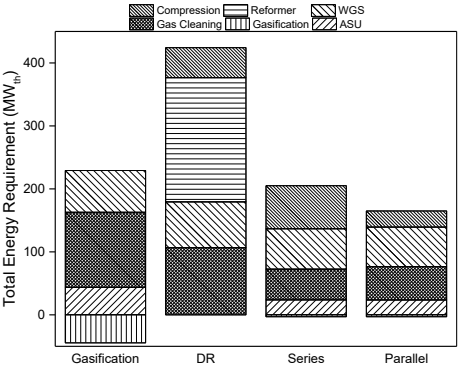


Figure 4. Section wise energy consumption for the standalone and integrated designs

The result shows that the series design has the highest total capital cost while the parallel design offers the lowest capital cost. The high capital cost of series design is due to the series integration of the gasification and DR processes which resulted in the large equipment size. The results also reveal that the direct cost is in the range of 55 – 63% of the total capital cost. The operating cost calculation includes the raw materials, utilities, labor and supervisory costs, maintenance and general and administrative costs. The integrated series design offers the TAC of 100.4 M$ which is 16% higher than the integrated parallel design cost. The economic analysis shows that the series design is cost competitive with the standalone DR design. Since all the designs produce the same amount of syngas, per unit cost of syngas for the integrated designs is lower compared to the standalone designs as shown in figure 5.

4. Conclusions

In this study, two integrated models are proposed to produce 10,000 kmol/h of syngas with a H_2/CO ratio of 2 delivered at a pressure of 28 bar. The integrated designs eliminate the energy requirement of dry reformer by supplying the heat energy from the gasifier. The total energy requirement for series and parallel design is 202 and 162 MW respectively, compared to 424.3 MW energy requirement for the standalone DR process. The low energy requirement by the integrated designs is reflected in terms of high process efficiency of 83.63 % and 86.66 % for the series and parallel designs, respectively. The economic analysis shows that the raw materials cost has the highest share in the TAC for all the designs. Parallel design requires less CAPEX and OPEX of 17.7 M$ and 12.2 M$/y respectively compared to the series design. Among all the designs (standalone and integrated), parallel offers the lowest TAC of 86.4 M$/y which translates into unit product cost of $ 0.99.

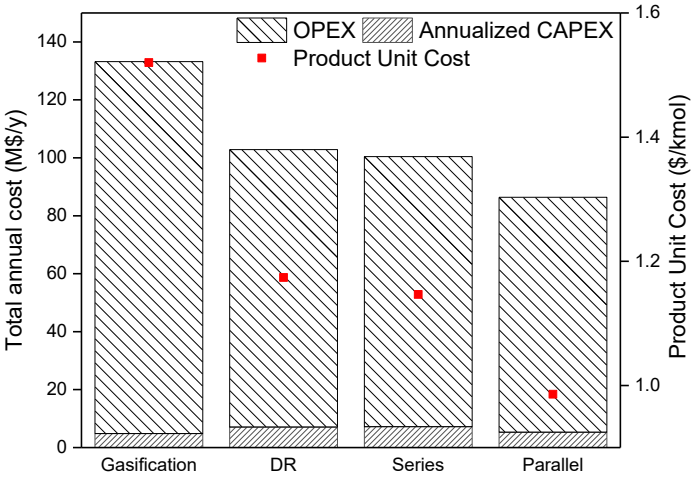


Figure 5. TAC and product cost for the standalone and integrated designs

References

Afzal, S., Sengupta, D., Sarkar, A., El-Halwagi, M., Elbashir, N., 2018. Optimization Approach to the Reduction of CO2 Emissions for Syngas Production Involving Dry Reforming. ACS Sustain. Chem. Eng. 6, 7532–7544.

Alenazey, F., AlOtaibi, B., Otaibi, R.A.L., Alyousef, Y., Alqahtania, S., Qazaq, A., Zahid, U., Vo, D.-V.N., Adesina, A., 2021. A Novel Carbon-Resistant Perovskite Catalyst for Hydrogen Production Using Methane Dry Reforming. Top. Catal. 1–9.

Alibrahim, H.A., Khalafalla, S.S., Ahmed, U., Park, S., Lee, C.-J., Zahid, U., 2021. Conceptual design of syngas production by the integration of gasification and dry-reforming technologies with CO2 capture and utilization. Energy Convers. Manag. 244, 114485.

Alibrahim, H.A., SeedAhmed, S., Ahmed, U., Zahid, U., 2019. Comparative analysis of gasification and reforming technologies for the syngas production, in: Computer Aided Chemical Engineering. Elsevier, pp. 1759–1764.

Balasubramanian, P., Bajaj, I., Hasan, M.M.F., 2018. Simulation and optimization of reforming reactors for carbon dioxide utilization using both rigorous and reduced models. J. CO2 Util. 23, 80–104.

Carapellucci, R., Giordano, L., 2020. Steam, dry and autothermal methane reforming for hydrogen production: A thermodynamic equilibrium analysis. J. Power Sources 469, 228391.

Chen, J., Yang, S., Qian, Y., 2019. A novel path for carbon-rich resource utilization with lower emission and higher efficiency: An integrated process of coal gasification and coking to methanol production. Energy 177, 304–318.

Elbashir, N.O., Challiwala, M.S., Sengupta, D., El-Halwagi, M.M., 2018. System and method for carbon and syngas production. World Intellect. Prop. Organ. Prod. Syst. Method Carbon Syngas, issued.

Man, Y., Yang, S., Zhang, J., Qian, Y., 2014. Conceptual design of coke-oven gas assisted coal to olefins process for high energy efficiency and low CO2 emission. Appl. Energy 133, 197–205.

Medrano-García, J.D., Ruiz-Femenia, R., Caballero, J.A., 2018. Multi-objective Optimization of a Carbon Dioxide Utilization Superstructure for the Synthesis of Formic and Acetic Acid, in: Computer Aided Chemical Engineering. Elsevier, pp. 1419–1424.

Qian, Y., Man, Y., Peng, L., Zhou, H., 2015. Integrated process of coke-oven gas tri-reforming and coal gasification to methanol with high carbon utilization and energy efficiency. Ind. Eng. Chem. Res. 54, 2519–2525.

Summa, P., Samojeden, B., Motak, M., 2019. Dry and steam reforming of methane. Comparison and analysis of recently investigated catalytic materials. A short review. Polish J. Chem. Technol. 21, 31–37.

Wang, H., Su, Y., Wang, D., Jin, S., Wei, S., Shen, W., 2020. Optimal Design and Energy-Saving Investigation of the Triple CO2 Feeds for Methanol Production System by Combining Steam and Dry Methane Reforming. Ind. Eng. Chem. Res. 59, 1596–1606.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50122-6

Network optimization of the electrosynthesis of chemicals from CO_2

Ana Somoza-Tornos[a*], Omar J. Guerra[b], Wilson A. Smith[a,b], Bri-Mathias Hodge[a,b*]

[a]*Renewable and Sustainable Energy Institute, University of Colorado, Boulder, CO 80309, United States*
[b]*National Renewable Energy Laboratory, Golden, CO, United States*
[*]*ana.somozatornos@colorado.edu, brimathias.hodge@colorado.edu*

Abstract

Carbon dioxide electroreduction (ECO2R) is gaining attention due to its capacity to mitigate CO_2 emissions while using electricity that would otherwise be curtailed. Its foreseeable industrial implementation requires of holistic methods to assess the technological and economic performance of ECO2R processes and integrate them in current chemical supply chains and power systems.
Here, we combine techno-economic assessment and mathematical programming to find the optimal paths to electroreduce CO_2 into valuable chemicals under variable electricity prices. The proposed approach is tested with a case study addressing the CO_2 capture from flue gas or direct air and its electricity-powered reduction into carbon monoxide, formic acid or multi-carbon compounds. The results obtained demonstrate the ability of the framework to build ECO2R networks and provide operation profiles that respond to fluctuating electricity prices.

Keywords: electroreduction, carbon dioxide, techno-economic assessment, superstructure, optimization.

1. Introduction

The electroreduction of carbon dioxide (ECO2R) can play a pivotal role in the transition to carbon-free chemicals. Catalyst and reactor design have led to an interesting range of electrosynthesized chemicals from single-carbon syngas and formic acid to multi-carbon commodities (e.g. ethylene, ethanol or propanol). However, the integration of these chemicals into the current market remains a challenge, mainly due to the electricity-intensive nature of ECO2R. Renewably-powered ECO2R can lead not only to the mitigation of CO_2 emissions, but also contribute to the use of electricity that would otherwise be curtailed. Conversely, techno-economic estimations show that under current electricity prices, the cost shares to power the electrolyzer can represent up to a 78% of the total cost (Somoza-Tornos et al., 2021). Thus, current research efforts should focus on scaling up the technology to industrially-relevant scales while integrating it with renewable energy systems in an economically viable manner.

In this work, we present an optimization framework to assess the large-scale implementation of ECO2R and its integration with renewable energy systems based on process modeling, techno-economic assessment and network optimization. It is built upon previous experimental work on ECO2R (Ma et al., 2021); studies on the modeling and

assessment of CCU and ECO2R (Jouny et al., 2018; Orella et al., 2019; Roh et al., 2020; Shin et al., 2021; Sisler et al., 2021; Zimmermann et al., 2020), hybrid fossil- and CO2-based routes (Ioannou et al., 2020) and supply chain optimization of carbon utilization processes (Leonzio et al., 2020; Zhang et al., 2020).

2. Modeling and techno-economic assessment of ECO2R processes

One of the main challenges of assessing the implementation of emerging ECO2R is the accurate evaluation of the costs that the technology will entail in short to mid-term time frames, when it is expected to be operated at industrial scales. To increase the accuracy of existing ECO2R techno-economic models, our framework includes a rigorous calculation of the mass balances of the systems under study, leading to better separation cost estimations. That is done through the calculation of the cell outlets based on the cell design and reaction mechanisms. Eq. 1, for instance, represents the global mass balance of the electrolyzer cell.

$$m_{cathode}^{in\,(gas)} + m_{electrolyte}^{in\,(aq)} = m_{cathode}^{out\,(gas)} + m_{electrolyte}^{out\,(aq)} + m_{anode}^{out\,(gas)} \quad (1)$$

The flows and compositions of the industrial-scale electrolyzer are later used in the techno-economic assessment to evaluate the unitary capital and non-energy operational costs, as well as the electricity and heat requirements. This data is then used as inputs for the optimization model in the next section.

3. Optimization model

In this section, the network optimization model for the combined design and operation of ECO2R systems is formally defined.

3.1. Problem statement

The problem addressed can be stated as follows: given are a set of CO_2 emissions sources (i.e. flue gas or air) and a set of technologies that enable the electricity-powered transformation of these emissions into valuable products (i.e. point source capture, direct air capture, and electroreduction of CO_2 and CO into chemicals including product purification) with their corresponding data on performance, cost and energy consumption. Given is also a yearly profile with hourly detail of electricity prices and other relevant economic parameters. Our goal is to find the optimal paths to best utilize CO_2 by its upcycling into commodity chemicals, tackling both design capacities and their operation according to fluctuating prices.

The elements of the network are represented through a superstructure, that provides all the paths to go from CO_2 emissions to carbon sinks, allowing for one and two-step electrosynthesis of products.

3.2. Mathematical formulation

The problem is formulated as a mixed-integer linear program (MILP) where the main variables are the capacity of technology *j* built (C_j) and its operational level at time period *t* (O_{jt}). The objective function to be minimized is the total cost (Eq. 2) including the capex and non-energy opex (calculated through the parameters for the unitary capex γ_j and opex σ_j), the electricity consumption (where θ_j denotes the unitary electricity consumption of each technology and ε_t the cost of electricity for time period *t*) and the heating costs (likewise, with ρ_j and τ). Eq. 3 denotes the material balance for each compound *i*, including the CO_2 sources utilized U_{it}, the transformation $\sum_j \mu_{ij} O_{jt}$ (where μ_{ij} is a

parameter linking the consumed and produced yields of compound *i* for each technology *j*) and the sales of final products P_{it}. Eqs. 4-7 represent the algebraic constraints to the decision variables: the set of demands of compounds *i* δ_i should be satisfied (Eq. 4) while not exceeding the available CO_2 sources (Eq. 5); the capacity of each technology cannot exceed a maximum level which is based on feasible plant sizes ($\bar{C}_j^{max}$) (Eq. 6); and the operation of each technology in each time period has to be below or equal to its capacity (Eq. 7).

$$\min Cost = \sum_j C_j \gamma_j + \sum_{jt} O_{jt} \sigma_j + \sum_{jt} O_{jt} \theta_j \varepsilon_t + \sum_{jt} O_{jt} \rho_j \tau \tag{2}$$

s.t.

$$U_{it} + \sum_j \mu_{ij} O_{jt} = P_{it} \qquad \forall i, \forall t \tag{3}$$

$$\sum_t P_{it} \geq \delta_i \qquad \forall i \in I^{products} \tag{4}$$

$$\sum_t U_{it} \leq \varphi_i \qquad \forall i \in I^{feeds} \tag{5}$$

$$C_j \leq \bar{C}_j^{max} \qquad \forall j \tag{6}$$

$$O_{jt} \leq C_j \qquad \forall j, \forall t \tag{7}$$

4. Case study

To illustrate the capabilities of the model, we define a case study on the CO_2 capture from flue gas or direct air and its electrically-powered reduction into carbon monoxide, formic acid or multi-carbon chemicals (including routes from CO_2 and CO to evaluate the one-step and two-step electrolysis).

The costs for the economic models of CO_2 capture from flue gas and direct air are retrieved from reports on their simulation and techno-economic assessment (James et al., 2019; Keith et al., 2018) and the experimental results needed for the modelling and assessment of ECO2R technologies are gathered from diverse lab-scale contributions (Chen et al., 2020; J. Li et al., 2019; Y. C. Li et al., 2019; Ma et al., 2020; Wang et al., 2020) identified at the review by Ma et al. (2021).

As an illustrative example, flue gas from a natural gas power plant with a capacity of 1,036 MW and annual CO_2 emissions of 3.5 M tonnes (United States Environmental Protection Agency (EPA), 2021) is used as a flue gas source, together with an unlimited supply of air for direct air capture. The demands of one multi-carbon product (ethylene) and one single-carbon (formic acid) are also defined. A target on the demand of ethylene of a 10% of the maximum stoichiometric yield of that CO_2 stream is used. And since the same assumption for formic acid would exceed the national yearly demand, the latter is assigned. The electricity prices profile chosen for the assessment is the hourly purchase prices for Texas in 2019, to avoid the effect of the 2020 Covid pandemic.

5. Results

The model is implemented in GAMS and solved with CPLEX 20.1. The MILP involves 307,489 continuous variables and 210,872 constraints and was solved in 42.2 CPUs on an Intel® Core™ i5-8250U processor operating at 1.60GHz.

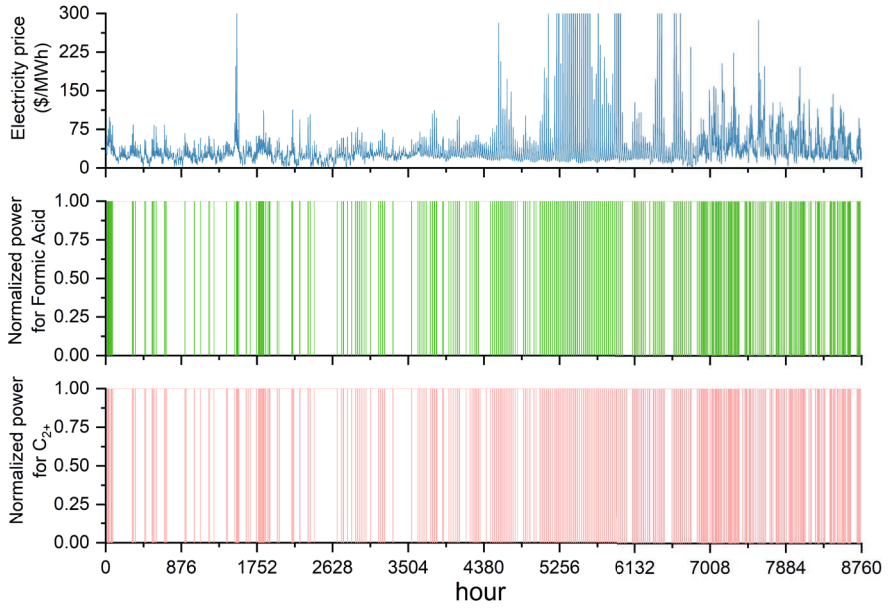


Figure 1. Hourly operation levels of the two selected technologies with respect to the electricity price profile.

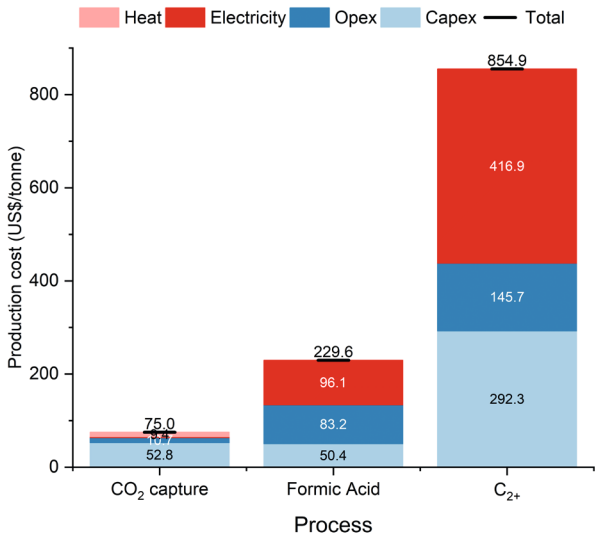


Figure 2. Production cost of the capture and ECO2R technologies resulting from the optimization.

The three technologies selected to satisfy the demands of ethylene and formic acid are the CO_2 capture from flue gas, the CO_2 reduction into formic acid, and one of the routes for

the CO_2 reduction into multi-carbon products (with mass yields of 54% ethylene, 27% acetic acid, 13% ethanol, 3% n-propanol and 3% hydrogen). Figure 1 shows the result for the optimal operation of the two ECO2R plants as a result of the variable electricity prices, with capacity factors of 85% for the formic acid production and 84% for the multi-carbon plant. The threshold for the electricity price that causes plant shutdowns for the present case is 49 USD/MWh. Figure 2 depicts the production cost breakdowns for the three resulting technologies. While the cost driver of the CO_2 capture from flue gas is its capital costs (52.8%), electricity is the main driver of ECO2R processes (42% for the reduction into formic acid and 48% for the reduction into multi-carbon products).

6. Conclusions

The results obtained show that the model is a useful tool to identify the most promising routes to convert CO_2 into valuable chemicals and its operative response to fluctuating electricity prices. Such methodology is valuable for the multiple actors involved in the mitigation of emissions and the alternative synthesis of chemicals: from private companies, who can identify the most economically promising processes; to policymakers, who can identify which routes should be incentivized to promote carbon capture and utilization.

ECO2R-specific conclusions can also be drawn from the studied case. CO_2 capture is preferred over direct air capture due to its higher cost-effectiveness. Also, one-step electrolysis is selected over the two-step route, since the higher efficiency of the latter does not compensate the higher capital expenses of two electrolyzers.

Future work will address the assessment of the main sources of uncertainty at the electrolyzer operating parameters and the economic model and the integration of environmental criteria in the assessment.

References

Chen, C., Yan, X., Liu, S., Wu, Y., Wan, Q., Sun, X., Zhu, Q., Liu, H., Ma, J., Zheng, L., Wu, H., Han, B., 2020. Highly Efficient Electroreduction of CO 2 to C2+ Alcohols on Heterogeneous Dual Active Sites. Angew. Chemie Int. Ed. 59, 16459–16464. https://doi.org/10.1002/anie.202006847

Ioannou, I., D'Angelo, S.C., Martín, A.J., Pérez-Ramírez, J., Guillén-Gosálbez, G., 2020. Hybridization of Fossil- and CO 2 -Based Routes for Ethylene Production using Renewable Energy. ChemSusChem cssc.202001312. https://doi.org/10.1002/cssc.202001312

James, R.E., Kearins, D., Turner, M., Woods, M., Kuehn, N., Zoelle, A., 2019. Cost and Performance Baseline for Fossil Energy Plants Volume 1: Bituminous Coal and Natural Gas to Electricity. https://doi.org/10.2172/1569246

Jouny, M., Luc, W., Jiao, F., 2018. General Techno-Economic Analysis of CO 2 Electrolysis Systems. Ind. Eng. Chem. Res. 57, 2165–2177. https://doi.org/10.1021/acs.iecr.7b03514

Keith, D.W., Holmes, G., St. Angelo, D., Heidel, K., 2018. A Process for Capturing CO2 from the Atmosphere. Joule 2, 1573–1594. https://doi.org/10.1016/j.joule.2018.05.006

Leonzio, G., Foscolo, P.U., Zondervan, E., 2020. Optimization of CCUS Supply Chains for Some European Countries under the Uncertainty. Processes 8, 960. https://doi.org/10.3390/pr8080960

Li, J., Wang, Z., McCallum, C., Xu, Y., Li, F., Wang, Y., Gabardo, C.M., Dinh, C.-T., Zhuang,

T.-T., Wang, L., Howe, J.Y., Ren, Y., Sargent, E.H., Sinton, D., 2019. Constraining CO coverage on copper promotes high-efficiency ethylene electroproduction. Nat. Catal. 2, 1124–1131. https://doi.org/10.1038/s41929-019-0380-x

Li, Y.C., Wang, Z., Yuan, T., Nam, D.-H., Luo, M., Wicks, J., Chen, B., Li, J., Li, F., de Arquer, F.P.G., Wang, Y., Dinh, C.-T., Voznyy, O., Sinton, D., Sargent, E.H., 2019. Binding Site Diversity Promotes CO 2 Electroreduction to Ethanol. J. Am. Chem. Soc. 141, 8584–8591. https://doi.org/10.1021/jacs.9b02945

Ma, W., He, X., Wang, W., Xie, S., Zhang, Q., Wang, Y., 2021. Electrocatalytic reduction of CO 2 and CO to multi-carbon compounds over Cu-based catalysts. Chem. Soc. Rev. https://doi.org/10.1039/D1CS00535A

Ma, W., Xie, S., Liu, T., Fan, Q., Ye, J., Sun, F., Jiang, Z., Zhang, Q., Cheng, J., Wang, Y., 2020. Electrocatalytic reduction of CO2 to ethylene and ethanol through hydrogen-assisted C–C coupling over fluorine-modified copper. Nat. Catal. 3, 478–487. https://doi.org/10.1038/s41929-020-0450-0

Orella, M.J., Brown, S.M., Leonard, M.E., Román-Leshkov, Y., Brushett, F.R., 2019. A General Technoeconomic Model for Evaluating Emerging Electrolytic Processes. Energy Technol. 1900994. https://doi.org/10.1002/ente.201900994

Roh, K., Bardow, A., Bongartz, D., Burre, J., Chung, W., Deutz, S., Han, D., Heßelmann, M., Kohlhaas, Y., König, A., Lee, J.S., Meys, R., Völker, S., Wessling, M., Lee, J.H., Mitsos, A., 2020. Early-stage evaluation of emerging CO 2 utilization technologies at low technology readiness levels. Green Chem. 22, 3842–3859. https://doi.org/10.1039/C9GC04440J

Shin, H., Hansen, K.U., Jiao, F., 2021. Techno-economic assessment of low-temperature carbon dioxide electrolysis. Nat. Sustain. https://doi.org/10.1038/s41893-021-00739-x

Sisler, J., Khan, S., Ip, A.H., Schreiber, M.W., Jaffer, S.A., Bobicki, E.R., Dinh, C.-T., Sargent, E.H., 2021. Ethylene Electrosynthesis: A Comparative Techno-economic Analysis of Alkaline vs Membrane Electrode Assembly vs CO 2 –CO–C 2 H 4 Tandems. ACS Energy Lett. 6, 997–1002. https://doi.org/10.1021/acsenergylett.0c02633

Somoza-Tornos, A., Guerra, O.J., Crow, A.M., Smith, W.A., Hodge, B.-M., 2021. Process modeling, techno-economic assessment, and life cycle assessment of the electrochemical reduction of CO2: a review. iScience 102813. https://doi.org/10.1016/j.isci.2021.102813

United States Environmental Protection Agency (EPA), 2021. Emissions & Generation Resource Integrated Database (eGRID), 2019. Office of Atmospheric Programs, Clean Air Markets Division, Washington, DC.

Wang, Y., Wang, Z., Dinh, C.-T., Li, J., Ozden, A., Golam Kibria, M., Seifitokaldani, A., Tan, C.-S., Gabardo, C.M., Luo, M., Zhou, H., Li, F., Lum, Y., McCallum, C., Xu, Y., Liu, M., Proppe, A., Johnston, A., Todorovic, P., Zhuang, T.-T., Sinton, D., Kelley, S.O., Sargent, E.H., 2020. Catalyst synthesis under CO2 electroreduction favours faceting and promotes renewable fuels electrosynthesis. Nat. Catal. 3, 98–106. https://doi.org/10.1038/s41929-019-0397-1

Zhang, S., Zhuang, Y., Liu, L., Zhang, L., Du, J., 2020. Optimization-based approach for CO2 utilization in carbon capture, utilization and storage supply chain. Comput. Chem. Eng. 139, 106885. https://doi.org/10.1016/j.compchemeng.2020.106885

Zimmermann, A.W., Wunderlich, J., Müller, L., Buchner, G.A., Marxen, A., Michailos, S., Armstrong, K., Naims, H., McCord, S., Styring, P., Sick, V., Schomäcker, R., 2020. Techno-Economic Assessment Guidelines for CO2 Utilization. Front. Energy Res. 8. https://doi.org/10.3389/fenrg.2020.00005

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50123-8

A robust design of heat exchanger network for high temperature electrolysis systems

Hua Liu[a], Lasse Røngaard Clausen[b], Ligang Wang[c], and Ming Chen[a*]

[a] *Department of Energy Conversion and Storage, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark*
[b] *Department of Mechanical Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark*
[c] *Institute of Energy Power Innovation, North China Electric Power University, 100193, Beijing, China*
[*] *minc@dtu.dk*

Abstract

The solid oxide electrolysis cell (SOEC) has been recognized as a promising technology for producing green hydrogen utilizing renewable energy. However, due to various degradation phenomena, the electrochemical performance of SOEC will deteriorate over time. This degradation leads to varying heat duty and inefficient operation. Based on previously reported durability test data, we simulate and optimize a SOEC system with a robust heat exchanger network to address the degradation issue. Overall, a 45.9% redundancy heat exchanger area is designed into the system, which leads to a levelized cost of hydrogen (LCOH) of 4.23 \$/kg H_2. The system efficiency is 77.4% initially but drops to 63.2% due to degradation. The vaporizer consumes the most energy in the heat exchanger network.

Keywords: Solid oxide electrolysis cells; Potentiostatic; Degradation; System design.

1. Introduction

Climate neutrality is one of the most essential European policies today, and using green hydrogen is indispensable to achieve this goal (Espegren et al., 2021). Green hydrogen is produced from water electrolysis powered by solar or wind energy. Among various electrolysis technologies, the solid oxide electrolysis cell (SOEC) is the most promising in the future due to its high electrochemical efficiency compared to its low temperature competitors (Ebbesen et al., 2014). The SOEC is a multi-layer unit consist of fuel electrode, electrolyte, and oxygen electrode. Its efficiency is strongly impacted by the operating voltage, necessitating careful control. When SOEC operates at thermoneutral voltage (~1.285 V), it has near 100% stack efficiency as the Joule heat fully compensates for the heat demand of the endothermic electrolysis reaction. Operating the SOEC stacks at thermoneutral voltage makes it easier to manage heat balance because no extra heat is needed from or released to the system. It also reduces potential cell cracking caused by large temperature gradients and excessive thermal stress within the stack.

Even though constant voltage operation has several advantages, it is not immune to the problem of cell degradation. The most likely causes for the cell degradation include formation of ZrO_2 nano-particles in the fuel electrode, which deteriorates the active triple phase boundary, and O_2 bubble formation in the electrolyte near electrode/electrolyte interface (Chen et al., 2013; Knibbe et al., 2010). Yang et al. recently reported a

potentiostatically operated SOEC durability test at thermoneutral voltage, which reveals the details of long-term cell degradation (Yang et al., 2021). As illustrated in Fig. 1, current density decreases dramatically over the first 300 hours and then stabilizes. Throughout the test, several unwanted interrupts and four load cycles occur, and the current density remains stable after each load cycle. This indicates that while SOEC degraded primarily during the initial stage, it is capable of remaining stable over time even when subjected to numerous load cycles.

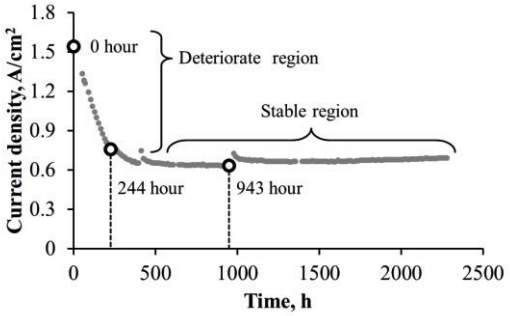


Fig. 1. Evolution of current density of an SOEC cell operated at 750°C and 1.29V for electrolysis of steam. Replotted based on the data reported by Yang et al., 2021.

From a system perspective, a decrease in current density results in a decrease in H_2 production and a change in the heat capacity of downstream flow. For efficiency and safety reasons, such uncertainty must be factored into the design stage. Given that the installed plant's SOEC effective cell area and heat exchange area are fixed, it requires a robust system designed to address the degradation issue and ensure the system's appropriate operation throughout its life. However, by designing an oversized heat exchange network, the system can achieve robustness efficiently and cost-effectively. (Chin et al., 2020; Kemp and Lim, 2020).

2. Process description

This paper proposes a robust SOEC system modified after the original design (AlZahrani and Dincer, 2017) as shown in Fig. 2. The system process is simulated and optimized based on the SOEC durability test profile (Yang et al., 2021). Table 1 presents the results of the durability test, including the current density. While data from the entire test period was used in the simulation, only those at the most critical time are listed here. The initial condition is set to 0 h, and the stable condition is set to 943 h. Data at 244 h are used solely for comparison purposes. The designed power $P_{\text{el,des}}$ is assumed to be 10 MW. As the stable region occupies most of the life cycle, the current density at the stable stage J_{stb} is used to estimate the cell area:

$$A_{\text{cell}} = \frac{P_{\text{el,des}}}{V_{\text{th}} J_{\text{stb}}} \tag{1}$$

During the test, the feedstock of the cathode is a mixture of 10% H_2 and 90% H_2O at 800 °C. Cathode feed stock flow rate is determined by the initial current density at thermal neutral voltage. To be more precise, the water molar flow rate $n_{\text{cat,H2O}}$ is:

$$n_{\text{cat,H}_2\text{O}} = \frac{J_{\text{ini}} A_{\text{cell}}}{2F \times SC_{\text{des}}} \tag{2}$$

where J_{ini} is initial current density, A_{cell} is cell area, F is Faraday constant, and SC_{des} is design steam conversion ratio, which is 85%. The feedstock flowrate remains constant throughout the test period. Meanwhile, because of degradation, current density decreases, along with the actual steam conversion ratio SC_{real} and electrolyze power $P_{el,real}$:

$$SC_{\text{real}} = \frac{J_i A_{\text{cell}}}{2F \times n_{\text{cat,H}_2\text{O}}} \tag{3}$$

$$P_{\text{el,real}} = J_i A_{\text{cell}} V_{\text{th}} \tag{4}$$

where J_i is current density at time i.

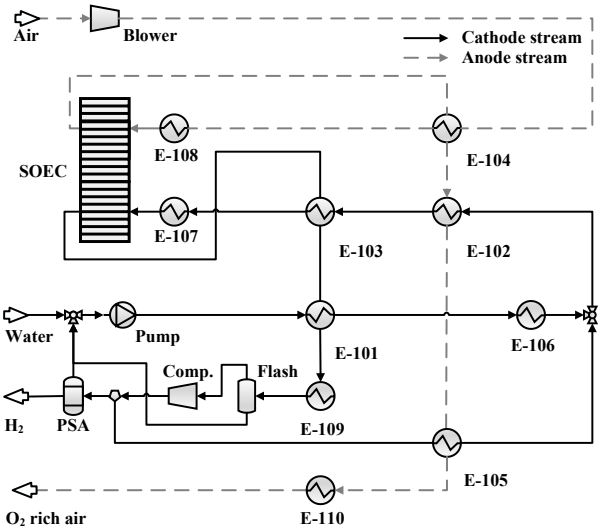


Fig. 2. Process flowsheet of the SOEC hydrogen production system.

Table 1. SOEC experiment profile from the durability test (Yang et al., 2021) and simulation respond at different time.

Time, h	Voltage, V	T, °C	J, A/cm^2	$n_{cat,H2O}$, mol/s	SC_{real}, %	$P_{el,real}$, MW
0	1.29	800	1.54	109.1	85.0	24.3
244	1.29	800	0.76	109.1	41.9	12.0
943	1.29	800	0.64	109.1	35.0	10.0

Before entering the SOEC at 800 °C, the water and air feedstocks are heated via a series of heat exchangers and an electric heater. Five of these heat exchangers (E-101~E-105) are configured for heat recovery. Given the current density deterioration, the optimal exchanger area will also shift over time. To get the optimal heat exchanger area, the temperature boundary of these five heat exchangers is optimized as a nonlinear constrained programme. The constraints include the mass and energy balance and the minimal temperature difference in the heat exchanger. The objective function is the total heat exchanger area. Then, for each heat exchanger, the design with the largest optimal exchange area over its life is chosen. Finally, the temperature settings in simulation are revoked, and the final oversized heat exchange areas are specified. Even though this increases the cost, it assures the system continues to run reliably over its entire life.

Besides the heat exchangers that improve robustness, the system involves other components such as evaporators, pre-heater, compressor, blower, and pump. The evaporator E-106 is an electric heater that vaporizes liquid water into steam. E-107 and E-108 are electric pre-heaters that ensure cathode and anode feedstock temperatures reach 800 °C. Anode product (O_2 rich air) is emitted into the environment after cooling down to room temperature by E-110. Considering the cost of purification, O_2 is inappropriate for separation and storage. Meanwhile, high commercial value product H_2 is treated another way. After being cooled by E-109, the cathode product is separated by a flash into two phases: the vapor and the liquid phase. The vapor phase is a mixture of H_2O and 98.2% H_2, which required a further purification by PSA. Liquid water from the flash and PSA are mixed with the makeup water before being fed to the system. A portion of the vapor phase from flash is heated and recycled to dilute the cathode feedstock to 90% H_2O.

3. Result and discussion

In this paper, a 10 MW SOEC system for hydrogen production is simulated in Aspen Hysys. Pinch analyses are evaluated in Aspen Energy Analyser. The optimization of the heat exchanger network is calculated in MATLAB. Optimal temperature set up and oversized UA is introduced from MATLAB into Aspen Hysys via COM interface following Hysys Customization Guide (AspenTech, 2011).

The grand composite curve of the SOEC system at both initial and steady stage are compared in Fig. 3. In both cases, the SOEC is operated in potentiostatic mode at thermal neutral voltage. As illustrated, water vaporization is the most energy intensive part in the heat exchanger network. The electricity consumption of vaporization accounts for 29.5% of the total system electricity consumption, while SOEC accounts for 67.0%. Apart from improving the efficiency of the SOEC stack, another way to save energy in this system is to rely on a cheap vaporization approach. The operation cost would be reduced if a less expensive heat source that can vaporize the water is available, such as abundant heat from fuel synthesis (Wang et al., 2018). To separate the water and hydrogen, the cathode product must be cooled to room temperature. Condensation cooling demand increases over time as the steam conversion ratio decreases and more water stays in the cathode product.

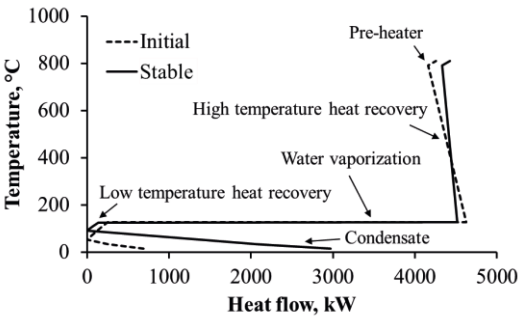


Fig. 3. Grand composite curve of the SOEC system in each degradation stage.

Due to degradation, both the composition and the heat capacity of the cathode product change, resulting in a variation in heat recovery duty, as well as a different optimal exchanger area for heat recovery. The varying optimal exchange areas of five heat recovery heaters (E-101~E-105) are depicted in Fig. 4. For example, the overall heat

transfer coefficient (UA) of E-101 and E-103 decreased by 28.7% and 35.9%, respectively. Also, not all heat exchanger areas need to be reduced with degradation. At 244 hours, the optimal UA of E-102 is 4.09 kJ/°C/h, but it is only 1.84 kJ/°C/h at the start. This emphasizes the necessity of doing detailed robust heat exchanger network design as feasible for all stages of SOEC performance.

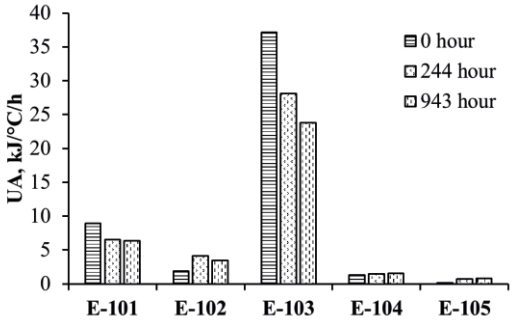


Fig. 4. Overall heat transfer coefficient (UA) of the heat exchangers at each time.

System efficiency (LHV) of the system is shown below, which is also depicted in Fig 5:

$$\eta_{sys} = \frac{\text{H}_2 \text{ combustion heat}}{\text{total electric power input}} \tag{5}$$

The SOEC system efficiency is deteriorating simultaneously as the current density. In the first 300 hours, the system efficiency also decreases by 14.2 %. Overall, the current density drops by 58.8 % from initial stage to stable stage, and the system efficiency decreases from 77.4% to 63.2%. Although the system efficiency reduces sharply over the first 300 hours, this problem can be mitigated in practice. The situation occurs because the initial water flow rate is too high and applying a lower water flow rate can improve efficiency. After each load cycle, SOEC recovers a little bit, which result in a rise of system efficiency. When the SOEC is restarted at 1000 h, the current density rises 0.09 A/cm^2, and the system efficiency enhances by 2.66%.

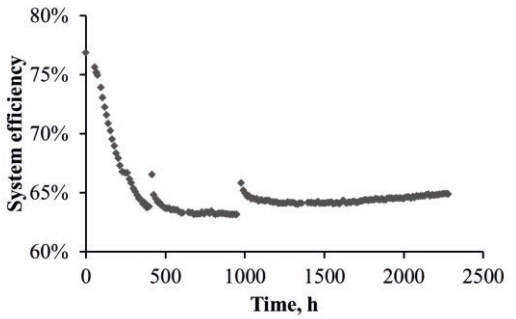


Fig. 5. System efficiency of the SOEC hydrogen production process.

This work also examines the economics of the robust SOEC system. It should be noted that the total capital cost (TPC) is calculated based on the oversized capacity of the components (particularly the heat exchanger). As a result, the system has a greater TPC than the conventional design. Simultaneously, a drop in current density results in a reduced rate of hydrogen production, thereby diminishes the revenue. The robust

optimization prolongs the system's life by 1.6% at the cost of an 11% increase in equipment costs. In the life cycle, the levelized cost of hydrogen (LCOH) is \$4.25/kg H_2 and the net present value (NPV) is \$13.5 million.

4. Conclusions

In summary, this paper presents a SOEC hydrogen production system incorporates oversized heat exchangers to handle degradation concerns. SOEC degradation manifests as a decrease in current density during potentiostatic operation. Both electric power and system efficiency decline in lockstep with the trend of decreasing current density. Additionally, the optimal heat exchange area changes because of degradation, and 45.9% more heat exchange area can handle this uncertainty. Meanwhile, vaporizers consume most of the electricity in the heat exchange network. When other techniques for reducing evaporating costs are considered, the system still has the potential to increase profitability. In future studies, we will look at how to improve the SOEC system design and operating strategy to increase durability.

References

K. Espegren, S. Damman, P. Pisciella, I. Graabak and A. Tomasgard, 2021. The role of hydrogen in the transition from a petroleum economy to a low-carbon society. Int. J. Hydrogen Energ. 46, 23125–23138.

S. D. Ebbesen, S. H. Jensen, A. Hauch and M. B. Mogensen, 2014. High Temperature Electrolysis in Alkaline Cells, Solid Proton Conducting Cells, and Solid Oxide Cells. Chem. Rev. 114, 10697–10734.

M. Chen, Y.-L. Liu, J. J. Bentzen, W. Zhang, X. Sun, A. Hauch, Y. Tao, J. R. Bowen and P. V. Hendriksen, 2013. Microstructural Degradation of Ni/YSZ Electrodes in Solid Oxide Electrolysis Cells under High Current. J. Electrochem. Soc. 160, F883–F891.

R. Knibbe, M. L. Traulsen, A. Hauch, S. D. Ebbesen, & M. Mogensen, 2010. Solid Oxide Electrolysis Cells: Degradation at High Current Densities. J. Electrochem. Soc. 157, B1209.

Y. Yang, X. Tong, A. Hauch, X. Sun, Z. Yang, S. Peng and M. Chen, 2021. Study of solid oxide electrolysis cells operated in potentiostatic mode: Effect of operating temperature on durability. Chem. Eng. J. 417, 129260.

H. H. Chin, B. Wang, P. S. Varbanov, J. J. Klemeš, M. Zeng, and Q. Wang, 2020. Long-term investment and maintenance planning for heat exchanger network retrofit. Appl. Energ. 279, 115713.

I. C. Kemp, & J. S. Lim, 2020. Pinch Analysis for Energy and Carbon Footprint Reduction. 131–194

A. A. AlZahrani and I. Dincer, 2017. Thermodynamic and electrochemical analyses of a solid oxide electrolyzer for hydrogen production. Int. J. Hydrogen Energ. 42, 21404–21413.

AspenTech, 2010. Aspen HYSYS: Customization Guide.

L. Wang, M. Pérez-Fortes, H. Madi, S. Diethelm, J. V. herle and F. Maréchal, 2018, Optimal design of solid-oxide electrolyzer based power-to-methane systems: A comprehensive comparison between steam electrolysis and co-electrolysis, Appl Energ, 211, 1060–1079.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50124-X

Techno-economic Assessment of Upstream and Downstream Process Alternatives for the Production of Monoclonal Antibodies

Sara Badr, Kozue Okamura, Nozomi Takahashi, Hirokazu Sugiyama*

Department of Chemical System Engineering, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, 113-8656, Tokyo, Japan
sugiyama@chemsys.t.u-tokyo.ac.jp

Abstract

The rising demand for monoclonal antibodies is increasing the pressure for improving production efficiency and lowering costs. A high-resolution assessment of two bottleneck units is provided. Within the main cultivation unit, different cell lines were assessed at various production scales, and operating modes. In addition to operating costs, and production time, the generation of impurities, e.g., host cell proteins was also estimated using a hybrid model. A newly established cell line was shown to be highly productive, but presenting a trade-off between production costs and time, and generated impurity levels. A superstructure was built combining design and operating parameters of the capture chromatography unit. Scenarios regarding variations in inlet conditions based on upstream disturbances were analyzed. Process robustness was evaluated and mitigation measures suggested. This work provides a more comprehensive assessment framework that extends beyond cost and time to include quality, and potential operational problems. This could help guide process design and control efforts.

Keywords: Cell cultivation; Capture Chromatography; Continuous production; Process assessment; Biopharmaceutical production; Therapeutic proteins.

1. Introduction

Demand for therapeutic proteins, and especially monoclonal antibodies (mAb)s, is rising, having already increased many folds in the past decade (Walsh, 2018). Applications for mAbs are expanding for the treatment of diseases such as cancer and even most recently for COVID-19. The pandemic has placed the efficiency of biopharmaceutical manufacturing under scrutiny. Increasing productivity is now necessary to avoid bottlenecks and keep up with the rising demand, especially for plants already operating at near capacity. Another major challenge is to reduce production costs to increase accessibility and to fend off rising competition from biosimilars.

A typical mAb production process involves upstream cell cultivation to produce the required antibody product and downstream purification units to reach the required product profile as shown in Figure 1. Efforts to increase productivity include cell line modifications, manipulating operating conditions to improve cell productivity, or changing operating modes from batch to continuous. Continuous operations can potentially offer higher productivities, more flexibility as well as being more suitable for unstable products. However, they are also susceptible to increased operational difficulties, such as clogging, which could lead to fluctuations in downstream loads and longer process downtimes.

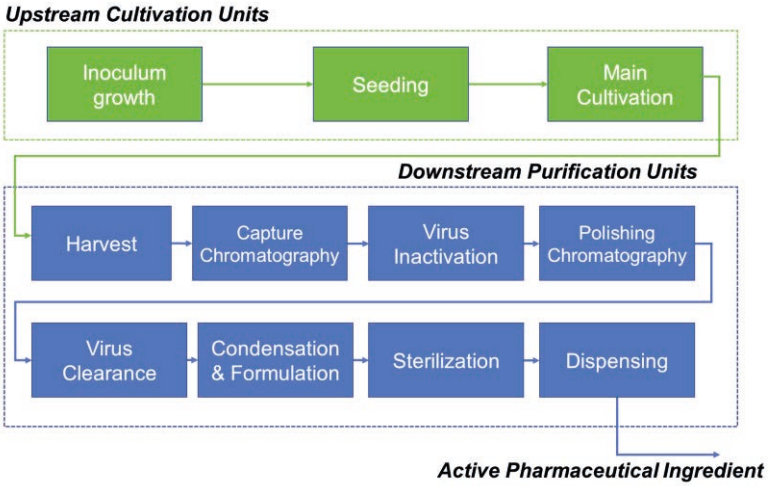


Figure 1: Typical mAb production process

Most studies comparing batch and continuous operations focus on ideal operations. A comprehensive framework is still needed for comparison including product quality and production robustness in addition to production time and costs.

This work offers an assessment of different upstream process alternatives involving variations in cell lines, operating conditions, and cultivation modes. A comparison is presented in terms of operating costs and expected product quality. Downstream capture chromatography is also investigated regarding differences in design and control parameters. Different scenarios of upstream changes are presented and their influence on capture costs and productivity. A discussion of measures needed to maintain productivity is also provided.

2. Simulation of process alternatives

2.1. Main cultivation unit

Simulated alternatives described in this work include fed-batch and perfusion cultivation (Karst et al., 2018). Fed-batch cultivation is where nutrient levels are maintained during operation to avoid cell death without removing product from the reactor. Perfusion cultivation is where continuous addition of nutrients is balanced with product and by-product removal to reach steady operation. Three different Chinese hamster ovary (CHO) cell lines are depicted in this work. Cell A is from a newly established cell line, which is proclaimed to be more productive than traditional cell lines. Further details can be found in Okamura et al. (2022). Simulations with Cell B were based on the fed-batch experiments presented in detail by Badr et al. (2021) at and Cell C on that presented by Xu and Chen (2016) with perfusion operation.

Cell cultivation models used in this work were based on the model presented by Badr et al. (2021). The model presented there was validated with experimental data generated from the Kobe GMP consolidated lab of the Manufacturing Technology Association of Biologics. The newly established cell line used as the basis for the Cell A simulations are more sensitive to lactate concentrations. A novel hybrid modelling approach is introduced in Okamura et al. (2022), where a data-driven module accounts

for the impact of changes in experimental conditions on model parameters and acts to provide a dynamic correction for lactate concentrations. The generation of host cell proteins (HCP)s and other process-related impurities was accounted for through description of cell death and dissolution (Maruhashi et al., 1994).

2.2. Downstream capture chromatography

A comparison of batch and continuous operations in the capture unit is presented. While chromatography is typically conducted as batch, continuous operation can be achieved through running several columns together. An example of continuous capture chromatography is periodic counter current (PCC) operations, where one column is loaded while others go through cycles of washing and regeneration (Baur et al., 2016). A sensitivity analysis was carried out varying the column design and control parameters. Within this analysis a superstructure of process options was built, where operational costs and productivity of the capture unit were compared. The superstructure was used to identify promising alternatives. In addition to varying the operating mode from batch to continuous, the super structure also included changes in the number of columns, column dimensions, superficial velocity in the column, and the column switching criterion in continuous mode. The model presented in Badr et al. (2021) was used for flow description within the column and for estimating the costs and productivity.

Four different scenarios regarding inlet concentrations and flowrates were also tested in this analysis as shown in Figure 2. The scenarios represent different potential disturbances that can arise in upstream units. In the baseline scenario Sc. 1, a constant inlet was simulated based on the expected flowrates from a typical perfusion or fed-batch cultivation operation. Sc. 2 represents the impact of the start-up and shutdown phases within the continuous operation in a more realistic portrayal of the inlet form perfusion mode. In the absence of active control in the upstream units, concentration peaks and fluctuations can be observed as seen in Sc. 3 and 4, respectively. The impacts of such disturbances were investigated along with mitigation measures in the design and operation of the capture columns.

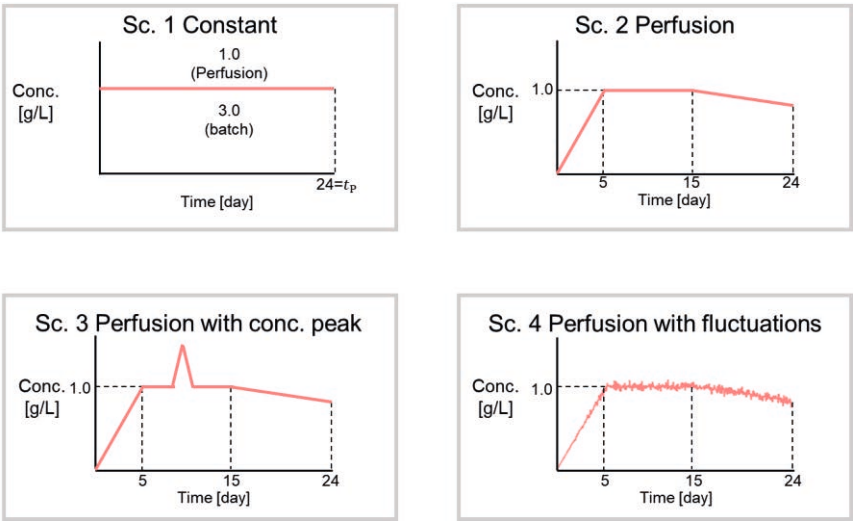


Figure 2: Downstream inlet scenarios

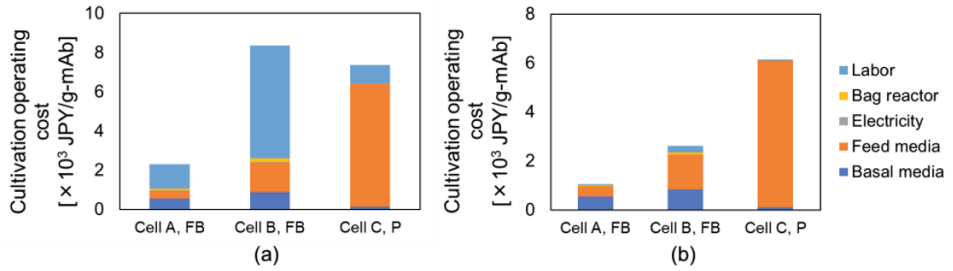


Figure 3: Cultivation operating costs for scenarios (a) 2 kg production scale in 50 L reactor and (b) 200 kg production scale in 2000 L reactor

3. Results and discussion

3.1. Cultivation unit

Figure 3 shows a comparison of cultivation operating costs and their breakdown at different production scales, reactor volumes, operating modes and cell lines. The newly established cell line (Cell A) resulted in the lowest overall costs per gram mAb produced. The higher productivity of this cell line also resulted in a reduced cultivation time (~4.4 times shorter than that of Cell B). The highest productivity and lowest cultivation time were achieved by the perfusion mode, followed closely by that of Cell A in fed-batch mode. However, perfusion mode also yielded the highest costs per gram at commercial scales, mainly due to the higher media consumption. At smaller scales and reactor volumes, labor costs were more pronounced for fed-batch units.

Figure 4 shows the calculated concentrations of HCPs in the reactor for Cells A and B. HCP concentrations as well as other process related impurities, such as DNA, were found to be higher for Cell A than for the other tested alternatives. Therefore, the load to downstream purification units from Cell A can be higher than the other candidates.

Further improvements can still be tested, for example, by using the more productive Cell A in perfusion mode. This could result in additional productivity gains, while maintaining the generated impurity concentrations at lower levels.

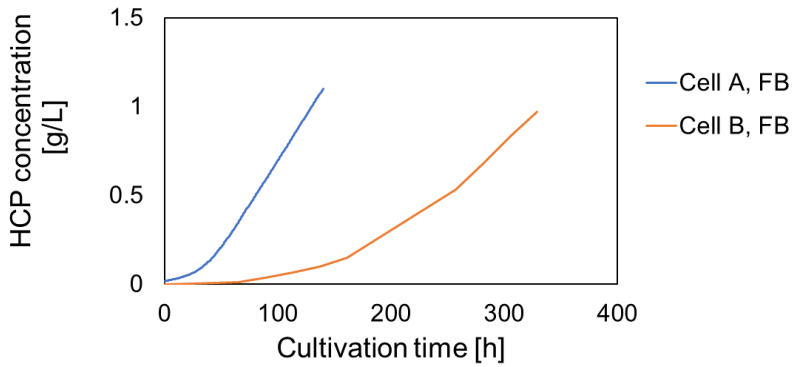


Figure 4: Simulated host cell protein concentrations in the reactor

3.2. Capture chromatography

In total 1,440 combinations of process alternatives were tested in the developed superstructure. Figure 5 shows the resulting pareto optimal alternatives considering operating capture costs and productivity for integrated batch options with fed-batch inlet and integrated continuous processes with perfusion inlet.

The superstructure results show that continuous options can offer lower costs compared to the batch alternatives. The lower productivity can also be attributed to the lower titers of the perfusion cultivation units, which results in a need to process larger volumes downstream to achieve the same product mass. Optimizing operating conditions can help increase productivity, e.g., increasing superficial velocity. However, care should be taken to avoid product losses especially when high velocities are coupled with short columns and slow switching between columns in PCC operations. Product losses lower the yield per cycle and lead to significant increases in costs. Product losses should thus be minimized for robust design. This effect is particularly highlighted when upstream disturbances are taken into account, especially for scenario Sc. 3 with the sudden concentration peaks. Therefore, the results from all scenario analysis should be carefully considered to reach a compromise between maintaining high productivity and robustness with upstream changes.

Ideally, one of the advantages of integrated continuous operations, would be eliminating the need for intermediate surge tanks. This could have a big impact on achieving reductions in process capital costs. However, surge tanks can still be beneficial for mitigating upstream disturbances, in particular concerning fluctuations in inlet flowrates and concentrations. Careful determination of the required surge tank volume is necessary to avoid suboptimal operation. The conducted scenario analysis is thus a crucial tool towards achieving this goal.

Overall, the analysis showed that hybrid alternatives with fed-batch upstream and continuous downstream options can outperform integrated batch or continuous operations. This is achieved by combining the advantages of both, with higher titers from upstream operations and more efficient continuous downstream conditions. However, there is still room for improvement with developments upstream, e.g., with using the more productive Cell A in perfusion mode, higher titers could be achieved.

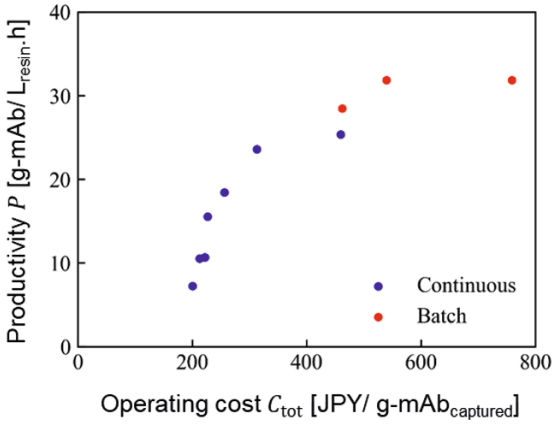


Figure 5: Pareto optimal design alternatives for the capture chromatography unit

4. Conclusions

In this work a high-resolution analysis was undertaken for two bottleneck units in mAb production. Within the upstream cultivation unit, the newly established cell line gave superior results in terms of operating costs and reasonable operating times even in fed-batch operation. The performance of this cell line should be investigated in more novel perfusion operation or in different experimental conditions to further enhance its performance in the established fed-batch operations. Cell A, however, resulted in a higher concentration of process related impurities (e.g., HCPs). This could cause increased loads on the downstream units. This work does not differentiate between different HCP structures though. A unified measure of quality between upstream and downstream units should still be determined. As downstream processing difficulties do not only arise from the increased volumes and concentrations but could also be influenced by similarities between the impurities to be removed and the main product.

The importance of incorporating expected upstream disturbances in the design of downstream units has been demonstrated. The analysis in this work was extended beyond costs and time to also include quality aspects in the upstream units and robustness downstream. This approach provides a comprehensive overview of performance under varying production scenarios, production scales, and possible operational issues. The analysis offers a deeper understanding of influential process parameters, the sources of disturbances in the system, their magnitudes, and potential mitigation measures.

Acknowledgements

This work was supported by the Japan Agency for Medical Research and Development (AMED) [grant No. JP21ae0121015, JP20ae0101064, JP20ae0101058].

References

Badr, S., Okamura, K., Takahashi, N., Ubbenjans, V., Shirahata, H., Sugiyama, H., 2021. Integrated design of biopharmaceutical manufacturing processes: Operation modes and process configurations for monoclonal antibody production. Comput. Chem. Eng. 153, 107422.

Baur, D., Angarita, M., Müller-Späth, T., Steinebach, F., Morbidelli, M., 2016. Comparison of batch and continuous multi-column protein A capture processes by optimal design. Biotechnol. J. 11, 920–931.

Karst, D.J., Steinebach, F., Morbidelli, M., 2018. Continuous integrated manufacturing of therapeutic proteins. Curr. Opin. Biotechnol. 53, 76–84.

Maruhashi, F., Murakami, S., Baba, K., 1994. Automated monitoring of cell concentration and viability using an image analysis system. Cytotechnology 15, 281–289.

Okamura, K., Badr, S., Murakami, S., Sugiyama, H., 2022. Hybrid modelling of CHO-MK cell cultivation in monoclonal antibody production. Submitted to the 14th International Symposium on Process Systems Engineering – PSE 2021+.

Walsh, G., 2018. Biopharmaceutical benchmarks 2018. Nat. Biotechnol. 36, 1136–1145.

Xu, S., Chen, H., 2016. High-density mammalian cell cultures in stirred-tank bioreactor without external pH control. J. Biotechnol. 231, 149–159.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50125-1

Biomethane liquefaction followed by CO_2 solidification based biogas upgrading process

Ahmad Naquash[a], Muhammad Abdul Qyyum[b], and Moonyong Lee[a*]
[a]School of Chemical Engineering, Yeungnam University, Gyeongsan 712-749, Rep. of Korea
[b] Department of Petroleum & Chemical Engineering, Sultan Qaboos University, Muscat, Oman
mynlee@yu.ac.kr

Abstract

Energy consumption and climatic changes are challenging issues these days. Unlike fossil fuels, the utilization of renewable fuels such as biogas is a promising option to meet these challenges. Biogas upgraded form (biomethane) is an emerging potential alternative to natural gas. However, biomethane production is itself challenging because of the pros and cons of each biogas upgrading technology. In this study, a cryogenic technology is adopted because of its dual benefits; carbon dioxide (CO_2) removal and biomethane precooling due to its low-temperature operation. The CO_2 is removed from biogas through the CO_2 solidification process. The phase behavior of CO_2 is investigated and the specified conditions for CO_2 solidification (-68°C and 5.17 bar) are applied for biomethane production. After CO_2 removal, biomethane is liquefied. The refrigeration duty for upgrading and liquefaction is provided by a parallel nitrogen expansion cycle adopting pure nitrogen as a refrigerant. Aspen Hysys® v11 is used as a commercial simulator for process simulation and to evaluate CO_2 freezing behavior in the proposed study. The mixed optimization technique is employed to optimize the design variables of the proposed process. The proposed process shows energy and exergy savings of 17.4 and 29.7%, respectively. It is evaluated that the proposed integrated process depicts superior results than the conventional studies.

Keywords: Biomethane production; Liquified biomethane; CO_2 solidification; Anti-sublimation; Mixed optimization; Parallel nitrogen expansion cycle.

1. Introduction

Due to increasing energy demand and environmental challenges, the use of biogas has seen an increasing trend at a rate of 11.5% annually (International Energy Agency, 2020). Owing to its increasing production, transportation of biogas is becoming a matter of concern especially because of production at atmospheric pressure. Transportation at this pressure is uneconomical due to its low energy density (Krich et al., 2005). This energy density can be enhanced either through compression or liquefaction. Compression is beneficial only for shorter distances. For longer distances, liquefaction is an economical approach. However, the impurities must be removed before liquefaction to avoid carbon dioxide (CO_2) freezing or maintenance issues. Typically, CO_2 below 50 ppm is recommended in the upgraded biogas i.e., biomethane (Fan et al., 2018).

Biogas upgrading can be carried out through absorption, adsorption, membrane, and cryogenic technology. Amongst these technologies, cryogenic technology is the most

viable technology if integrated with the liquefaction process, known as liquefied biomethane (LBM) production (Naquash et al., 2021). The cryogenic technology is of two types, depending on the state of CO_2 separated. The cryogenic distillation process is adopted to separate CO_2 in liquid form (Yousef et al., 2017) while anti-sublimation technology is adopted to separate CO_2 in solid-state (Spitoni et al., 2019). Separation of CO_2 in solid form is preferable because of high product purity. Various researchers have explored these technologies technically and economically. For example, the cryogenic and absorption-based biomethane production integrated with a liquefaction process was studied by (Pellegrini et al., 2018). Their results showed that the absorption upgrading, and liquefaction process requires 34.8% of the total energy to liquefy biomethane, which is significantly higher than cryogenic processes. Similarly, the absorption and cryogenic biogas upgrading processes integrated with liquefaction were analyzed by (Hashemi et al., 2019). Their results showed that the cryogenic process is more efficient, with lower specific energy consumption (SEC) of 2.07 kWh/kg bio-LNG than that of the absorption process (3.35 kWh/kg) (Hashemi et al., 2019). In another study, the CO_2 solidification-based biogas upgrading integrated with the liquefaction process was investigated and the reported SEC was 1.45 kWh/kg (Baccioli et al., 2018). Similarly, the CO_2 solidification-based biogas upgrading integrated with the liquefaction process under various feed compositions was also studied recently and the reported SEC ranged from 1.093 to 1.574 kWh/kg (Spitoni et al., 2019). Recently, another study was proposed in which a biogas upgrading process through CO_2 solidification followed by LBM production was investigated. The total SEC was 0.495 kWh/kg (Naquash et al., 2021).

It is analyzed from the literature review that the cryogenic upgrading process is energy-intensive and complex. In this study, a simple and energy-efficient process is proposed that adopts a single refrigeration cycle utilizing pure refrigerant to produce high purity LBM. The refrigeration cycle i.e., parallel nitrogen expansion cycle (PNEC) is a simple cycle utilizing pure N_2 as a refrigerant. The application of PNEC has already been explored in the liquefaction of natural gas (He et al., 2019). In this study, the application of PNEC is evaluated for cryogenic biogas upgrading and LBM production. The process is simulated in Aspen Hysys® v11. To explore energy-saving potential, the process is optimized through Aspen Hysys® v11 in-built optimizer using mixed optimization technique. The proposed process is analyzed through energy and exergy analyses to identify and locate process inefficiencies.

2. Process design and simulation

2.1. Process description

The process flow diagram of the proposed process is shown in Figure 2. The feed stream (Biogas) at 1.0 bar is first compressed to 6.97 bar by compressors (K1 and K2) assisted with aftercoolers (E1 and E2) before entering in CHX1. After CHX1, the feed stream temperature is dropped to -66℃ at 5.17 bar in stream (5). The stream (5) is entered into the cold box (CB1) where the temperature is further decreased to -68℃. The CB1 conditions are selected according to the solidification conditions of CO_2. The phase diagram of the feed stream is prepared according to the data calculated by Aspen Hysys® v11, as shown in Figure 1. According to Figure 1 (enlarged version), the feed conditions at -68℃ and 5.17 bar are in the region of CO_2 solidification.

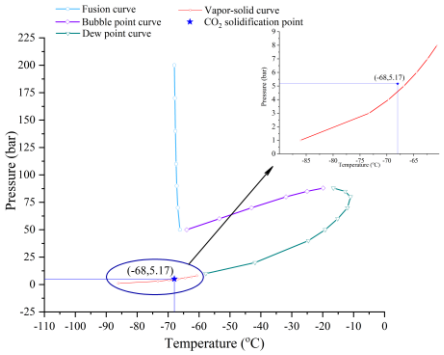


Figure 1 Pressure-Temperature phase diagram of biogas feed (CH_4/CO_2: 0.5/0.5)

In CB1, the CO_2 is separated in solid form and biomethane is separated from the top of CB1. For simplification, only one cold box is shown in Figure 2. Considering continuous operation of solidification process, another cold box configuration can be installed. Biomethane is then sent to CHX2 for liquefaction. At the outlet of CHX2, the biomethane is 100% liquefied. The pressure of LBM is reduced to 1.2 bar by passing through an expander (T3) for storage at liquid conditions. The cooling duty for biogas upgrading and biomethane liquefaction is provided by PNEC. In PNEC, pure nitrogen (N_2) is used as a refrigerant, which is compressed to 16.83 bar in a series of compressors. After compression, stream (18) is cooled to -66°C by passing through CHX1. After CHX1, the stream (19) is split into streams (20 and 21) by TEE2. Stream 21 pressure is reduced to 4.20 bar which further enters MIX2. Stream (20) is further cooled to -137.3°C by passing through CHX2 in stream (23). Stream (23) pressure is then reduced to 4.20 bar which reenters CHX2 to exchange its cold energy. Stream (25), leaving CHX2, is mixed with stream (22) in MIX2 and reenters CHX2 to further exchange its cold energy. After leaving CHX2, stream (27) at -88°C enters CHX1 to assist in temperature decrease of stream (4) and stream (18). After CHX1, stream 28 at 37.14°C and 3.90 bar is recycled back to complete the refrigeration loop.

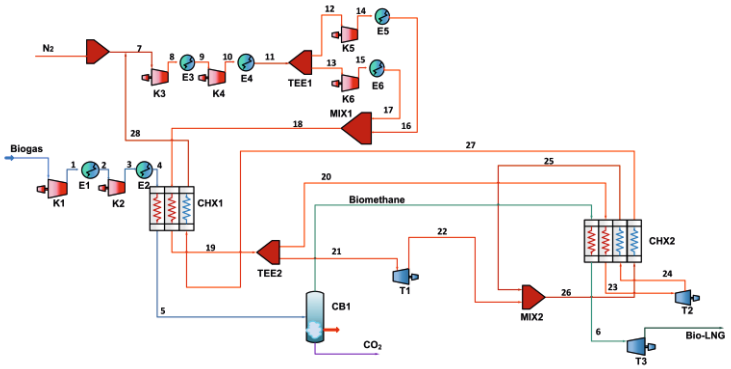


Figure 2 Process flow diagram of the proposed process

2.2. Process simulation

The process is simulated in Aspen Hysys® v11. Peng-Robinson is used as the equation of state (Peng and Robinson, 1929). The feed conditions and design parameters have opted from the Base case (Naquash et al., 2021). The following assumptions (Naquash et al., 2021) were taken in the simulation of this study:

1. The pressure drop in coolers is 0.25 bar.
2. The pressure drop in CHXs is 1.0 and 0.1 bar for the hot and cold sides, respectively.
3. It was assumed that CO_2 is completely solidified at -68°C and 5.17 bar.

The design parameters of the proposed process are presented in Table 1.

Table 1 Design parameters and constraints of proposed process

Design Parameters	**Values** (Naquash et al., 2021)
Feed biogas conditions	T: 35 °C P: 1 bar Flowrate: 308.2 kg/h
Feed biogas composition (mole)	CH_4: 0.5 CO_2: 0.5
Compressor efficiency	80%
Turbine efficiency	90%
Design constraints	
MITA value (°C)	1.0 ~ 3.0
Inlet temperature of compressors	$>T_{dew}$

3. Process optimization

The process was optimized in the in-built optimizer of Aspen Hysys® v11. In this study, a mixed (combination of BOX and SQP) optimization method was adopted to calculate the optimal values of design variables keeping the design constraint value within the range. The objective function is to reduce SEC while keeping the design constraint i.e., minimum internal temperature approach (MITA) in the range of 1 to 3°C. The values of refrigerant flow rate, suction and discharge pressure of refrigeration cycle, and split ratios of TEE1 and TEE2 are the design variables optimized through mixed technique.

4. Results and discussion: Process analysis

The energy and exergy analyses are performed which are discussed in the following sections.

4.1. Energy analysis

The energy analysis of a liquefaction process is typically described in terms of SEC and refrigerant flowrates. The values of refrigerant flowrates, SEC along suction and discharge pressure of refrigeration cycles of the base case and the optimized proposed process is presented in Table 2. It can be seen from Table 2 that the proposed process is energy efficient with 17.4% energy savings compared to the base case process. In the base case process, two separate refrigeration cycles were used whereas in the proposed process, a single refrigeration cycle is used with pure refrigerant to make a simple, and energy-efficient biomethane production and liquefaction process. However, as compared to the base case, the process consumes more refrigerant flowrate i.e., 2398 kg/h which is 33.4% more.

Table 2 Design variables values of base case (Naquash et al., 2021) and proposed case

Design variables	**Base case**	**Proposed process**
Biogas upgrading section		
m_{C1} (kg/h)	66.6	-
m_{C3} (kg/h)	87.5	-

m_{CO2} (kg/h)	718.1	-
Total refrigerant flowrate (kg/h)	872	-
Suction pressure (bar)	11.8	-
Discharge pressure (bar)	50.7	-
Specific energy consumption (kWh/kg)	0.162	-
Bio-LNG section		
m_{N2} (kg/h)	85	2398
m_{C1} (kg/h)	376	-
m_{C2} (kg/h)	165	-
m_{C3} (kg/h)	98	-
Total refrigerant flowrate (kg/h)	724	2398
Suction pressure (bar)	3.1	4.2
Discharge pressure (bar)	68.7	17.1
Specific energy consumption (kWh/kg)	0.333	
Total specific energy consumption (kWh/kg)	**0.495**	**0.408**
Relative energy savings (%)		**17.4**

4.2. Exergy analysis

The exergy analysis of the optimized process is performed. The equations adopted for exergy analysis are taken from (Venkatarathnam and Timmerhaus, 2008)**.** The values of exergy destruction for each piece of equipment are presented in Figure 3.

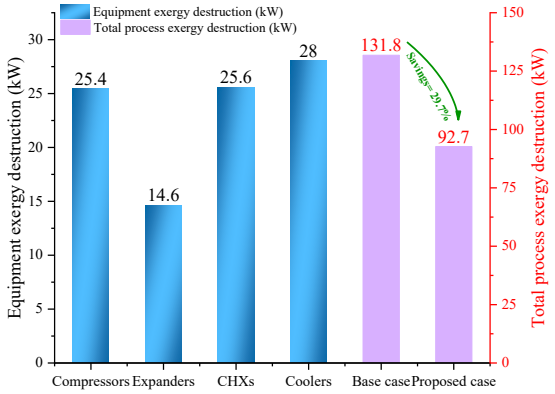


Figure 3 Total process and equipment exergy destruction values

The highest exergy destruction (28 kW) is attributed to coolers followed by compressors (25.4 kW) whereas expanders exhibit the lowest exergy destruction (14.6 kW). The main reason for high exergy destruction in compressors and coolers is due to high SEC by refrigeration cycle compressors. This high exergy destruction shows that there is large potential available for improvement. As compared to the base case, the overall exergy destruction is significantly lower with 29.7% savings.

5. Conclusions

The cryogenic biogas upgrading, and LBM production process is a complex and energy-intensive process. To reduce high SEC and process complexity, a simulation study considering CO_2 solidification-based biogas upgrading process followed by LBM production is proposed. The proposed process is simulated in Aspen Hysys® v11. The CO_2 solidification conditions are assessed from the phase diagram of biogas. The CO_2 is

solidified at -68℃ at 5.17 bar. The high purity biomethane is liquefied assisted by PNEC. The proposed process is analysed from energy and exergy aspects. From the results, it is concluded that the process has a low SEC of 0.408 kWh/kg with 17.4% energy savings compared to the base case. In terms of exergy, the proposed process is 29.7% efficient owing to less equipment and low SEC.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2021R1A2C1092152) and by Priority Research Centers Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2014R1A6A1031189).

References

Baccioli, A., Antonelli, M., Frigo, S., Desideri, U., Pasini, G., 2018. Small scale bio-LNG plant: Comparison of different biogas upgrading techniques. Appl. Energy 217, 328–335. https://doi.org/10.1016/j.apenergy.2018.02.149

Fan, Q.H., Li, H.Y., Jia, L.X., 2018. DESIGN AND ANALYSIS OF A SMALL-SCALE BIOGAS LIQUEFACTION CYCLE. https://doi.org/10.1063/1.2908468

Hashemi, S.E., Sarker, S., Lien, K.M., Schnell, S.K., Austbø, B., 2019. Cryogenic vs. absorption biogas upgrading in liquefied biomethane production – An energy efficiency analysis. Fuel 245, 294–304. https://doi.org/10.1016/j.fuel.2019.01.172

He, T., Liu, Z., Ju, Y., Parvez, A.M., 2019. A comprehensive optimization and comparison of modified single mixed refrigerant and parallel nitrogen expansion liquefaction process for small-scale mobile LNG plant. Energy 167, 1–12. https://doi.org/10.1016/j.energy.2018.10.169

International Energy Agency, 2020. Renewables Information - Overview (2020 Edition). IEA Stat. 497.

Krich, K., Augenstein, D., Batmale, J., Benemann, J., Rutledge, B., Salour, D., 2005. Storage and Transportation of Biogas and Biomethane. Biomethane from Dairy Waste A Sourceb. Prod. Use Renew. Nat. Gas Calif. 71–80.

Naquash, A., Qyyum, M.A., Haider, J., Lim, H., Lee, M., 2021. Renewable LNG production: Biogas upgrading through CO2 solidification integrated with single-loop mixed refrigerant biomethane liquefaction process. Energy Convers. Manag. 243, 114363. https://doi.org/10.1016/j.enconman.2021.114363

Pellegrini, L.A., De Guido, G., Langé, S., 2018. Biogas to liquefied biomethane via cryogenic upgrading technologies. Renew. Energy 124, 75–83. https://doi.org/10.1016/j.renene.2017.08.007

Peng, D., Robinson, D.B., 1929. A new equation of state. Nature 123, 507.

Spitoni, M., Pierantozzi, M., Comodi, G., Polonara, F., Arteconi, A., 2019. Theoretical evaluation and optimization of a cryogenic technology for carbon dioxide separation and methane liquefaction from biogas. J. Nat. Gas Sci. Eng. 62, 132–143. https://doi.org/10.1016/j.jngse.2018.12.007

Venkatarathnam, G., Timmerhaus, K.D., 2008. Cryogenic mixed refrigerant processes. Springer.

Yousef, A.M., Eldrainy, Y.A., El-Maghlany, W.M., Attia, A., 2017. Biogas upgrading process via low-temperature CO2 liquefaction and separation. J. Nat. Gas Sci. Eng. 45, 812–824. https://doi.org/10.1016/j.jngse.2017.07.001

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50126-3

Importance of interannual renewable energy variation in the design of green ammonia plants

Nicholas Salmon[a] and René Bañares-Alcántara[a*]

[a]*Department of Engineering, University of Oxford, Parks Road, Oxford, OX1 3PJ, UK*
**rene.banares@eng.ox.ac.uk*

Abstract

Green hydrogen and ammonia are critical technologies in our decarbonisation toolkit, but at present remain more expensive than traditional energy vectors. In order to increase their competitiveness with respect to fossil fuels, many authors have optimised production systems based on hourly solar and wind profiles and a range of technologies to maximise production and minimise costs. This optimisation problem, however, is enormous in scale: it requires consideration of a large number of possible production sites and their performance over many years. Failure to consider both spatial and temporal variation in green ammonia production costs may exclude excellent locations, or include sites that are unreliable due to interannual variation. In this work, we examine three techniques which can reduce the complexity of input data: time aggregation, hierarchical clustering, and K-means clustering. We compare the suitability of each of these approaches based on the extent to which they accelerate the solution of a green ammonia plant design optimisation problem, and the error between the simplified and actual solutions to the problem. Using these simplification approaches, we demonstrate the importance of considering interannual variation in green ammonia plant design.

Keywords: Data Clustering, MILP Optimisation, Green ammonia production, Renewable energy storage

1. Introduction

The decarbonisation of national power grids poses new energy storage challenges to provide reliable power during periods of high demand, or when renewable energy availability is low. While a range of technologies can be deployed on a very large scale for storage on an hours-days timescale (e.g. demand response, battery to grid, highly interconnected electrical grids with geographical diversity of energy sources), fewer technologies are available for storage on a months-years timescale.

Storage on a months-years timescale is necessary to provide sufficient storage (i) to meet seasonal peaks in energy usage, and (ii) to account for interannual variation in renewable energy availability. In the UK, for instance, the energy shortage at the start of autumn 2021 was exacerbated by a long-term wind-drought, the scale of which had not been observed since the early 1960s (Ambrose 2021); the frequency of such wind droughts is forecast to increase with climate change (Dawkins 2019).

Green ammonia is a promising solution for storage over large timescales. Compared to liquid hydrogen, it can be stored cheaply, and with much lower boil-off. Unlike gaseous hydrogen stored in salt caverns, it has much higher density, allowing import via ship. Recent publications have optimised the cost of green ammonia production and have demonstrated its cost is rapidly falling, and will soon be competitive with other technologies (Nayak-Luke and Bañares-Alcántara 2020; Fasihi et al. 2021).

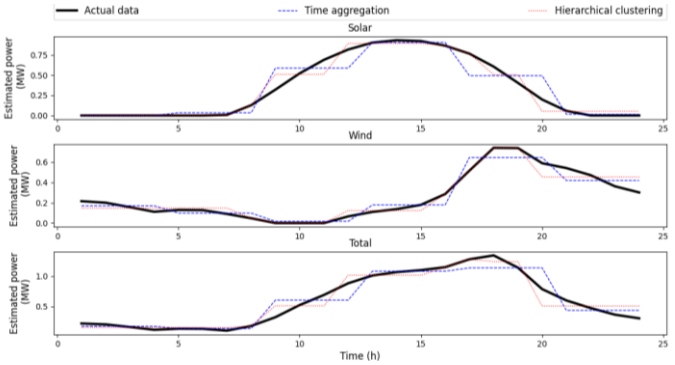


Figure 1 - Comparison of time aggregation and hierarchical clustering

However, the optimisation of green ammonia plants is computationally intensive, forcing researchers to either consider a comparatively small number of locations (Nayak-Luke and Bañares-Alcántara 2020), or to use only one year of weather data for plant design (Fasihi et al. 2021). Further complications emerge if other energy sources (e.g. electricity grids or hydro/nuclear power) are considered. Given the interannual weather variation described above, and the intended use of ammonia to store energy between years, larger volumes of weather data need to be processed to develop a robust ammonia economy.

This work explores three data clustering approaches which reduce the volume of weather data. It demonstrates that it is possible to significantly accelerate calculation without sacrificing accuracy. Several applications of accelerated calculation are then discussed.

2. Methodology

Green ammonia production uses renewable energy (from wind and/or solar plants) to power three core units: a water electrolyser to make hydrogen, an air separation unit to make nitrogen, and a Haber-Bosch loop which synthesises them into ammonia. Since it operates at high temperatures and pressures, the Haber-Bosch process requires back-up storage of power and hydrogen to sustain production above minimum rates. Here, back-up power is supplied from batteries, or from a fuel cell which cannibalises some of the stored hydrogen. The model may also connect the plant to the electricity grid, from which it can buy and sell power (capped at 175 MW by the limits of grid connection).

Rigorous existing models for ammonia plant design use hourly weather data to size each piece of equipment, subject to material and energy balances and technical constraints on equipment, mostly related to the inflexibility of the ammonia plant. However, in other energy optimisation analyses, some data clustering techniques have been attempted.

The Balmorel energy systems model, a popular open-source application, uses time aggregation, whereby larger time steps are used to reduce the data size (Wiese et al. 2018). Palys and Daoutidis (Palys and Daoutidis 2020) modelled a grid-based energy system which included ammonia that used a hierarchical approach, in which the most similar input data points are iteratively clustered until the data is reduced in size by a pre-specified factor. Not all time steps are the same size using this approach, so the model affords them different weights to determine the impact of a time-step on ammonia production, and the levels in energy and mass storage equipment. Figure 1 shows the difference between these two techniques. The former is simpler for both the coder and the computer to implement; the latter captures more dynamic variation in weather with the same number of time steps.

A third approach in the literature simplifies the data using representative days (van der Heijde et al. 2019); one technique used by Gabrielli et al. (Gabrielli et al. 2018) for selecting such days is K-means clustering, which uses a principal component analysis to classify days into clusters, and then represents each day in the cluster by its medoid.

To compare these options (time-aggregation, hierarchical aggregation and K-means clustering), we define the simplification factor (*SF*), which is the ratio of the number datapoints in the raw hourly data to the number of datapoints in the clustered data; for instance, data using a two-hourly time-step would have *SF = 2*. The model has eleven variables defined at all time steps (nine for power flows between plant components, and one for each of the inventories of the battery/hydrogen storage), nine continuous variables for the size of each unit in the process, and a binary variable indicating if grid connection is used. Thus for a model considering *n* years of data, the total number of variables is:

$$\frac{11}{SF}(8760n) + 10 \tag{1}$$

For K-means clustering, the storage inventories cannot be represented by their cluster's medoid, since this will cause discontinuities where two consecutive days are not in the same cluster. Therefore for the K-means cases, the number of variables is:

$$\left(\frac{9}{SF} + 2\right)(8760n) + 10 \tag{2}$$

An 8-core desktop computer with an i7 processor and 16 GB of RAM was used to solve the model for 701 locations in Australia (spread in a grid pattern across the country). Only three cores were parallelised; if more cores were used, very large convergence times were observed as the computer hit RAM limits. The concurrent and barrier methods of the Gurobi optimisation solver were used for the root and node relaxations respectively; these settings were found to give the fastest solutions.

3. Results

3.1. Comparison of performance

Figure 2 compares the techniques at different simplification factors to results at a simplification factor of 1. Each point represents results at one location; if the points form 45° lines, this indicates good agreement between the simplified model and actual results.

Time aggregation and hierarchical clustering perform well at low simplification factors. Performance starts to degrade around a simplification factor of 12, because the diurnal variation of solar panels begins to be smoothed into a near constant supply of electricity, which will underestimate the amount of energy storage required. As expected, the performance of time aggregation degrades more quickly than hierarchical clustering. In general, both techniques tend to slightly underestimate the LCOA; this is because the smoothing inherent to reducing the size of the time data provides more reliable electricity and therefore reduces the need for batteries, hydrogen storage, or back-up grid power.

The performance of K-means clustering is very poor at all simplification factors. While other techniques exist by which representative days can be selected, it is unlikely that a different selection of medoids would radically change the poor performance of this approach. To some extent, the poor performance is caused by the long-term nature of ammonia plant design, meaning the representative day is too short a time frame over which to base plant design. For instance, plant operation may need to be different on two days with identical weather based on the inventories of the hydrogen storage and battery.

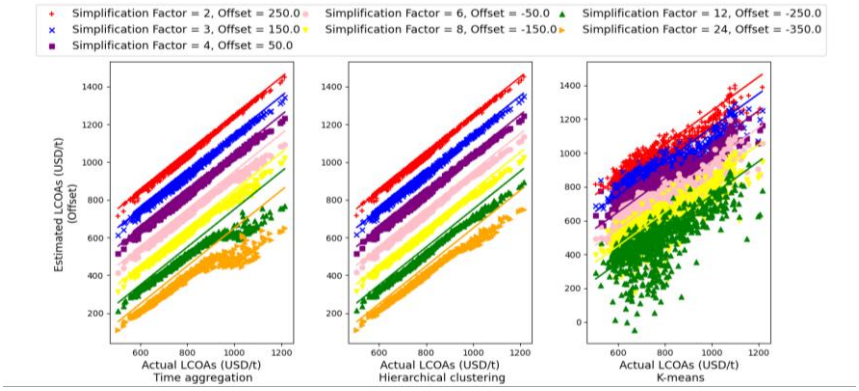


Figure 2 - x-y plots for estimated LCOA from simplified approaches compared to results from a one-hour time-step. Data are offset for readability; see legend. Each point represents a different location; offset x-y 45º lines are included for each series (a) - Left: Time aggregation. (b) - Centre: Hierarchical clustering. (c) - Right: K-means (Simplification Factor = 24 excluded for readability)

3.2. Comparison of speed

Figure 3 shows the relationship between the time taken for the model to converge in all of the 701 locations considered, and both the average and maximum error observed in the results compared to the result obtained using a one-hour time step.

Predictably, because of the results shown in Figure 2, the results for K-means clustering generate very high errors; additionally, they also take a long time to converge. This is because the dimensionality of the hydrogen and battery storage variables is not reduced by representative day clustering (since the model still needs to build a continuous storage profile). For the other two clustering techniques, higher simplification factors tend to result in shorter solution times and higher errors. In all cases, the error observed using the hierarchical clustering was lower than that observed using time aggregation at the same simplification factor, but the time required for convergence was greater.

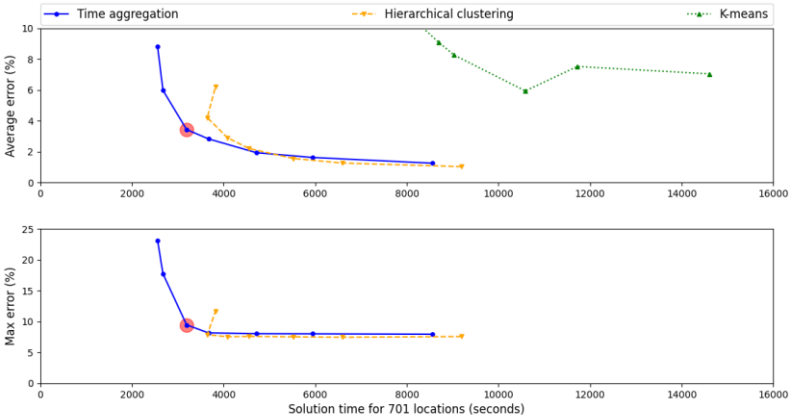


Figure 3 - Relationship between error and solution time. The highlighted marker is the option selected as optimal for further analysis. (a) – Top: Average error across all locations. (b) – Bottom: Maximum error across all locations; the axes in this subplot do not encompass K-means clustering due to the large errors observed with this method.

At low simplification factors, the majority of time taken for the model to converge is taken by the optimisation solver. However, as the simplification factor increases, the computational time required to simplify the data itself begins to increase, and the time required for the optimisation solver decreases. This limits the time taken for the model to converge to the time taken to perform the data clustering itself. Because the hierarchical clustering approach is more computationally challenging than time aggregation, this minimum limit on convergence time is higher; indeed, at the highest simplification used here (24), the model begins to take longer to converge than at the lower factor of 12.

The best option for further analysis will deliver a fast solution with an acceptably small error. The point selected is shown on Figure 3; it combines acceptable errors (~3.5% on average, 9.5% at most) with quick solution times (~3,500 s). More accurate results could have been obtained with a time penalty of ~ 20%, but the improvement is not large, particularly given the error implicit in the model due to input parameter estimation.

4. Historical Data Analysis

Green ammonia plants for energy storage need to operate under different conditions over a large time period (~30 years). Using time aggregation with a simplification factor of 8, the model was converged repeatedly using different starting years, and considering time periods of 1, 3, 5 and 10 years. Since long-term historical grid data is not available, grid connections were not allowed for these estimates. The LCOA results for a single representative location are plotted in Figure 4. They demonstrate that as more years of data are considered, the optimum value for the LCOA tends to converge around a single value; if only one year of data is used, there is a wide spread in the LCOA estimates.

Considering all locations for the cases where only one year of data was analysed, the average range between the minimum and maximum LCOAs estimated was 15% (substantially higher than the error introduced by clustering); in the worst case, it was more than a third of the total ammonia cost. When a larger number of years of data are considered, the results are much more stable over different time periods. Using ten years of input data, the average range between the minimum and maximum cases across all locations was only 2%, and the largest error observed was only 6%.

If the long term average LCOA is taken from the cases where a single year of data is used, the cheapest location for ammonia production is located near in central Western Australia. However, this site only has the cheapest production cost in ten of thirty years of individual data; in one poorly performing year, it was the 50th ranked site of the 701 considered.

Figure 4 - Optimum LCOA for a given location in South Australia using different input weather data; data are offset from each other to improve readability.

There is a weak relationship between the number of years of data used for the optimization and its outcome. In 677 of the 701 cases considered, there was a positive correlation between the number of years of data considered and the LCOA; i.e. considering more years of data slightly increases the LCOA. Although this relationship is small (on average, the LCOA estimated increases by 10 USD/t when comparing average results from ten individual years to the results from one ten-year period), it does suggest that an engineering plant designed over a single year alone will be underdesigned.

5. Conclusions

To play a robust role as an energy storage medium on a months-years timescale, green ammonia plants must consider large time periods in system design. This research analysed three techniques to reduce the size of large weather datasets to enable plant design to be optimised rapidly over large time scales and in many locations.

The results demonstrate that a 'representative day' approach (e.g. K-means clustering) for data size reduction is not suitable for plant design; too much relevant data is excluded with little benefit to convergence time. However, both time aggregation and hierarchical clustering accelerate convergence while providing meaningful results. Since hierarchical clustering itself is slow, time aggregation provided the optimum balance of speed and accuracy for this work, but hierarchical clustering may be better suited for other problems.

Using the time aggregation approach, ammonia plant designs were considered for a range of different years of data. They demonstrated that considering only a single year of data may undersize ammonia plants, and may fail to identify the optimum production location.

The accelerated solutions enabled by time aggregation could facilitate significant further research. Two options would include (a) Monte-Carlo simulation of green ammonia plant design given the significant unknowns surrounding equipment cost and performance, and (b) consideration of climate forecasting (which returns a large array of possible outcomes) in future-proofing ammonia plant designs against various possible scenarios.

References

J. Ambrose, 2021, 'What caused the UK's energy crisis?', *The Guardian*.

L. Dawkins, 2019, 'Weather and Climate Related Sensitivities and Risks in a Highly Renewable UK Energy System: A Literature Review', Met Office, Accessed May 2021. https://nic.org.uk/app/uploads/MetOffice_NIC_LiteratureReview_2019.pdf.

M. Fasihi, R. Weiss, J. Savolainen, and C. Breyer, 2021, 'Global potential of green ammonia based on hybrid PV-wind power plants', *Applied Energy*, 294: 116170.

P. Gabrielli, M. Gazzani, E. Martelli, and M. Mazzotti, 2018, 'Optimal design of multi-energy systems with seasonal storage', *Applied Energy*, 219: 408-24.

R. Nayak-Luke, and R. Bañares-Alcántara, 2020, 'Techno-economic viability of islanded green ammonia as a carbon-free energy vector and as a substitute for conventional production', *Energy & Environmental Science*, 13: 2957-66.

M.J. Palys, and P. Daoutidis, 2020, 'Using hydrogen and ammonia for renewable energy storage: A geographically comprehensive techno-economic study', *Computers & Chemical Engineering*, 136.

B. van der Heijde, A. Vandermeulen, R. Salenbien, and L.Helsen, 2019, 'Representative days selection for district energy system optimisation: a solar district heating system with seasonal storage', *Applied Energy*, 248: 79-94.

F. Wiese, R. Bramstoft, H. Koduvere, A.P. Alonso, O. Balyk, J.G. Kirkerud, A. G. Tveten, T.F. Bolkesjø, M. Münster, and H. Ravn, 2018, 'Balmorel open source energy system model', *Energy Strategy Reviews*, 20: 26-34.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50127-5

Integrating Carbon Negative Technologies in Industrial Clusters

Elizabeth J. Abraham, Dhabia M. Al-Mohannadi*, Patrick Linke

Department of Chemical Engineering, Texas A&M University at Qatar, Education City, PO Box 23874, Doha, Qatar
dhabia.al-mohannadi@qatar.tamu.edu

Abstract

Negative emission technologies (NETs) are an emerging innovation essential for tackling the climate crisis since decarbonization on its own will no longer suffice. While these technologies are crucial in limiting temperature rise by the end of the century, they bring their own set of unique feasibility and efficiency challenges. An innovative approach to overcoming these barriers is the development of eco-industrial parks (EIPs) or clusters, which can integrate several processes and technologies to achieve economic, environmental, and social benefits collectively. Furthermore, through the multiple interactions that can arise from exchanging resources and technical knowledge in these systems, technology spill overs can occur, allowing these parks to serve as a hub for process innovation. Thus, this work explores the integration of NETs into industrial clusters that can simultaneously exchange multiple resources over time. The multi-period model optimized network configurations for maximum net present value and determined park design across time. Two NETs, bioenergy with carbon capture and storage (BECCS) and direct air carbon capture and storage (DACCS), were integrated into a cluster to provide the carbon dioxide needed for methanol production, produced through either carbon hydrogenation or electrolysis. The optimization revealed that electrolysis required improved energy efficiency to compete with hydrogenation. Furthermore, significant reductions in capital costs are necessary for the NETs to become active in EIPs.

Keywords: Multi-period, Resource integration, Negative emission technologies, Industrial parks, Optimization

1. Introduction

Climate change mitigation is a daunting challenge whose slow progress now requires the reduction of greenhouse gases (GHGs) and their removal from the atmosphere to limit global warming temperature rise to less than 2 °C above pre-industrial levels (Gasser et al., 2015). While several strategies and technologies can effectively reduce GHGs, negative emissions technologies (NETs) that offset positive emissions have only begun to gain traction. To lower their atmospheric concentrations, NETs remove GHGs, particularly carbon dioxide (CO_2), from the atmosphere through various existing or proposed techniques and subsequently store them (McLaren, 2012). The extensive research on NETs currently explores different technologies, their feasibility, and comparative performance with other technologies, where two NETs, bioenergy with carbon capture and storage (BECCS) and direct air carbon capture and storage (DACCS) make up a significant portion of this literature. BECCS utilizes biomass or plant material that absorbs atmospheric CO_2 as it grows to produce bioenergy and then captures and stores any emissions released during production in geological formations (IPCC, 2018).

DACCS, on the other hand, directly extracts CO_2 from the atmosphere, typically using solid sorbents or basic aqueous solutions as the capture media, before storing them underground (Keith et al., 2018). While both NETs are in early commercialization stages, they encompass many uncertainties. Geological storage is a fundamental aspect of NETs, which though seemingly feasible at present, entails high costs and uncertainties over long-term storage (Norhasyima and Mahlia, 2018). In these scenarios, there is a proclivity to utilize and valorize the CO_2 captured (Norhasyima and Mahlia, 2018), which calls for the integration of NETs with carbon capture utilization and storage (CCUS), as shown by Tan et al. (2021) in carbon management networks integrated with enhanced weathering and biochar application.

Systematic methods that synthesize CCUS strategies by considering various CO_2 sources and sinks include optimization methods (Al-Mohannadi et al., 2016), graphical source-sink allocation approaches (Lameh et al., 2020), and others like the Mini-MAC approach that uses marginal abatement costs (Lameh et al., 2021). While these approaches focus on CO_2 as the primary material of interest, most utilization technologies require multiple material and energy resources. The recent resource integration approach developed by Ahmed et al. (2020) considers both material and energy resources simultaneously to design industrial clusters or eco-industrial parks (EIPs). This work uses a multi-period model based on this novel approach to identify challenges in integrating NETs and CCUS in clusters that can capture emissions from the atmosphere and itself.

2. Approach

The multi-period resource integration approach described in this work adopts the method proposed by Ahmed et al. (2020) to optimize industrial clusters while simultaneously considering all material and energy flows. The extension allows for assessing the evolution of such clusters over a time horizon, where each process considered, including the NETs, have a set of associated resources. The optimization is solved as a mixed-integer linear program (MILP) that determines the existence and capacities of processes along with the flow of its associated resources throughout the cluster in each period considered. The net present value (NPV) of the cluster, which accounts for the time value of money, is maximized as the objective function in this work as given by Eq. (1):

$$NPV_{EIP} = \sum_{Period} \frac{\sum_{Process} Revenue - Operating\ Cost - Capital\ Cost}{(1+Discount\ Rate)^{Number\ of\ periods\ considered}} \quad (1)$$

The proposed model has several parameters that govern resource and cash flows throughout the cluster, such as process-specific cost and performance parameters, resource line specifications, resource prices, inflation rates, and NPV discount rates. Simply put, each process has specific capital and operating cost parameters, in addition to specific mass and energy balance parameters that define the flow of resources to and from it. These resources flow in the cluster through resource lines characterized by a unique set of specifications (temperature, pressure, quality, etc.). Thus, processes obtain input resources from certain resource lines and send output resources to their corresponding resource lines. In this way, resource lines can be associated with multiple processes with the same resource specifications, thereby facilitating resource integration. Variables in this model include process capacities and flows of fresh feed and output resources in each period considered. Furthermore, quality and inequality constraints are placed on the resource line balances and variables respectively to ensure non-negative values, while binaries implement logical decisions within the model. The optimization problem has been solved using LINDO “What’sBest!16.0.2.6” in Microsoft Excel 2019.

3. Illustrative example

3.1. Process data

The multi-period model is to design an industrial cluster that must integrate two NETs, bioenergy (BEC) and direct air (DAC) carbon captures, to capture the CO_2 required to produce 500,000 tons per period (tpp) of methanol (CH_3OH). For this, the cluster considers two innovative methanol production routes, namely, the emerging hydrogenation (CHR) route and the novel infeasible CO_2 electrolysis (CEM) route. The CHR process requires hydrogen (H_2), provided from a steam methane reformer (SMR) or water electrolyzer (WSE), in addition to CO_2, while the CER only requires CO_2. Power requirements for the cluster are met by a natural gas power plant (NGP) and photovoltaic system (PV). The CHR, SMR and NGP processes each have their own carbon capture units with a 90% capture efficiency, while a sequestration unit (SQU) stores any unutilized CO_2. Lastly, an air separation unit (ASU) supplies the DAC's oxygen (O_2) demand. Data on process specifics, namely, reference products, CAPEX parameters, and maximum capacities, are listed in Table 1, where the maximum capacity indicates the maximum allowable capacity of a process across the periods considered.

Table 1: Process reference products, CAPEX parameters and operational capacities

Resource	Reference Product (RP Unit)	CAPEX Parameter ($/RP Unit)	Maximum Capacity (RP Unit per period)
ASU	O_2 (t)	20.30	80,000
BEC	CO_2 (t)	34.00	3,000,000
CER	CH_3OH (t)	13.03	500,000
DAC	CO_2 (t)	102.06	1,000,000
CHR	CH_3OH (t)	11.64	500,000
CHR Capture	CO_2 (t)	6.98	60,000
NGP	Electricity (kWh)	10.00	20,000,000,000
NGP Capture	CO_2 (t)	12.49	750,000,000
PV	Electricity (kWh)	20.00	20,000,000,000
SQU	CO_2 (t)	9.02	1,250,000
SMR	H_2 (t)	13.09	125,000
SMR Capture	CO_2 (t)	1.08	1,250,000
WSE	H_2 (t)	623.20	125,000

The only inputs to the cluster are air, biomass, CO_2, natural gas, water, additives for DAC, and other utilities besides power. The given cluster was analyzed across three periods, where each period spans five years. The operational capacities and capital cost or CAPEX parameters remain fixed across all periods for all processes except BEC and DAC, whose capacities increase with their expected increase in deployment capacities. Specifically, the capture capacities of BEC increases from 1,000,000 tpp to 3,000,000 tpp, while that of DAC increases from 20,000 tpp to 1,000,000 tpp across the periods considered. Prices and process parameters for the BEC, CER, DAC, NGPP and SMR can be found in Table 2, while those for all other processes are obtained from Abraham et al. (2021). The NET parameters in Table 2 were acquired from Bhave et al. (2017) and Keith et al. (2018), while data for the CER, NGPP and SMR processes were found in Jouny et al. (2018), Biliyok and Yeung (2013), and Spath and Mann (2001) respectively.

Table 2: Process parameters (in Unit/RP Unit) and resource prices in ($/Unit)

Resource (Unit)	Price	BEC	CER	DAC	NGPP	SMR
Air (t)	-	-4.98	-	-1467.84	-0.01	-
BEC emissions (t)	-	4.72	-	-	-	-
Biomass (t)	60	-0.74	-	-	-	-
Calcium carbonate (t)	330	-	-	-0.02	-	-
Carbon dioxide (t)	40	1.00	-0.73	1.00	-	-
Hydrogen (t)	900	-	0.02	-	-	1.00
Methanol (t)	400	-	1.00	-	-	-
Natural Gas (t)	136	-	-	-0.12	-	-2.91
NGPP emissions (t)	-	-	-	-	0.01	-
Oxygen (t)	-	-	0.39	-0.35	-	-
Slurry (t)	-	-	-	-5.30	-	-
SMR emissions (t)	-	-	-	-	-	10.73
Treated air (t)	-	-	-	1474	-	-
Wastes (t)	-	-	-	0.70	-	0.20
Waste water (t)	-	-	-	21.58	-	-
Water (t)	0.02	-	-0.68	-3.11	-	-14.10
Cooling water (t)	0.03	-	-	-3.25	-0.11	-
Electricity (kWh)	-	574	-11532	-366	1.00	-316
MP Steam (t)	-	-	-	-	-	5.08
Process water (t)	0.02	-	-	-20.36	-	-4.80

3.2. Results

For illustrative purposes, the breakthroughs needed in process parameters were analyzed at inflation and discount rates of 10% and 5%, respectively. Under various scenarios, the optimization revealed different designs to meet methanol production demands over time. The activated processes in each scenario and period are shown in Figure 1.

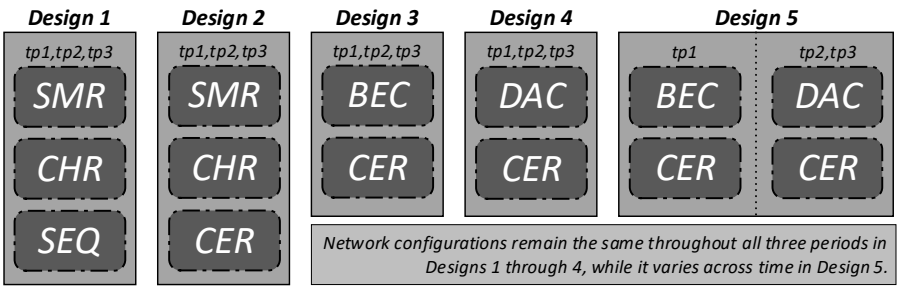


Figure 1: Activated processes for the designs analyzed

The optimization identifies methanol production via hydrogenation or CHR, where H_2 is provided by SMR and CO_2 by SMR capture, to be most profitable across time. Sequestration is a fundamental part of this design scenario, called D1, in storing the CO_2 captured from the CHR and SMR processes. Thus, when sequestration is unavailable, the cluster configuration changes to a new design, D2. Here, instead of being sequestered, the CO_2 captured from SMR unused in CHR, and the CO_2 captured from CHR itself are utilized in the CER or carbon electrolysis process. It is clear from these two cases, where

the cluster design remains fixed across all periods considered, that the NETs are not an attractive CO_2 source, even as both methanol production routes activate.

To bring the NETs online, solely improving only their prospects, was thus not sufficient due to the high costs of the auxiliary processes supporting them. Of the two production routes, CER has a significantly higher power demand than CHR, considering even the additional units CHR must activate for H_2 production, carbon capture, and sequestration. Furthermore, with WSE being an expensive process, H_2 will continue to come from SMR, whose emissions, along with those from CHR, will always need to be captured and sequestered. With sequestration constrained, these captured emissions cannot leave the cluster and must be utilized in the CER process. In this way, methanol production via both routes prevents the activation of the NETs since there is no need to obtain CO_2 from the atmosphere as the cluster itself generates the required amount. Therefore, the CER process must become the sole methanol production route to integrate the NETs into the cluster, and to do so, the CER process requires a 72% reduction in its power requirement.

With the power efficiency of CER improved, design scenarios 3 and 4 aimed to activate either of the NETs as the sole source of CO_2. BEC met the CO_2 demand for all three periods with a 98% capital cost reduction in design D3. On the other hand, even with a 90% reduction in capital costs, DAC only fully supplied the second and third periods in design D4. DAC partially met the demands of the first period, however, due to its capacity limits, the cluster's CO_2 feed met the rest. When the model integrated both NETs, supplies of the first period were met by BEC, while DAC covered the remaining periods as in design D5. The capacity limitations of DAC prevent its activation in the first period of D5; however, in later periods when both NETs can meet the CO_2 demand, the DAC is more feasible due to its lower operating costs. The capacities across all three periods for each of the designs are summarized in Table 3.

Table 3: Cumulative process capacities across all periods for network configurations

Network	CHR (t CH_3OH)	CER (t CH_3OH)	BEC (t CO_2)	DAC (t CO_2)	NPV (Million $)
N1	1,500,000	-	-	-	734
N2	1,163,185	336,815	-	-	696
N3	-	1,500,000	1,092,275	-	772
N4	-	1,500,000	-	748,183	773
N5	-	1,500,000	364,092	728,184	774

To achieve such significant capital reductions is a challenge that requires a thorough understanding of the technology and parameters that define its efficiency such as selectivity and yield. Understanding these parameters will aid in identifying the pathways by which these cost reductions can be achieved. Thus, by incorporating these parameters into the multi-period model, it can therefore, determine the learning curves that can make these technologies competitive (Feriolia et al., 2009).

4. Conclusion

EIPs are an opportune means to integrate NETs with CCUS to derive greater collective benefits. The feasibility of integrating these technologies in EIPs was analyzed from a holistic perspective in terms of cash and resource flows as illustrated through the multi-

period resource integration described here. The optimization determined that CER needs more energy efficiency, while both BEC and DAC require a reduction in capital costs to be deployed successfully. Thus, through the multi-period analysis, the proposed approach can analyze the deployment of emerging technologies and their varying parameters simultaneously considering operations in other periods during optimization, and without performing numerous iterations. Furthermore, the model can be extended to incorporate uncertainties of different parameters to create more resilient and sustainable systems.

References

E.J. Abraham, F. Ramadan, D.M. Al-Mohannadi, 2021, Synthesis of Sustainable Carbon Negative Eco-Industrial Parks, Frontiers in Energy Research, 9, 345.

D.M. Al-Mohannadi, P. Linke, 2016, On The Systematic Carbon Integration of Industrial Parks for Climate Footprint Reduction, Journal of Cleaner Production, 112, 5, 4053-4064.

R. Ahmed, S. Shehab, D.M. Al-Mohannadi, P. Linke, 2020, Synthesis of Integrated Processing Clusters, Chemical Engineering Science, 221, 115922.

A. Bhave, R.H.S. Taylor, P. Fennell, W.R. Livingston, N. Shah, N. Mac Dowell, J. Dennis, M. Kraft, M. Pourkashanian, M. Insa, J. Jones, N. Burdett, A. Bauen, C. Beal, A. Smallbone, J. Akroyd, 2017, Screening and Techno-economic Assessment of Biomass-Based Power Generation with CCS Technologies to Meet 2050 CO_2 Targets, Applied Energy, 190, 481–489.

C. Biliyok, H. Yeung, 2013, International Journal of Greenhouse Gas Control, 19, 396–405.

F. Feriolia, K. Schoots, B.C.C. van der Zwaanab, 2009, Use and Limitations of Learning Curves for Energy Technology Policy: A Component-learning Hypothesis, Energy Policy, 37, 7, 2525-2535.

T. Gasser, C. Guivarch, K. Tachiiri, C.D. Jones, P. Ciais, 2015, Negative Emissions Physically Needed to Keep Global Warming Below 2 °C, Nature Communications, 6, 1, 1–7.

IPCC, 2018, Global Warming of 1.5 °C.

M. Jouny, W. Luc, F. Jiao, 2018, General Techno-Economic Analysis of CO_2 Electrolysis Systems, Industrial and Engineering Chemistry Research, 57, 6, 2165–2177.

D.W. Keith, G. Holmes, D. St. Angelo, K. Heidel, 2018, A Process for Capturing CO_2 from the Atmosphere, Joule, 2, 8, 1573–1594.

M. Lameh, D.M. Al-Mohannadi, P. Linke, 2020, Graphical Analysis of CO_2 Emissions Reduction Strategies, Cleaner Engineering and Technology, 1, 100023.

M. Lameh, D.M. Al-Mohannadi, P. Linke, 2021, Minimum Marginal Abatement Cost Curves (Mini-MAC) for CO_2 Emissions Reduction Planning, Clean Technologies and Environmental Policy, https://doi.org/10.1007/s10098-021-02095-y

D. McLaren, 2012, A Comparative Global Assessment of Potential Negative Emissions Technologies, Process Safety and Environmental Protection, 90, 489–500.

R.S. Norhasyima, T.M.I. Mahlia, 2018, Advances in CO_2 Utilization Technology: A Patent Landscape Review, Journal of CO2 Utilization, 26, 323–335.

P.L. Spath, M.K. Mann, 2001, Life Cycle Assessment of Hydrogen Production via Natural Gas Steam Reforming.

R.R. Tan, K.B. Aviso, S. Bandyopadhyay, 2021, Pinch-Based Planning of Terrestrial Carbon Management Networks, Cleaner Engineering and Technology, 4, 100141.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50128-7

Flexibility analysis of chemical processes considering overlaying uncertainty sources

Christian Langner[a,*], Elin Svensson[b], Stavros Papadokonstantakis[c], Simon Harvey[a]

[a] *Chalmers University of Technology, Gothenburg, Sweden*
[b] *CIT Industriell Energi AB, Gothenburg, Sweden*
[c] *TU Wien, Vienna, Austria*
* *christian.langner@chalmers.se*

Abstract

Chemical processes are often subject to uncertainty. Consequently, the evaluation of the flexibility of a process with respect to variations in inlet conditions is vital to identify bottlenecks in current process flowsheets or new flowsheet design proposals. The flexibility index is a well-established concept to perform flexibility analysis of chemical processes. In this paper, we propose novel reformulations of the flexibility index problem to account for overlaying uncertainty sources which interfere with each other. The aim of the paper is to identify and define overlaying uncertainty sources and to overcome shortcomings of existing approaches when calculating the flexibility index in such situations.

Keywords: Flexibility Analysis, Flexibility Index, Overlaying Uncertainty, Types of Uncertainty, Chemical Process Design.

1. Introduction

Flexibility analysis of chemical processes has been investigated since the 1980s when the flexibility index was introduced by Swaney and Grossmann (1985). Although established in the 1980s, the mathematical formulation of the flexibility index has also been applied in recent publications (e.g., da Silva et al. 2021). The flexibility index is defined as the ratio between the maximum feasible variation range (defined by operational constraints) and the expected variation range. Therefore, feasibility is guaranteed if the flexibility index is larger than or equal to 1.

Several extensions and reformulations have been suggested in literature to adapt the original formulation by Swaney and Grossmann (1985) to specific applications. Pulsipher and Zavala (2018) suggested to incorporate multivariate Gaussian random variables in the flexibility index problem to define expected variation ranges with ellipsoidal shape (compared to the hyperrectangular shape used traditionally) in order to capture correlations. In addition, Langner et al. (2021) suggested to study the geometrical shape of the distribution of operating points to explicitly define the expected variation range using linear boundary functions. Ochoa and Grossmann (2020) extended the original flexibility index problem to account also for uncertain parameters which cannot be measured and therefore cannot be subject to control actions during operation.

Traditionally, when formulating the flexibility index problem, the expected variation range is formulated for each uncertain parameter by accounting for the largest positive and/or negative deviation from a nominal operating point which has been observed

(retrofit problem) or which is expected (greenfield problem) over a specific timeframe (e.g., lifetime of a plant). Consequently, the expected variation range is independent of time.

However, this is problematic in the case when sources of uncertainty interfere with each other, e.g., when the flexibility with respect to short-term operational disturbances is affected by an uncertain singular event (see e.g., Marton et al. 2021). Examples of such singular/rare events can be a switch in feedstock, a change of operational parameters required to comply with new emission legislation and/or a change in the production rate. All such events have in common that they can have a significant impact on the nominal operating conditions, i.e., the nominal operating point is likely to vary temporarily or even change permanently. On the other hand, it is very unlikely that short-term operational disturbances will disappear when the nominal operating point is subject to uncertainty. Consequently, in such cases we face overlaying sources of uncertainty which interfere with each other, meaning these singular/rare events divide the above-mentioned timeframe into intervals which need to be accounted for in the flexibility index problem.

This paper discusses possible approaches to account for overlaying uncertainty sources in the flexibility index problem and present the necessary reformulations of the flexibility index problem as originally reported by Swaney and Grossmann (1985).

2. Flexibility index and suggested reformulations

The formulation of the flexibility index problem reported by Swaney and Grossmann (1985) is given in Eq. (1) where d is the vector of design variables, z relates to the control variables, and the varying inlet conditions or uncertain parameters are denoted by θ.

$$\begin{aligned} &FI = \max \delta \\ &s.t. \max_{\theta \in T(\delta)} \min_{z} \max_{j \in J} f_j(d, z, \theta) \leq 0 \\ &\quad T(\delta) = \left\{ \theta_i \middle| \theta_{i,N} - \delta \Delta\theta_i^- \leq \theta_i \leq \theta_{i,N} + \delta \Delta\theta_i^+ \right\} \forall\, \theta_i \in \theta \\ &\quad \delta \geq 0 \end{aligned} \tag{1}$$

The maximum feasible variation range is described by $T(\delta)$ which states that if $\delta = 1$, the maximum feasible and the expected variation range are similar. Hereafter, we include overlaying uncertainty sources (which interfere with each other) in the original formulation of the flexibility index (see Eq. (1)) and present the shortcomings. We then derive reformulations to the original flexibility index problem to overcome these shortcomings. For readability, the uncertainty sources are classified as shown in Table 1.

Table 1: Different classes of uncertainty based on the source or origin of the uncertainty

Conventional operational disturbances (included in the original flexibility index formulation, see Eq. (1))	Uncertainty class A
"Other" sources of uncertainty (i.e., nominal operating point varies or changes)	Uncertainty class B

2.1. Original flexibility index formulation and overlaying uncertainty sources

In line with the original formulation of the flexibility index problem, it may be intuitive to include (additional) uncertainty sources (of any kind and nature) in a similar fashion to that proposed for uncertainty class A by Swaney and Grossmann (1985). Thus, the vector

θ would represent all uncertain parameters (of class A and class B: $\theta_{class\ A}$ and $\theta_{class\ B}$) and $T(\delta)$ is formulated as in Eq. (1). This approach yields a single scalar which expresses the maximum feasible variation/change for each uncertain parameter in θ. Note that a physical uncertain parameter (e.g., an uncertain temperature) may be present in both sets, $\theta_{class\ A}$ and $\theta_{class\ B}$, while the expected variation range or change ($\Delta\theta^-$ and $\Delta\theta^+$) differs for each class. However, with such an approach it is not possible to gain information on the feasibility of operational short-term disturbances (class A) when the nominal operating point varies or changes (class B) since the uncertainty sources are analyzed independently of each other. Such information is, however, vital in design and planning processes and to overcome this problem, Eq. (1) needs to be formulated for different discrete nominal points (θ_N) and the resulting formulations can then be solved in an iterative scheme. This iterative scheme can be impractical and time-consuming; thus, we suggest reformulating Eq. (1).

2.2. Suggested reformulation of the flexibility index problem

In our approach, we aim to find the maximum feasible variation/change for uncertain parameters of class B which ensures that a pre-defined flexibility target of the uncertain parameters of class A (operational short-term disturbances) is feasible, e.g., the expected variations of the uncertain parameters of class A are exactly feasible. Thus, we distinguish between the two uncertainty classes and suggest reformulating Eq. (1) to Eq. (2).

$$
\begin{aligned}
&FI = \max \delta_B \\
&s.t. \max_{\theta \in T(\delta_A, \delta_B)} \min_{z} \max_{j \in J} f_j(d, z, \theta_{class\ A}, \theta_{class\ B}) \leq 0 \\
&T(\delta_A, \delta_B) = \begin{Bmatrix} \{\theta_i | \theta_{i,N} - \delta_A \Delta\theta_i^- \leq \theta_i \leq \theta_{i,N} + \delta_A \Delta\theta_i^+\} \ \forall\ \theta_i \in \theta_{class\ A} \\ \{\theta_k | \zeta_k - \delta_A \Delta\theta_{k,A}^- \leq \theta_k \leq \zeta_k + \delta_A \Delta\theta_{k,A}^+\} \\ \{\theta_k | \theta_{k,N} - \delta_B \Delta\theta_{k,B}^- \leq \zeta_k \leq \theta_{k,N} + \delta_B \Delta\theta_{k,B}^+\} \end{Bmatrix} \forall\ \theta_k \in \theta_{class\ B} \\
&\delta_A = specified\ target, \delta_B \geq 0
\end{aligned}
\tag{2}
$$

Eq. (2) includes several reformulations compared to Eq. (1). When distinguishing between uncertain parameters of class A and of class B, the maximum feasible variation/change for each class is respected individually by defining a scalar δ for each class (δ_A, δ_B). As aforementioned, we suggest searching for the maximum feasible variation/change for uncertain parameters of class B which allows for a pre-defined flexibility target of the uncertain parameters of class A. Consequently, the constraint $\delta_A = specified\ target$ is included in Eq. (2) while searching for the maximum value of δ_B is formulated as the objective function. Furthermore, the uncertainty of the nominal value(s) for the uncertain parameters of class B is respected by defining the variables $\zeta_k\ \forall\ \theta_k \in \theta_{class\ B}$. Consequently, the first and the second line in $T(\delta_A, \delta_B)$ guarantee that expected short-term disturbances remain feasible when the nominal operating point varies or changes. The maximum feasible variation/change of the nominal operating point is found by including the third line in $T(\delta_A, \delta_B)$.

For solving Eq. (2), we suggest using the active constraint strategy proposed for solving Eq. (1) by Grossmann and Floudas (1987). This requires that an upper bound for the operational flexibility target value is pre-defined ($\delta_A = specified\ target$). This upper bound can be obtained in a first step by considering only the uncertain parameters of class A and thus formulating and solving Eq. (1). In a second step, the uncertain parameters of class B are added, and Eq. (2) can be formulated and solved.

If the expected variation range of operational uncertainty (class A) is itself expected to change (when the nominal operating point varies/changes), $\Delta\theta_{k,A}^{-}$ and $\Delta\theta_{k,A}^{+}$ can be defined as functions depending on the nominal operating point ζ_k (e.g., if the expected variation range of short-term disturbances is expected to be a percentual share of the nominal operating point, such as ± 5%).

3. Illustrative Example

Consider a process with a feasible operating region that can be described by a set of given constraints (Eq. (3a) to (3f)) where θ_1 and θ_2 correspond to two uncertain parameters and x_1 denotes an additional state variable.

$$\theta_1 - 2x_1 = 0 \tag{3a}$$
$$2x_1 - \left(67 - \frac{2}{3} * \theta_2\right) \leq 0 \tag{3b}$$
$$10 - 0.25\theta_2 - \theta_1 \leq 0 \tag{3c}$$
$$\theta_1 - (0.5\theta_2 + 5) \leq 0 \tag{3d}$$
$$2.5 + 0.05\theta_2 - 2x_1 \leq 0 \tag{3e}$$
$$\theta_1 - (22.5 - 0.05\theta_2) \leq 0 \tag{3f}$$

Let us further assume a nominal point ($\theta_{1,N}$, $\theta_{2,N}$) = (12, 65) and due to operational disturbances, θ_1 and θ_2 are expected to vary in the range of (±4, ±8). Following the suggested two-step solution algorithm, Eq. (1) is formulated and solved for the illustrative example, and the flexibility index is calculated to be $FI_{Eq1} = \delta_{A,max} = 1.25$.

For illustrative purposes, we further assume that the nominal values of θ_1 and θ_2 are expected to change due to a singular event, but that the potential changes are not independent, i.e., ζ_1 is correlated with ζ_2. We assume that the maximum (expected) change of ζ_2 is $\Delta\theta_{2,B,max}^{-} = -40$ and the corresponding change of ζ_1 is given by $\zeta_1 = 23 - 0.17 * \zeta_2$. Solving Eq. (2) for $\delta_A = 1$ (expected operational disturbances are feasible), yields $\delta_B = 0.6$ meaning that ζ_2 may only decrease by $\Delta\theta_{2,B,feasible}^{-} = -24.0$ to ensure that the operational disturbances, (±4, ±8), can be handled by the process equipment also at the new operating point. The analysis is illustrated in Figure 1 in which the expected variation range of the operational disturbances (uncertainty class A) are shown as a black, hatched rectangle around the current nominal operating point, and the expected change of the operating point (uncertainty class B) is shown as solid black line. Additionally, the maximum feasible change of the nominal operating point is indicated by a grey arrow pointing from the current nominal operating point to the shifted operating point. The potential change of the nominal operating point is limited by the expected operational disturbances (visualized as grey, hatched rectangle around the shifted nominal point).

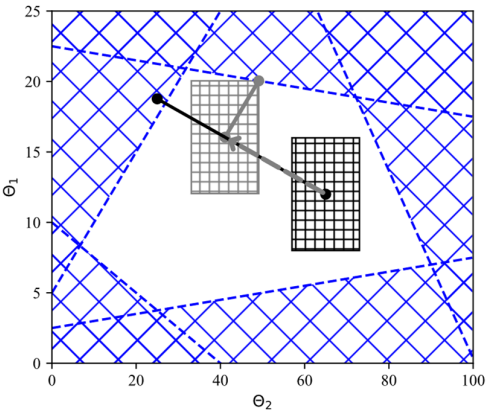


Figure 1: Visualization of flexibility analysis for illustrative example (operational disturbances overlaying with change of nominal operating point)

By means of an additional case, we now illustrate how the proposed approach can be used when the expected variation range of operational uncertainty (class A) is expected to change (when the nominal operating point varies/changes). An illustrative example, very similar to the previously presented example was developed, but in contrast to the previous example the operating variation does not remain constant but depends on the nominal point, i.e., $(\Delta\theta_{1,A}, \Delta\theta_{2,A}) = (\pm 0.3\zeta_1, \pm 0.3\zeta_2)$. The analysis of the additional case is shown in Figure 2.

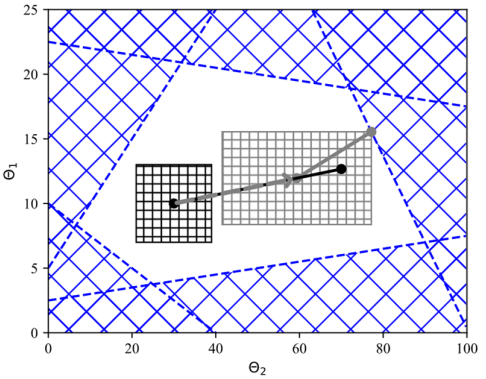


Figure 2: Visualization of flexibility analysis for additional case of illustrative example (operational disturbances depending on nominal point).
Solution of Eq (1): $FI_{Eq1} = \delta_{A,max} = 1.33$
Correlation between ζ_1 and ζ_2: $\zeta_1 = 8 + 0.06 * \zeta_2$
Solution of Eq. (2) for $\delta_A = 1$: $\delta_B = 0.73$

4. Industrial case study

To illustrate the practical application of the proposed reformulations, an industrial case study taken from Marton et al. (2021) was investigated. The case study involves three process streams which are part of two different process units of an oil refinery located in Sweden. The flow sheet of the case study is shown in Figure 3 and process data (including the variation range) is presented in Table 1 in Marton et al. (2021). For the analysis, it was assumed that the UA-value of heat exchangers HX 1-1 and 1-2 is 850 kW/K and 110kW/K for the exchanger HX 2. With the given design data and the variation data in Table 1 in Marton et al. (2021), the flexibility index was calculated to be $FI_{Eq1} = \delta_{A,max} = 1.38$ using Eq. (1). In Marton et al. (2021), it was also discussed that the refinery operating company is considering introducing biomass feedstock in the oil refinery, which would cause a substantial increase of the nominal flow rates of streams 1 and 2 (see Figure 3). Assuming a flow rate increase of streams 1 and 2 of 50-100% compared to current operation (numbers given in Figure 3), the reformulations suggested in this paper can be used to analyse if the process configuration would remain feasible (operational disturbances remain constant). Formulating and solving Eq. (2) yields $\delta_B = 0.5$ (for $\delta_A = 1$) for an expected increase in flow rates of streams 1 and 2 of 100%. Consequently, it can be concluded that the process configuration would be able to handle an increase of 50% in the flow rates while for any larger increase in flow rates the expected operational short-term uncertainty may not be feasible.

5. Influence of operational flexibility target value

The influence of the operational flexibility target value ($\delta_A = specified\ target$ in Eq. (2)) was investigated by means of sensitivity analyses. Eq. (2) was solved for different target values of δ_A for both cases of the illustrative example and the case study. We identified non-linear dependencies between δ_A and δ_B for the case study (due to non-linearities in the constraints describing the equipment) and the case when operational disturbances depend on the nominal point (percentual change itself is non-linear). For the first case of the illustrative example, we identified a linear dependency between δ_A and δ_B which is dependent on the limiting or active constraint (i.e., when the limiting/active constraint changes, the parameters of the linear dependency also change).

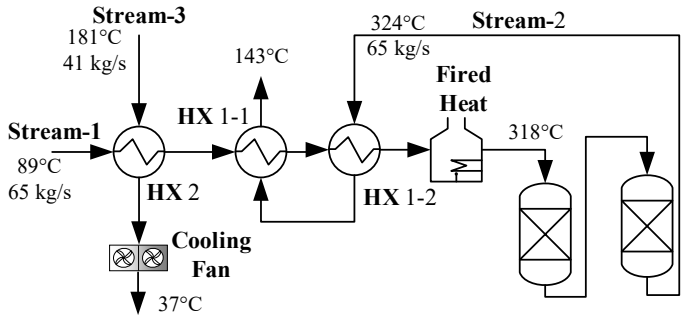


Figure 3: Flowsheet of heat exchanger network case study adapted from Marton et al. (2021).

6. Conclusion

This paper presented novel reformulations of the flexibility index problem to efficiently account for overlaying uncertainty sources which interfere with each other. We defined that overlaying uncertainty sources are present when the timeframe for which the flexibility analysis is aimed to be valid needs to be divided into intervals. We then established the need for the aforementioned reformulations by firstly classifying different (possibly overlaying) uncertainty sources and secondly presenting the shortcomings of the original flexibility index formulation when exposed to such overlaying uncertainty sources (i.e., iterative scheme). In this paper, we focused on the deterministic calculation of the flexibility index (i.e., based on expected variations). However, distinguishing between overlaying uncertainty sources should also be considered when operational disturbances and/or uncertainty of the nominal operating point are expressed via probability density functions. An illustrative example and an industrial case study were investigated and for both, the feasible change of the nominal operating point (due to a singular event) could be determined respecting short-term operational disturbances.

References

R. E. Swaney, I. E. Grossmann, 1985. An index for operational flexibility in chemical process design. Part I: Formulation and Theory. AIChE Journal 31 (4).

P. R. da Silva, M. E. Aragão, J. O. Trierweiler, and L. F. Trierweiler, 2021. A systematic approach for flexible cost-efficient hydrogen network design for hydrogen management in refineries, Chem. Eng. Res. Des., 172.

S. Marton, C. Langner, E. Svensson, and S. Harvey, 2021. Costs vs. flexibility of process heat recovery solutions considering short-term process variability and uncertain long-term development, Front. Chem. Eng.

I. E. Grossmann and C. A. Floudas, 1987. Active constraint strategy for flexibility analysis in chemical processes," Comput. Chem. Eng., 11 (6).

J. L. Pulsipher and V. M. Zavala, 2018. A mixed-integer conic programming formulation for computing the flexibility index under multivariate gaussian uncertainty," Comput. Chem. Eng., 119.

M. P. Ochoa and I. E. Grossmann, 2020. Novel MINLP formulations for flexibility analysis for measured and unmeasured uncertain parameters, Comput. Chem. Eng., 135.

C. Langner, E. Svensson, and S. Harvey, 2021. Flexibility analysis of chemical processes considering dependencies between uncertain parameters, Comput. Aided Chem. Eng., 50.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50129-9

Flexible and Sustainable Methanol Production Including Option with Green Hydrogen

Nga Thi Quynh Do[a*], Stephane Haag[a], Frank Castillo-Weltter[a], Armin Gunther[b]

[a]*Air Liquide Innovation Campus, Gwinnerstraße 27-33, 60388 Frankfurt, Germany*
[b]*Air Liquide Engineering & Construction, Olof-Palme-Str. 35, 60439 Frankfurt, Germany*
*ngathiquynh.do@airliquide.com

Abstract

The reduction of CO2 footprint, the valorisation of different carbon containing sources (stranded gas, biomass, plastic waste, high carbon-containing streams released by plants etc.) as well as the usage of renewable electricity are of increasing interest, especially in relation to global warming and the limitation of carbon containing natural resources. In this context, methanol is a key molecule; which can satisfy both environmental requirements and economic constraints; for energy storage, for clean fuels and as a building block for producing high value chemicals.

The utilization of hydrogen produced from electrolysis using renewable energy directly, the use of unconventional feedstock (CO2, off-gases etc.) with the changing of gas compositions and a higher focus on smaller units for methanol production are challenging and not only from an economical point of view but also regarding the catalyst degradation, the heat management, the by-product formation and the integration in an existing industrial complex. Therefore, the innovation philosophy has not only focused on finding a solution for the methanol synthesis loop but for the global process via new types of equipment (reactor, distillation, heat exchanger etc.), process intensification (layer management, intermediate product separation etc.), and smaller footprint (smaller equipment, new arrangement etc.) to optimize the CAPEX and OPEX.

In this contribution, an update of R&D methanol pilots is provided with some highlights on recent developments of the multistage reactor. These include the main pilot's features and capability, the campaign for validation of different scenarios i.e operating conditions, different design, model validation, by-product formation, dynamics, etc which specifically aims for off-gas and CO2 rich gases valorisation from the steel industry. The availability of the conventional pilot and the flexibility of the new developed methanol pilot at Air Liquide Innovation Campus Frankfurt paves a way for a physical and digital connection with advanced electrolysis pilots using renewable energy from the grid for H2 production and consequently to demonstrate a complete flexible and sustainable concept from power to X (i.e Fuel).

Keywords: Methanol; electrolysis; off-gas; carbon dioxide; hydrogen

1. Introduction

Carbon footprint reduction appears in the past few years as a key topic in multifaceted discussion in different industries and governments in reaching carbon neutrality. Until 2021, the 44 countries and the European Union have committed to meet a net-zero emissions target which accounts for around 70% of global CO2 emissions and GDP by 2050 (IEA, Net Zero by 2050). Supporting the sustainable development and climate

objectives, Air Liquide pledges to reach carbon neutrality by 2050 by increasing the use of low-carbon electricity for operations, implementing innovative carbon capture technologies, optimizing supply chains and improving the efficiency of our production units and products portfolio (Figure 1).

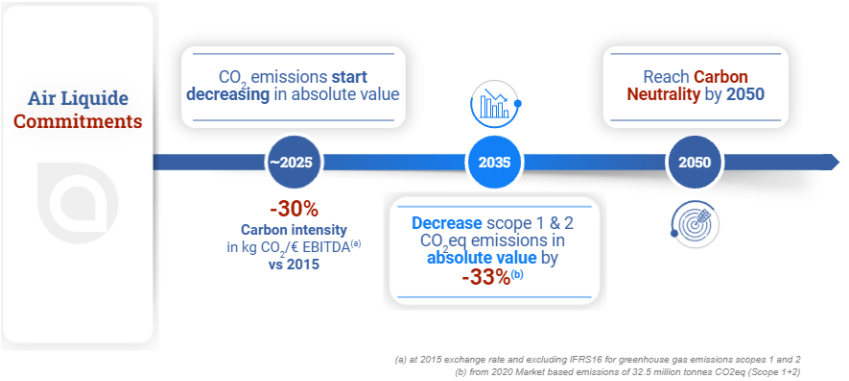


Figure 1: Air Liquide commitment and stepwise approach in reaching carbon neutrality until 2050 (Air Liquide Climate Objective)

Beside the direct electrification of renewable energy, it is believed that hydrogen from electrolysis is an excellent solution for the growth of renewable energy, enabling a novel way of renewable energy distribution to different industrial sectors i.e mobility, heat, chemical, steel… As renewable energy sources are widely dispersed and intermittent by nature, a storage and transportation solution is necessary to ensure the supply chain and the best green energy exploitation. Among different storage solutions, green hydrogen can be further combined with a C1 source i.e CO2, off-gas/flue gas with CO/CO2 content in producing methanol as a solution for carbon footprint reduction.

Air Liquide has a strong history in methanol production covering from R&D to various commercial products in its portfolio including the Lurgi Methanol and Lurgi MegaMethanolTM technologies. Since 1969, more than 68 licenses have been sold globally for a capacity of more than 55,5 million tonnes per year of methanol production. In a close collaboration with Engineering and Construction, R&D is well equipped with competency and pilots supporting different methanol research activities and benchmarking (N.T.Q. Do et al 2020; T. Oelmann et al, 2020). In the past years, different extensive studies on conventional based methanol and CO2-based methanol synthesis have been done at the Innovation Campus Frankfurt covering process development, new concept testing, catalyst validation, kinetics model and by-product model development (T. Svitnic et al, 2020) and strong focus recently is on off-gas valorisation and CO2 based methanol using green hydrogen (EU i3upgrade project funded by RFCS).

This paper aims to give an update on R&D methanol pilots with some highlights on recent developments of the multistage reactor development. These include the main pilot's features and capability, the campaign for validation of different scenarios i.e operating conditions, different design, model validation, byproduct formation, dynamics, etc which mainly aims for off-gas and CO2 rich gases valorization from the industry.

The availability of the conventional pilot and the flexibility of the new developed methanol pilot at Air Liquide Innovation Campus Frankfurt paves the way for a physical and digital connection with advanced electrolysis pilots using renewable energy from the

grid for H2 production and consequently to demonstrate a complete flexible and sustainable concept from power to X (i.e Fuel).

2. Results and Discussion

2.1 Conventional Methanol Unit

Figure 2: Conventional MeOH pilot plan

The first methanol pilot was constructed more than 10 years ago featuring MegaMethanol set up with a water cooled reactor (WCR) and gas cooled reactor (GCR). The maximum methanol capacity of this pilot plant is about 6 kg/h. The unit is well equipped with a thermocouple for tracking temperature along the reactor length, control and analytical tools for composition analysis. It was used intensively to validate different concepts and catalyst testing in the past. Examples can be seen in figure 3 in which two catalyst testing concepts are presented.

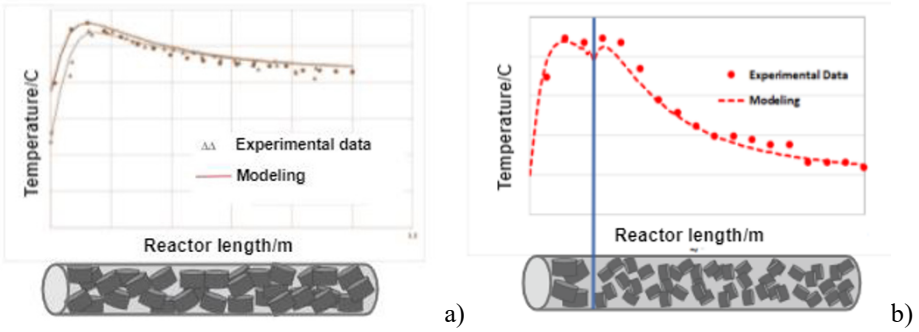


Figure 3: Concept testing at our first methanol pilot plant a) one layer catalyst testing b) two layers of catalyst testing

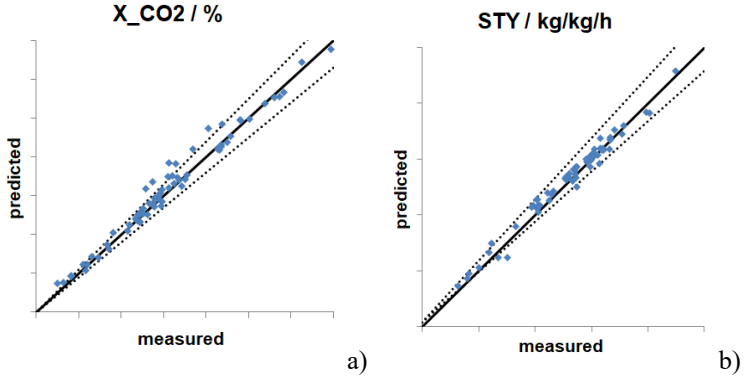


Figure 4: The fitting of modeling using CO2 kinetics and measure data a) CO2 conversion b) Space Time Yield (STY)

The temperature profile of one layer catalyst (figure 3a) and layer management (figure 3b) concept is shown in which one type of catalyst and two types of catalyst are installed inside the reactor correspondingly. There is an excellent agreement between experimental data and modeling of the two concepts which sets a good basis for scale up and

commercialization (N.T.Q. Do et al 2020). One concrete example of the usage of this asset is the development of the layer management (LM) concept with Air Liquide Engineering & Construction and with Clariant as described in figure 3b. Within the LM concept, a different optimization strategy can be tailored for CAPEX (less catalyst volume, compact distillation) and OPEX (longer catalyst lifetime, less recycle ratio, heat management, less by-product formation) saving. This concept can also open for customized and disruptive design focusing on flexibility of feedstock, high H_2 efficiency and compactness of the whole methanol loop. Another example is an extensive test campaign with CO2 rich feed stocks to develop a brand new kinetic model for CO2 based methanol and a refined deactivation model for the catalyst under these conditions (see Figure 4).

2.2 Multistage Methanol Pilot Unit

Figure 5: Multistage MeOH Pilot plant

The new multistage methanol reactor (figure 5) has been built and successfully started in 2020 under the framework of an EU project funded by RFCS namely i3-upgrade (grant agreement No 800659) with the objective to convert off-gases from steel plant into methane and methanol. In this case it is very important to respond quickly to changes in load and in gas compositions, to be able to deal with high amounts of inert like methane and nitrogen, as well as to use very effectively the hydrogen available.

The reactor has several tubes with several stages which can be flexibly adjusted for different concept testing. The system is well equipped with thermocouple, hot gas sampling, and online-offline analysis. The throughput is up to 35 m³N/h for feed gas and can produce methanol with capacity up to 20kg/h. The multistage reactor concept is promising to offer an optimal solution in the conversion of gases with high inert content since it allows very effective conversion rates at reduced recycle rates (down to <1), or even without recycling, and reduced costs (CAPEX & OPEX) for low reactive gases like CO2. As a first step toward validation, an experimental campaign has run in the multistage pilot for more than 1800 hours time on stream (TOS) to confirm for example byproduct content, temperature profiles and plant dynamics and to validate the process model setting as a basis for next step process design. The total byproduct value of the reactor under different pressure conditions can be seen in Table 1. This is crucial to ensure proper distillation column design and requested methanol grade.

Table 1: maximum temperature, hydrogen conversion and byproducts formation

Case	P/ bar	MUG/ Nm³/h	Tmax/°C	X_H2/%	By-products/wt-ppm
1	50	6.7	251	63.6	3808
2	70	6.7	261	72.7	3928
3	90	6.7	273	78.1	3909
4	110	6.7	279	81.9	3548

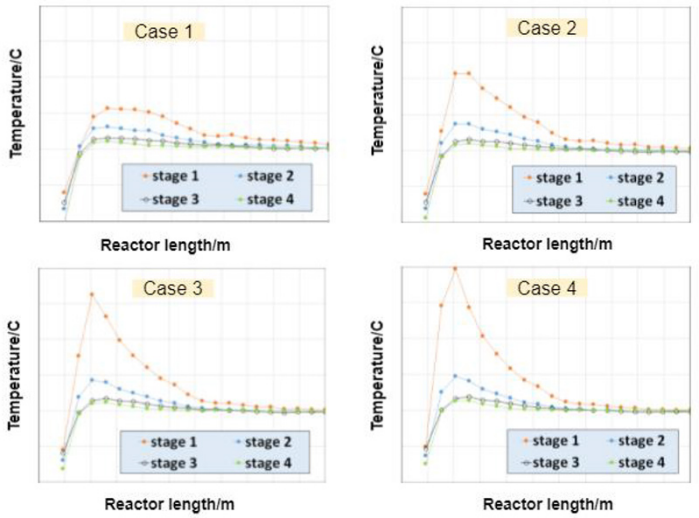


Figure 6: Temperature profile of 4 stages in 4 different cases studies as listed in table 1

The temperature profile decreases along with the stages and reaches a flat profile approaching the reactor end due to the reduction of exothermic reaction as the components approach equilibrium (figure 6) and also the cooling effect of the water jacket.

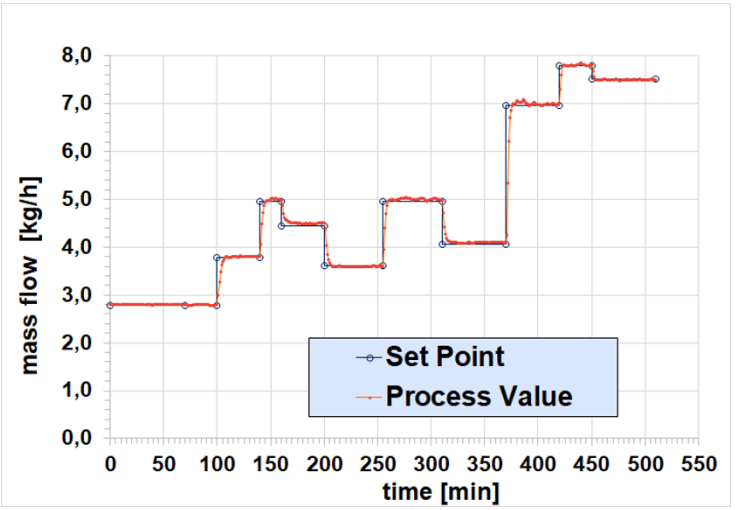


Figure 7: The load response of plant to set point

As renewable energy is very intermittent by nature. Additionally, offgas outlets in terms of composition and conditions can also vary depending on upstream feedstock and operation mode. Therefore, one of the challenges of integration from upstream fluctuation behavior to electrolysis and to hydrogen usage downstream like methanol is to know the bottle-neck(s) that determines the whole plant dynamics. This is to ensure the whole value chain operability and supply chain planning and optimization. Aiming to understand the dynamics, the campaign was done in the pilot by varying load in different steps and time

intervals in a once-through setup. Results in figure 7 show a very fast response (less than 10 minutes) of the process value to the set point values. This is a huge advantage in comparison to the classical setup in which a much longer time is needed to respond to changes in load and in gas compositions. In the first step this indicates very good synergy in the coupling solution of electrolysis and this new multistage methanol reactor system. Moreover, it paves the way for innovative advanced process control schemes to respond properly to the availability of hydrogen and the CO and CO2 present in different streams for the methanol synthesis.

3. Conclusions and Outlook

The reduction of carbon footprint via the valorization of CO2 rich gas, off gas, and flue gases to fuel or further usable chemicals are of increasing interest for many industries. Combined with green hydrogen from electrolysis using renewable energy, this C-1 gas source can be used for producing clean methanol as a direct fuel source or building block for producing high value chemicals. To enable the synergy of the whole value chain from upstream (energy) to downstream (production, usage, storage), the innovation philosophy has not only focused on finding an optimal solution of one process unit but the whole chain. Having a stepwise approach, identification of determining process units is important for operability and process design. In this contribution, the strong focus on the methanol process can be realized. In addition to the conventional methanol pilot, the recent developments on new process concepts and the multistage reactor development has provided some detailed information based on the experimental campaign of about 2000 hours of TOS. The by-product formation, the temperature profile and peak temperature are in the common range and the fast response of the reactor can be achieved. The dynamic behaviour of the new pilot shows a promising synergy with electrolysis in valuing the renewable energy fluctuation. The flexibility in stage operation provides an opportunity in the development of tailor-made solutions for diverse carbon sources, especially focusing on a circular economy and on the valorisation of off-gases and of CO2 rich streams from industries and the distribution of renewable energy.

References

International Energy Agency, Net Zero by 2050 A Roadmap for the Global Energy Sector, July 2021.
Air Liquide Climate Objective (accessed on 10.2021), 2019.
N.T.Q. Do *et al.*, Layer Management for Methanol Process, Oil & Gas Magazine, June 2020.
T. Oelmann *et al*., A new reactor concept for conversion of CO2 to Methanol, DGMK, Future of Chemicals and Fuels - Feedstocks and Process Technologies, October 2020.
T. Svitnik, *et al,* Data-driven approach for predictive modeling of by-product formation in methanol synthesis process, 30th Computer Aided Chemical Engineering, Volume 48, p 505-510, September 2020.
i3upgrade (https://www.i3upgrade.eu/), this project has received funding from the Research Fund for Coal and Steel, grant agreement No 800659; (accessed on 10.2021).

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50130-5

Optimization and Heat Exchanger Network Design of Diethyl Carbonate Two-step Synthesis Process from CO_2 and Propylene Oxide

Tsai-Wei Wu, I-Lung Chien*

Dept Chemical Engineering, National Taiwan University,
No.1, Sec. 4, Roosevelt Road, Da'an Dist.,10617 Taipei City, Taiwan;
ilungchien@ntu.edu.tw

Abstract

CO_2 has been deemed crucial as an alternative carbon source recently due to its low price and the urge to utilize it as a way to reduce its amount in the atmosphere. Diethyl carbonate (DEC) is a potential candidate for CO_2 conversion because it can serve as good fuel additive or solvent for lithium batteries. However, CO_2 conversion reaction often suffers from low conversion rate because of its very stable chemical activity. Not only catalyst development is essential, but the research from a process systems engineering (PSE) perspective is also necessary when considering the practical implementation in the future. In the previous work from our group (Chen & Chien, 2018), three processes of DEC synthesis from CO_2 and propylene oxide (PO) were proposed and compared. The results showed that the two-step process performed relatively better in CO_2 emission amount. The two step process includes PO reacting with CO_2 to produce propylene carbonate (PC) and PC reacting with ethanol to produce DEC. Nevertheless, after environmental evaluation throughout the two-step process, the positive net CO_2 emission amount indicated that it still couldn't be a strategy for CO_2 emission mitigation. In this study, the root cause of the high energy consumption of the process was reviewed, and it was found that the large excess ratio of the second step reaction led to massive energy consumption in the separation stage. Consequently, process optimization has been executed by sequential iterative procedure. Heat exchanger network according to pinch analysis along with thermally coupled configuration has been developed for the two-step DEC synthesis process, and the results showed that the new design can save at least 75% CO_2 emission amount for the process. The purpose of this work is to investigate the potentiality of the system to be a CO_2 emission reduction route.

Keywords: CO_2 Utilization; Diethyl Carbonate; Heat Exchanger Network; Optimization.

1. Introduction

Extreme climate all over the world due to the serious Greenhouse effect has caused many serious disasters nowadays. The reduction and utilization of CO_2 have become urgent research topics, and CO_2 valorisation can be a solution to consuming CO_2 profitably. Among the candidate products of CO_2 valorisation, DEC is a promising one for its wide applicability as fuel additive or solvent. In addition, no azeotropes exist in the system of DEC production process, which will be an advantage for the purification section.

Chen and Chien (2018) had investigated three routes for producing DEC by using CO_2 and PO as the feedstocks. Route 1 consists of three steps, which start from the production of PC, and the transesterification of which with methanol produces dimethyl carbonate

(DMC), leading to the final product DEC by reacting with ethanol. Route 2 has the same first step of producing PC, with the second step reacting PC with ethanol to produce DEC directly. Route 3 is a one-pot reaction of PO, CO_2 and ethanol, generating DEC with propylene glycol monoethyl ether (PEE) as by-product. After economic and environmental evaluations, route 2 was decided to be the process which exhibited larger potential for carbon dioxide utilization.

As a result, optimization, process intensification and heat exchanger network design were implemented in this study, in order to disclose the capability of reducing carbon emission through this application. Net CO_2 emission for route 2 in Chen and Chien (2018) was estimated to be 0.7 kg CO_2/kg DEC, and it is our target to reduce the value to less than 0 which can truly imply a reduction in carbon emission.

2. Process description and optimization

The process flowsheet of route 2 proposed by Chen and Chien (2018) is shown in Fig. 1, which demonstrates two reactors and four columns. CO_2 and PO are fed to a PFR first for the production of PC, which is purified in the following stripper. Subsequently, ethanol is fed into the system with PC in a CSTR for generating the desired product DEC. Ethanol excess ratio was set to be 8 to achieve a higher PC conversion, and this leads to high energy consumption to purify the unreacted ethanol for the following column. The last two columns are used for purifying DEC and the by-product PG.

Optimization by simple sequential iterative procedure has been carried out for choosing proper design parameters and operating conditions. Stripper pressure, stripper stage number, pressures, stage numbers and feed locations for the three purification columns were determined by the optimization procedure using total annual cost (TAC) as objective function. In this work, the overall TAC is calculated by Eq. (1), where TOC stands for total operating cost and TCC stands for total capital cost. The correlations for TAC calculation were from Luyben (2012).

$$\text{TAC} = \text{TOC} + \frac{\text{TCC}}{\text{payback period}} \tag{1}$$

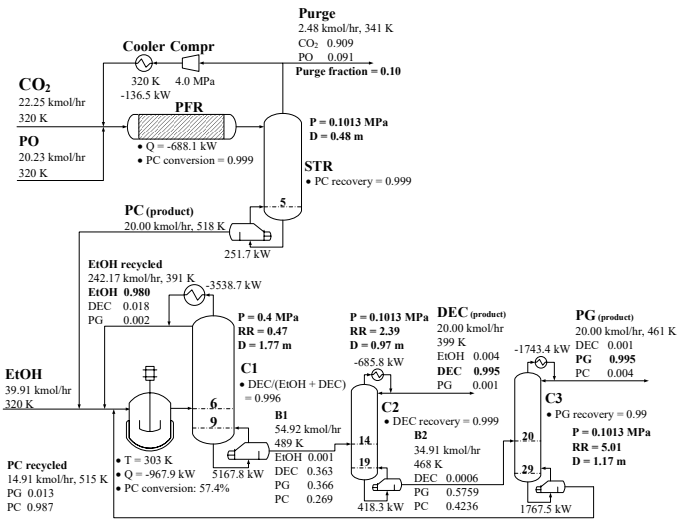


Fig. 1. DEC two-step production process flowsheet

After process optimization, in comparison to the base case in Fig. 1, pressures of stripper and C1 were adjusted lower to save reboiler duty and heat exchanger area. Number of stages in C3 has been increased considerably since PC and PG have close VLE curves which indicate more required stages for separation. Nevertheless, C1 remains to be the unit with the largest energy consumption.

3. CO_2 emission calculation

CO_2 emission amount throughout the process is calculated by the addition of CO_2 emission due to energy usage in each unit. In this process, steam of different grades for reboilers and electricity used in compressor were considered. The corresponding CO_2 emission amount of different utilities follows the calculation method proposed by Gadalla et al. (2005) and is listed in Table 1. The main correlation for calculating CO_2 emission amount is shown as Eq. (2), where Q_{Fuel} is the amount of fuel used; NHV is the net heating value of the fuel used in the boiler (LNG 51600 kJ/kg); C% is the carbon percentage of the fuel (75.4 for LNG); α is the molar weight ratio of CO_2/C (3.67).

$$[CO_2]_{Emiss} = (\frac{Q_{Fuel}}{NHV})(\frac{C\%}{100})\alpha \quad (2)$$

Table 1. Equivalent CO_2 emission amount

Utility	Equivalent CO_2 emission (kg/GJ)
Low-pressure steam, 0.5 MPa, 160 ºC	72.86
Medium-pressure steam, 1.0 MPa, 184 ºC	76.60
High-pressure steam, 4.2 MPa, 254 ºC	91.14
Electricity	96.95
Cooling water	-

4. Process intensification and heat integration

4.1. Thermally-coupled process

Thermally-coupled columns are commonly used for columns with obvious remixing effect (Li et al., 2015; Wang et al., 2014), which leads to unnecessary energy consumption and can be observed on the figure of composition profile in columns. After reviewing the composition profile in the purification columns C1 – C3, it was found that C1 and C2 both had obvious remixing effect in DEC and PG components respectively. C2 and C3 was chosen to implement the thermally-coupled streams because high-pressure steam was used in these two columns, while medium-pressure steam was used in C1.

It can be observed that remixing effect has been almost eliminated after implementing thermally coupling to C2 and C3 as shown in Fig. 2. PG composition fluctuates less and the overall reboiler duty has been saved for 3.26%. Noted that optimization when implementing thermally coupling is necessary since vapor side draw flow rate and liquid side stream feed location are influential to the energy saving result. Optimized design parameters for the thermally-coupled columns can be seen in Fig. 3.

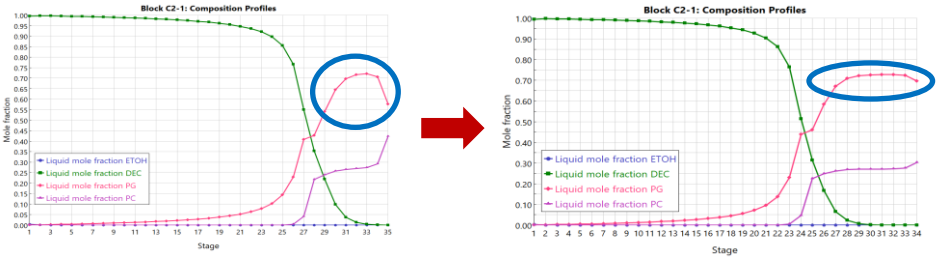


Fig. 2. Elimination of remixing effect in C2 after thermally coupling

4.2. Heat integration

By observing the energy consumption among each unit in Fig. 1, it is obvious that the most energy-consuming unit is C1 for ethanol purification. The ethanol purified specification was set to be 98 mol% for the sake of saving energy; however, the large excess ratio of ethanol to PC led to the significant amount of steam usage. Fortunately, this is an opportunity for using the top vapor in C1 as a heat source for the system. Final process flowsheet with thermally-coupled columns and heat integration is shown in Fig. 4. The top vapor stream of C1 was used to heat up the product stream of DEC reactor, and the top vapor streams of C2 and C3 joined the line to preheat the inlet stream fed to C1. The sequence of using different heat sources was determined by their temperature to guarantee the enough temperature difference for heat exchange. An auxiliary condenser is needed for C1, for the temperature of C1 top vapor stream is not high enough for complete heat exchange with reactor product stream. After optimization and heat integration, 60% of C1 reboiler duty can be reduced compared to the base case. A feed-effluent heat exchanger is further installed before the thermally-coupled columns C2 and C3.

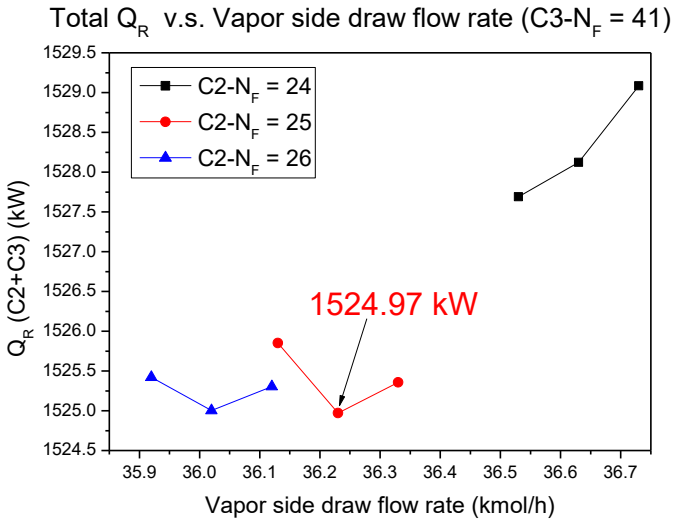


Fig. 3. Optimized design parameters for the thermally-coupled columns

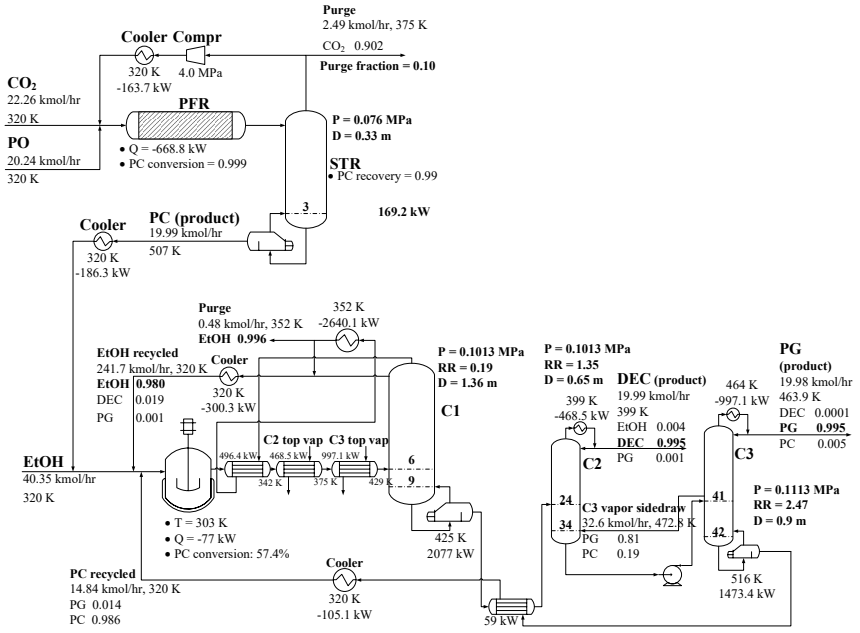


Fig. 4. Energy-saving process flowsheet

CO_2 emission amount of the base case (Fig. 1), optimized process and the final energy-saving process with heat integration (Fig. 4) is listed in Table 2. Optimization has successfully reduced the energy consumption in the both steps, and half of the CO_2 emission amount can be saved in this stage. After the implementation of thermal coupling and heat integration, significant CO_2 emission amount can be saved in the PC to DEC step along with the saving in energy consumption. A nearly zero value can be derived for the CO_2 net emission of the best process. Large excess ratio of ethanol and PC resulted in the failure of reducing CO_2 net emission to a negative value.

Table 2. CO_2 emission amount of the three processes

	Base case	**Optimized**	**Optimized & Heat-integrated**
CO_2 emission (kt/y)	Step A: 0.97	0.82	0.82
	Step B: 19.30	12.70	7.95
CO_2 consumption (kt/y)	7.04	7.04	7.04
Net CO_2 emission (kt/y)	13.22	6.57	1.82
DEC production (kt/y)	18.89	18.84	18.84
Net CO_2 emission (kg CO_2/kg DEC)	0.70	0.35	0.097
CO_2 (emission/consumption)	1.88	0.93	0.26

5. Conclusion

A CO_2 utilization process for producing DEC has been optimized and intensified by the implementation of thermally-coupled columns and heat integration in this study. In the base case, PC is generated by the reaction of PO and CO_2, subsequently, excess amount of ethanol is fed to a CSTR with PC to produce the desired product DEC. The largest disadvantage in the system is the large excess ratio of ethanol to PC, which leads to a unavoidable large reboiler duty in C1, accounting for up to 67% of CO_2 emission among the process. CO_2 emission amount was calculated to be 0.7 kg CO_2/kg DEC.

Sequential iterative procedure was used for the optimization of the base case. Results showed that CO_2 emission amount can be reduced to 0.35 kg CO_2/kg DEC by adjusting the operating pressure and design parameters in the system.

Thermal coupling was used as a strategy for process intensification. The remixing effect in C2 was almost eliminated and the overall heat duty can be saved for C2 and C3. Heat integration has also been investigated for the system, and four heat exchangers were installed to recover the heat from top distillate streams and a hot stream. The energy-saving process saves 86% of CO_2 net emission amount, leading to a nearly carbon neutral DEC production process.

With new development in catalyst in the future, it is expected that the conversion of PC to DEC can be enhanced, and the excess ratio of ethanol to PC can be reduced to a reasonable value. Additionally, green energy can be used for supplying the utilities used in this system, such as combined heat and power system or renewable energy. It is firm that researchers are going to explore in the field of green production, and there is high possibility that DEC production can be an approach dealing with CO_2 reduction and utilization in the near future.

References

Chen, M.K., & Chien, I. L., 2018, Potentials for CO_2 Utilization: Diethyl Carbonate Synthesis from Propylene Oxide, Computer Aided Chemical Engineering, 44133-138.

Gadalla, M. A., Olujic, Z., Jansens, P. J., Jobson, M., & Smith, R., 2005, Reducing CO_2 Emissions and Energy Consumption of Heat-Integrated Distillation Systems, Environmental Science & Technology, 39, 17, 6860-6870.

Li, L., Sun, L., Wang, J., Zhai, J., Liu, Y., Zhong, W., & Tian, Y., 2015, Design and Control of Different Pressure Thermally Coupled Reactive Distillation for Methyl Acetate Hydrolysis, Industrial & Engineering Chemistry Research, 54, 49, 12342-12353.

Luyben, W. L., 2012, Principles and Case Studies of Simultaneous Design: Wiley.

Wang, S.J., Cheng, S.H., Chiu, P.H., & Huang, K., 2014, Design and Control of a Thermally Coupled Reactive Distillation Process Synthesizing Diethyl Carbonate, Industrial & Engineering Chemistry Research, 53, 14, 5982-5995.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50131-7

Characterization of Industrial Flaring under Uncertainty for the Design of Optimum Flare Recovery and Utilization Systems

Monzure-Khoda Kazi[a], Fadwa Eljack[a*], Saad Ali Al-Sobhi[a], Vasiliki Kazantzi[b], Nikolaos Kazantzis[c]

[a]*Department of Chemical Engineering, College of Engineering, Qatar University, Doha P.O. Box-2713, Qatar*
[b]*Department of Business Administration, University of Thessaly, Larissa 41500, Greece*
[c] *Department of Chemical Engineering and Center for Resource Recovery and Recycling, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609, USA*
**Corresponding Author's E-mail: Fadwa.Eljack@qu.edu.qa*

Abstract

One of the main challenges in industrial applications is to optimally manage flare gases that are inevitably generated both in routine and non-routine process operations but can yet constitute valuable energy resources for process systems. A main challenge is to explore the best possible strategies for exploiting these valuable hydrocarbon streams and propose process design alternatives and operational solutions that achieve maximum recovery and use of flare gases at minimum total cost and considering the uncertainty variations associated with flaring incidents. This requires an understanding of the characteristics of flare streams that affect their recovery and reutilization potential as well as an examination of their impact on process system performance while recognizing that the inherently uncertain nature of flaring calls upon a probabilistic approach. In our study, we examine the impact of using a comprehensive probabilistic analysis framework for process flare streams' characterization on the design of an optimal recovery and utilization system. In particular, the work aims to explore the impact of uncertainty for key parameters on the design solutions, such as rate of flare occurrences that were assumed constant in other research works (Kazi et al., 2018). Suitable parametrized Monte Carlo (MC) simulations are employed for more accurate flare profile representations. A comparative study is conducted between the base case optimal design and values at risk solutions for cases where flaring variation increases may significantly affect the design features and economic performance of the process system. The proposed framework could inform decision makers' assessments of the impact of random variations in flaring profiles on process performance profile.

Keywords: Flare Management, Flaring Uncertainty, Monte Carlo, Flare Characterization.

1. Introduction

A dominant strategy in the circular economy era is material and energy recovery, recycling, and reuse, offering multiple economic, environmental, and social benefits when realized in an efficient manner. It has been shown that there are several flare management opportunities and enabling technologies for the recovery and utilization of flare streams from process industries that can significantly contribute to reducing the carbon footprint, increasing energy efficiency and enhancing economic process performance, thus supporting circular economy actions (Kamrava et al., 2015; Kazi et al., 2016; Kazi et al., 2015). In this respect, the main issue is to be able to explore the best possible strategies for scavenging these waste streams in a particular process system and

suggest the optimal process design and operation alternatives for attaining their maximum recovery and utilization at the minimum possible annualized total cost. Moreover, a holistic strategy needs to be developed to consider the maximum potential plant-wide benefits under uncertainty variations associated with flaring.

The ability to adequately characterize streams set for flaring in terms of their energy supply potential is of paramount importance in order to recognize flaring perspectives, as potential waste energy carrier. From a process systems engineering point of view, flare stream characterization will facilitate the establishment of optimum designs for the flare recovery and utilization system. The challenge is that the flare occurrences, as well as flowrates and compositions of hydrocarbon streams flared from various process sites, constitute highly uncertain inputs to the design model of a recovery and utilization system in a controlled and safer manner (Kazi et al., 2019a, 2019b). Hence, their occurrence rates, and their energy content vary significantly over time and with respect to the process site(s) from which they are generated (Kazi et al., 2018; Kazi et al., 2015). Although challenging, the process energy utilization opportunities are broadened as available streams from multiple process sites can be mixed and segregated to obtain the optimum possible waste energy load. While flare streams' characterization is obviously source dependent, attaining their combined probabilistic profiles can help in directly identifying 'the big picture' of the design alternatives and providing insights into a more systematic design approach. Hence the framework suggested in this work encompasses a methodology to systematically examine the improvement of the integrated process performance regarding economics, energy consumption and most importantly the environmental impact for routine and non-routine flare gases during abnormal situations. In addition, the extent of variations in the uncertain inputs has been examined within the scope of their impact on the overall process performance profile through an appropriately developed probabilistic sensitivity analysis.

2. Problem statement

Given is a productive system with known design and operational data under certain utility requirements. Available is also a historical database of flaring events (flare causes, sources, frequencies, compositions, and duration) and specific regulatory limits on GHGs. A COGEN unit with a specific boiler type and known capital and operating expenditures is considered as an efficient flare mitigation tool. A systematic approach to maximize the use of available flare streams by optimally sizing the COGEN unit is developed with the aim to ultimately gain environmental and economic benefits, as well as to achieve efficient energy recovery and sustainability enhancement. The need to explore the impact of incorporating a flare mitigation system on process performance while considering the uncertainties of the flaring incidents is reflected in particular, whereas an examination on how the extent of variations in the uncertain inputs (flare frequency, amount etc.) impact performance outcomes (energy-related, environmental, economic) is performed.

3. Flare streams characterization and optimization model

Probability distributions of flare properties comprising input variables can capture the uncertainty and variability explicitly when incorporated into the optimization model taking into account all sources of uncertainty at the same time. Especially non-routine flaring events are constantly faced with uncertainty and variability, and thus can not be adequately represented in an optimization model. Monte Carlo (MC) techniques are utilized in this work to probabilistically characterize and interpret such incidents that can subsequently be employed as input variables in process optimization models. MC simulations provide a means to risk analysis by building models of possible outcomes after sampling a range of values for any uncertain variable. Producing probability distributions of all possible outcomes, MC techniques furnish the decision maker with a more realistic way of describing uncertainty in input variables that can propagate through

the optimization modeling framework. They can also offer multiple insights from the whole process case scope and sensitivity analysis results.

In this study a previously developed multi-period optimization formulation is used, with the objective of minimizing total cost of the flare recovery and utilization system. The objective function is mathematically expressed as follows (Kazi et al., 2018):

Objective function:

$$\sum_{Min} \mathrm{TAC} = \underbrace{C_{op} \times H_{\gamma}}_{\text{Annual Operating Cost}} + \underbrace{k_f \times (C_{\mathrm{Boiler}} + C_{\mathrm{Turbine}})}_{\text{Annual Fixed Cost}} - \underbrace{P_e \times H_{\gamma}}_{\text{Annual Income}} - \underbrace{E_o \times C_{\mathrm{tax}}}_{\text{Carbon Tax Savings}} \quad (1)$$

$$\sum_{Min} \mathrm{TAC} = \underbrace{C_{op} \times (\mathrm{AOT} - \sum_{i=1}^{N} \boldsymbol{\alpha_i t_i})}_{\text{Annual Operating Cost}} + \underbrace{k_f \times (C_{\mathrm{Boiler}} + C_{\mathrm{Turbine}})}_{\text{Annual Fixed Costs}} - \underbrace{P_e \times (\mathrm{AOT} - \sum_{i=1}^{N} \boldsymbol{\alpha_i t_i})}_{\text{Annual Income}} - \underbrace{\left(\sum_{i=1}^{N} e_{ji}^{U} - e_{ji} t_i\right) \times \mathbf{C_{tax}}}_{\text{Carbon Tax Savings}} \quad (2)$$

i = Flaring events (e.g., PRC trip, Acetylene reactor trip, CGC trip etc.)

k = Flaring locations (e.g., flare A, flare B, ... flare G)

Subjected to

$$\sum_{i=1}^{N} \alpha_i = 1 \quad (3)$$

$$\text{Where, } \alpha_i = \frac{n_i}{n_{tot}} = \frac{\text{Expected number of event } i \text{ per year}}{\text{Expected number of total events per year}} \quad (4)$$

$$t_i = \frac{m_{tot,i}}{m_i} = \frac{\text{Annual total mass flowrate of event } i}{\text{Mass flowrate of event } i \text{ at correspondin } k} \quad (5)$$

Eq. 1 represents the objective function of the optimization model expressed in terms of the total annualized cost (TAC) considering the operating costs related to fuels, boiler and turbine of the COGEN system, the annualized capital cost of the COGEN, the annualized income from any excess generated power, and the environmental cost in terms of CO2 tax savings for total emissions. Similarly, Eq. 2 refers to the TAC with the same four economic terms, but while considering the uncertain variable of operating time based on the flare duration fluctuations with the inclusion of α_i and t_i obtained from MC simulation results (see Eqs. 3-5) as explained in Kazi et al. (2018) (Kazi et al., 2018). The ultimate objective is to design an optimum COGEN unit (e.g., the size of the boiler and turbine) for flare minimization under uncertainty.

4. Results

4.1. Probabilistic characterization

To organize and visualize the data sets available for the ethylene process system under consideration, a MC simulation approach was adopted aiming to capitalize on the occurrence values of the random variables and characterize the flare streams with respect to the extent of their energy supply and utility potential. Therefore, a detailed statistical characterization of flare stream occurrences was performed first by using standard MC simulation techniques and probability distribution outcomes for flare flowrates that were generated for each uncertain incident as shown in Fig.1 for the off-spec production.

Flare distribution profiles can provide values for the whole range of flare flowrates associated with their probabilities, including maximum and minimum flowrate values or ranges of flow rates for different scenarios. For example, Fig. 1 demonstrates that the probability of the flare flow rate values for the off-spec production to reach up to 1.2 $\times 1^6$ kg/yr is around 55%. Thus, if the probability/likelihood of each flaring incident per year is known, the expected flowrate value range of that incident can be estimated from its distribution profile. Next, these simulation results generated can be simultaneously fed to and propagated through the optimization model as random input variables, to generate a comprehensive set of solutions under different uncertainty scenarios. The set of solutions are obtained in the form of Pareto fronts representing heat and power combination points that dictate optimal designs of the COGEN system in the presence of flare uncertainties, as discussed in Kazi et al. (2018). Based on the results, the decision-maker can appropriately select, trade-off, and incorporate his design and operational preferences concerning the desired implementation goals and hierarchically implement and manage his preferred design solutions in a multi-objective decision-making approach.

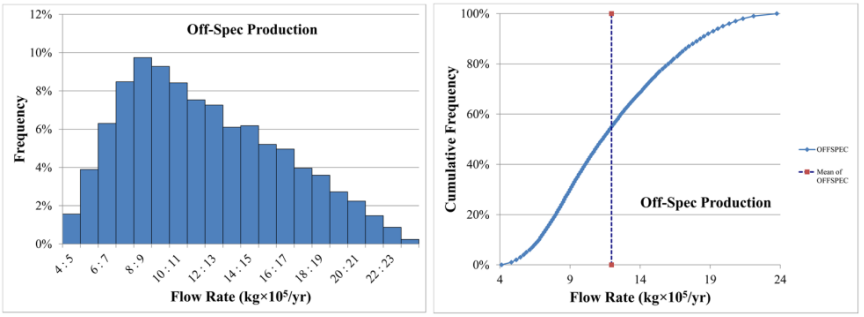


Figure 1. Frequency and cumulative distribution profiles of process upset during off-spec production

4.2. Sensitivity analysis

Next, sensitivity analysis was carried out to explore how the extent of variations in the uncertain inputs (frequencies, emissions, cost) impact performance outcomes. Initially the probability distribution for the simulated total flaring frequencies is depicted in Fig.2. Although it was found that the standard deviation of distribution is quite small in this case (0.0067), variations in the total amount of flowrates can be amplified, and hence greatly impact process performance outcomes such as, CO2 tax savings, and sizing the COGEN unit, as shown in Kazi et al. (2018).

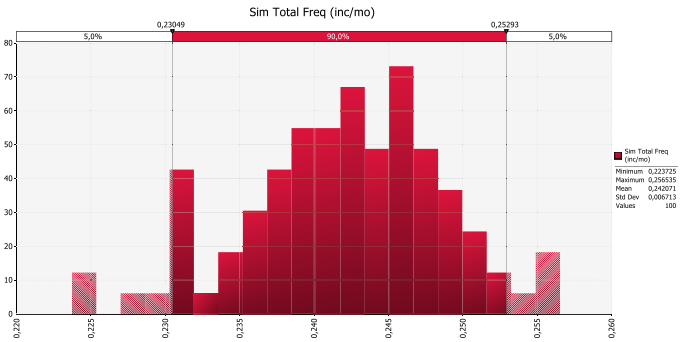


Figure 2. Simulated total frequencies (number of incidents/month)

The effect of the incident type on the simulated total Mean Time Between Incidents (MTBI) is next illustrated by employing the sensitivity analysis tool of the @Risk software, the results shown in Fig. 3. In particular, a dominant effect of incidents happening due to the off-spec production on total expected flaring events (due to the most frequent number of these incidents) was evident. The variability of the total amount of CO2 emissions generated in the ethylene production plant was also examined with respect to both incident type and time (year) of occurrence. In Fig. 4a it is shown that the highest variability in CO2 emissions was observed during the years 2005-2007, whereas the largest value is shown in 2012. All statistical reports are obtained to enable a full characterization of all the uncertain input variables for which probability distributions and variability analyses are produced.

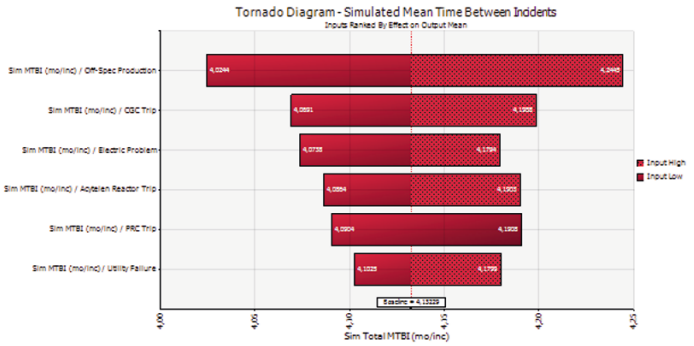
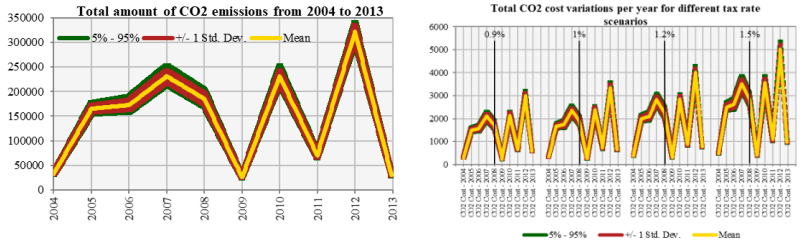


Figure 3. Impact of incident type on variability of total MTBI (mo/inc)

The analysis included examination of the impact of uncertainty in tax rates on expected total CO2 cost. Multiple scenarios with 0.9%, 1.2% and 1.5% annual tax rates were simultaneously examined, assuming a 0.5% annual tax rate increase. As seen in Fig. 4b, annual variability in expected CO2 cost exhibits the same pattern through the different scenarios, as expected, with a significantly accentuated effect for the 1.5% tax rate case. Uncertainty in economic and market conditions can thus be captured and propagated through the optimization model (Kazi et al., 2018). These uncertainty characteristics may influence and cause deviations from process objectives and hence need to be explicitly included in the model and explored for making more well-informed decisions at both the strategic (process design) and operational levels.



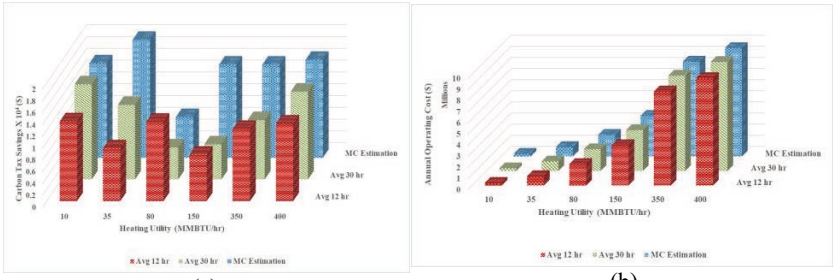
(a) (b)

Figure 4. (a) Variations of the total amount of CO2 Emissions through the decade (2004-2013) and (b) CO2 cost variations per year for different tax rate scenarios

4.3. Effects of uncertainties on process economics

Sensitivity analysis on the techno-economic and environmental factors was carried out in the light of a flaring uncertainty and it was observed that some of the components evaluated (i.e., CO2 tax savings) can significantly be influenced by mild changes in the

flaring profiles (see Fig. 5a). The flare mitigation options may be over/under designed when evaluating the typical values for flare durations. Fig. 5b shows that, if uncertainties relating to flare incidents are incorporated, OPEX values may change for different sizes of the COGEN unit.



(a) (b)

Figure 5. Effects of uncertainties on economic factors: (a) annual CO2 tax savings and (b) annual operating cost under different flare duration scenarios from sources

5. Conclusions

Flaring characterization is essential to effectively design a COGEN system and improve integrated process performance at an economic, operational and environmental level. The extent of variations in the uncertain inputs and its impact was the focus of this work. In particular, the proposed framework enables: (1) integration of key data sets facilitating further elaboration (insightfully summarizing/visualizing input data of uncertain nature, describing basic features, statistically exploring dominant characteristics and trends that may have been overlooked), (2) probabilistic characterization of available flare streams as energy sources by insightfully generating real flaring profiles using standard MC techniques, (3) direct incorporation of the risk associated with the uncertain nature of the flaring events in the multi period optimization model, and (4) examination of the economic, environmental and energy-related trade-offs derived by optimizing the energy recovery and utilization system while providing valuable information to multiple end-users. Although the usual limitations of the existence of only scarce and case-dependant data hinders a more complete exploration of the system's behaviour, the conceptual sensitivity analysis framework for visualizing and describing the degree of impact of inherently uncertain parameters' variability on performance objectives is highlighted, aiming at generating real data-driven insights for enhancing the empirical understanding of efficient energy recovery and utilization systems.

References

Kamrava, S., Gabriel, K. J., El-Halwagi, M. M., & Eljack, F. T. (2015). Managing abnormal operation through process integration and cogeneration systems. Clean Technologies and Environmental Policy, 17, 119-128.

Kazi, M.-K., Eljack, F., Al-Sobhi, S. A., Kazantzis, N., & Kazantzi, V. (2019a). Application of i-SDT for safer flare management operation. Process Safety and Environmental Protection, 132, 249-264.

Kazi, M.-K., Eljack, F., Al-Sobhi, S. A., Kazantzis, N., & Kazantzi, V. (2019b). Process Dynamic Analysis and Control Strategy for COGEN option Used for Flare Utilization. In A. A. Kiss, E. Zondervan, R. Lakerveld & L. Özkan (Eds.), Computer Aided Chemical Engineering (Vol. 46, pp. 1255-1260): Elsevier.

Kazi, M.-K., Eljack, F., Amanullah, M., AlNouss, A., & Kazantzi, V. (2018). A process design approach to manage the uncertainty of industrial flaring during abnormal operations. Computers & Chemical Engineering, 117, 191-208.

Kazi, M.-K., Eljack, F., Elsayed, N. A., & El-Halwagi, M. M. (2016). Integration of Energy and Wastewater Treatment Alternatives with Process Facilities To Manage Industrial Flares during Normal and Abnormal Operations: Multiobjective Extendible Optimization Framework. Industrial & Engineering Chemistry Research.

Kazi, M.-K., Mohammed, F., AlNouss, A. M. N., & Eljack, F. (2015). Multi-objective optimization methodology to size cogeneration systems for managing flares from uncertain sources during abnormal process operations. Computers & Chemical Engineering, 76, 76-86.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50132-9

DEVELOPMENT OF MICRO SCALE PRC USING LOW GRADE GEOTHERMAL THERMAL ENERGY

Ryosuke AKIMOTO [a], Yasuhiko SUZUKI[b], Yuki OGASAWARA[b], Masaru NAKAIWA[a] and Keigo MATSUDA[a*]

[a] Department of Chemistry and Chemical Engineering, Graduate School of Science and Engineering, Yamagata University, 4-3-16, Jonan, Yonezawa, Yamagata 992-8510, Japan

[b]Faculty of Chemistry and Chemical Engineering, Yamagata University, 4-3-16, Jonan, Yonezawa, Yamagata 992-8510, Japan

[*]Corresponding Author's E-mail: matsuda@yz.yamagata-u.ac.jp

ABSTRACT: In order to enhance the versatility, we proposed a design method in which a micro ORC of 10 kW class. The micro ORC is small scale power generation system for using renewable thermal energy such as hot springs. However, such thermal energy is of very low grade and involves temperature variations, so it is necessary to develop a versatile small power generation cycle. In this study, the design conditions of heat exchanger, expander, and pump were fixed for the micro ORC, and the power generation performance was investigated using six different working fluids for multiple heat source conditions. The heat source flow rate set to be 3,000 kg/h and the heat source temperature was 393 K. The heat sink flow rate set to be 18,000 kg/h and the temperature was 293 K. The pump discharge pressure was 1,500 kPa. As a result, Among the selected working fluids, butane showed the highest power generation performance of 12.9 kW. To predict the power generation performance of the micro ORC, we correlated the power generation of each working fluid with the thermophysical properties of the working fluid, but low correlation was obtained. Therefore, we introduced new parameter, Exergy Parameter Index (EPI), for predicting the power generation of a micro ORC. It was found to be a very reliable parameter for forecasting the power generation of a micro ORC.

Keywords: Power generation, Organic Rankine Cycle, Renewable energy

1. Introduction

Small ORCs are one-of-a-kind industrial products because the working fluid, evaporator, condenser, pump, and expander are designed to correspond to the heat source conditions. Therefore, it is not versatile and has a high cost. In order to design a small ORC, there are two ways of thinking about it: one is to create a one-of-a-kind device with the heat recovery rate as the objective function, and the other is to improve the power generation performance of the same device by changing the working fluid and pump discharge pressure. From the viewpoint of

process enhancement, concepts such as miniaturization are necessary in the design of ORCs. Recently, several experimental studies of compact ORCs have been reported (Akimoto *et al.* 2021, Li *et al.* 2021 and Uusitaro *et al.* 2020 and). Many researchers have discussed the performance of power generation by changing the specification of the equipment for a given heat source. To the best of our knowledge, the same equipment specification of a small ORC for various heat sources has not been studied at all.

In this study, we proposed a micro ORC as a waste heat recovery technology for the 10 kW class, assuming that the pump, expander, and heat exchanger can work properly even if the working fluid was different. By considering the micro ORC, we could expect a low-cost and high-efficiency small-scale exhaust heat recovery technology. In this paper, the micro ORC design and its power generation performance was discussed. The micro ORC was adopted by a model-based approach. We investigated how the power generation performance of micro ORC was affected by the exchange of working fluid and the control of the pump discharge pressure. We also proposed an index to represent the operation and design methods that match the heat source conditions.

2. Modeling

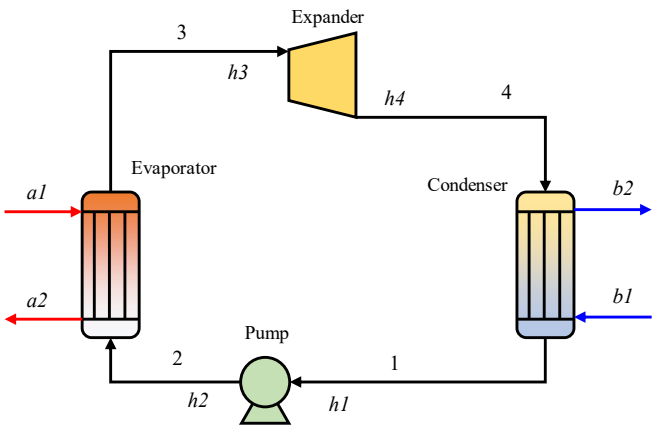


Fig. 1 Schematic diagram of the micro ORC

The schematic diagram of micro ORC with this study is shown in **Fig. 1**. Micro ORC consists of an evaporator, a condenser, an expander, a pump, and an injector of working fluid. The working fluid is supplied to the evaporator by the pump (1→2). In the evaporator, the working fluid is heated to the saturated vapor or superheated vapor by the heat source (2→3). The heated working fluid vapor is expanded through the expander to generate work (3→4). The expanded working fluid vapor is supplied to the condenser and condensed to the liquid by cooling water (4→1). Finally, the condensed the liquid is pumped back to the evaporator. **Table 1** shows the mass and

Table 1 Mass and Energy balance of each device

Device	Mass and energy balance equation
Evaporator	$\dot{m}_2 = \dot{m}_{a1}$ $\dot{m}_2(h_3 - h_2) = \dot{m}_{a1}(h_{a2} - h_{a1})$
Condenser	$\dot{m}_8 = \dot{m}_{b1}$ $\dot{m}_8(h_1 - h_8) = \dot{m}_{b1}(h_{b2} - h_{b1})$
Pump	$\dot{m}_1 = \dot{m}_2$ $W_p = \dot{m}_1(h_2 - h_1)$ $W_{p,net} = W_p / \eta_p$
Expander	$\dot{m}_4 = \dot{m}_5$ $W_{ex} = \dot{m}_4(h_5 - h_4)$ $W_{ex,net} = W_{ex}\,\eta_{ex}$

Table 2 Input parameter of design conditions for the micro ORC

Equipment	Design parameter	Value	Unit
Evaporators	Overall heat transfer coefficient	1,500	W/ (m^2·K)
	Heat transfer area	4.0	m^2
	Minimum temperature approach	5.0	K
Condenser	Overall heat transfer coefficient	1,000	W/ (m^2·K)
	Heat transfer area	15	m^2
	Minimum temperature approach	5.0	K
Pump	Efficiency	80	%
Expander	Isentropic efficiency	80	%
	Mechanical efficiency	80	%

Table 3 Input parameter of operation conditions for the micro ORC

Operation factor	Operation parameter	Value	Unit
Hot heat source	Flow rate	3,000	kg/h
	Temperature	393	K
	Pressure	200	kPa
Heat sink	Flow rate	18,000	kg/h
	Temperature	293	K
	Pressure	100	kPa
Pump	Discharge pressure	1,500	kPa

energy balance of each device. Operating conditions and design conditions of the micro ORC in this study are shown **Table 2** and **Table 3**. The working fluid was used propane, propylene, n-butane, isobutane, n-pentane, isopentane and R245fa.

3. Simulation results and discussion

Fig. 2 shows the power generation performance of each working fluid and relationship between

the key parameters, boiling point and latent heat, and the net power generation under the conditions of heat source temperature of 393 K, pump discharge pressure of 1,500 kPa, and heat source flow rate of 3,000 kg/h. Among the selected working fluids, butane showed the highest power generation performance of 12.9 kW, and the lowest power generation is propane, its value is 2.75 kW. In Fig. 5a, for R245fa, n-butane, and isobutane, the latent heat + sensible heat at 1,500 kPa is related to the amount of electricity generated, while for propane and propylene, it is not. In addition, as shown in Fig.5b, the boiling point is not related to the net power generation for all the working fluids. These results show that the net power generation of the micro ORC cannot be estimated only by the physical properties such as latent heat, sensible heat, and boiling point. Since the condensation pressure varies depending on the working fluid, the pressure difference between the inlet and outlet of the expander is different for each working fluid. It is known that the greater the pressure difference, the greater the amount of power generated by the expander. For components such as propane and propylene, which have a low boiling point of 233K or lower and a high condensation pressure of 1,000 kPa or higher, the pressure difference between the inlet and outlet of the expander is less than 600 kPa, resulting in less power generation. Therefore, it cannot be inferred from the physical property values alone. The power generation by the heat engine can be evaluated by the energy input to the process and the driving force. It should be able to be expressed in terms of enthalpy and driving force. The enthalpy is the sum of sensible and latent heat, and the driving force is the pressure difference before and after the expander. In this study, this relationship is proposed as an innovative index to estimate the power generation performance. We define this index as exergy parameter index (EPI), which estimates the power

Heat source temperature: 393 K, Heat source flow rate: 3,000 kg/h, Pump discharge pressure : 1,500 kPa

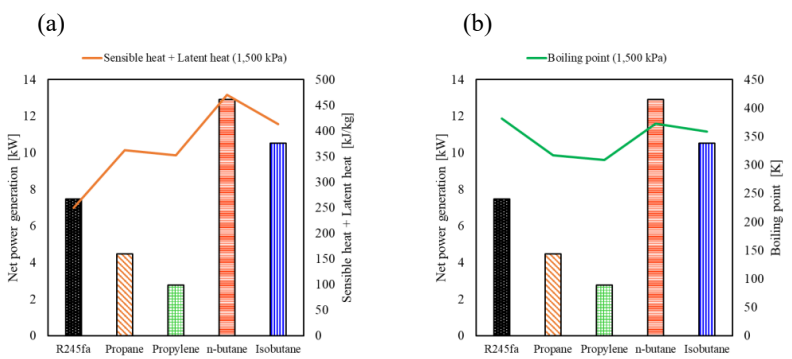


Fig. 2 The relationship between net power generation and: (a) Sensible + latent heat, (b) Boiling point (1,500 kPa)

generation performance of micro ORC using heat below 393 K. EPI is expressed by the following equation

$$EPI = \Delta P \cdot (h_{sensible} + h_{latent}) \quad (1)$$

where ΔP is pressure difference in the cycle, $h_{sensible}$ is sensible heat, and h_{latent} is latent heat. The pressure difference can be calculated as margin between the pump discharge pressure and the condensation pressure of the working fluid. $h_{sensible} + h_{latent}$ is heat required for the working fluid to become saturated vapor.

Fig. 3 show association between EPI and net power generation of micro ORCF. The EPI ranged between 1.22 and 5.91 each fluid. The larger the value of EPI, the higher the power tends to be generated: the EPI and power generation of n-butane were 5.91 (MPa·kg)/kJ and 12.9 kW, respectively. The EPI is good agreement with power generation performance each working fluids. Since the EPI considers not only the physical properties but also the mechanical parameters in the cycle, it could be predicted to the net power generation. It is clearly found that the EPI can be easy to estimate the net power generation for these conditions without any correction parameters. These findings will contribute to the spread of general-purpose small ORCs.

In this study, the EPI is a key index for predict the micro ORC performance and will be an index for future working fluid research, for example, objective function of materials informatics (MI).

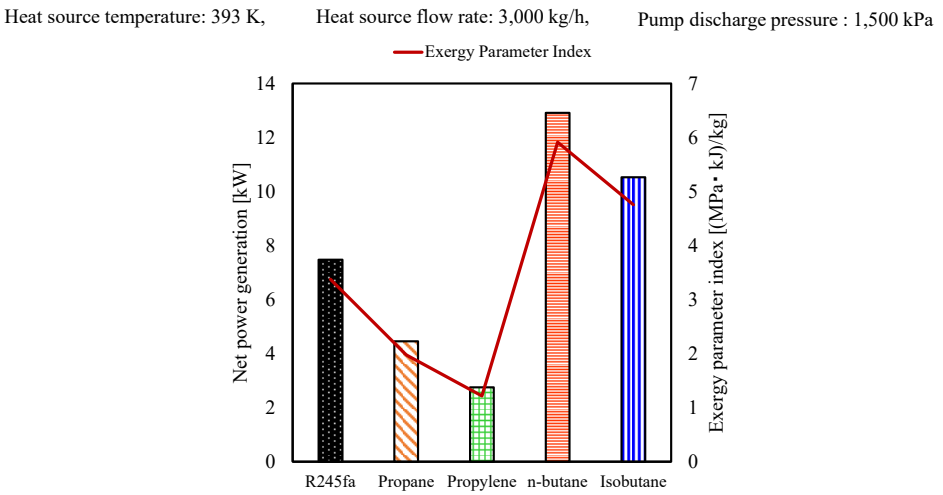


Fig. 3 The effect of net power generation on Exergy Parameter Index (EPI).

4. Conclusion

In this study, to improve the versatility of small scale ORCs and to clarify their design guidelines, a 10 kW class ORC (micro ORC) was developed on a model basis, an assumed that various working fluids operate in a single device, and the performance of each working fluid was investigated under fixed design conditions.

In the assumed ORC, n-butane showed the highest power generation performance among the selected working fluids. When the relationship between the power generation performance of each working fluid and the latent heat, sensible heat, and boiling point was investigated, no high correlation was found.

Therefore, we introduced a new parameter, Exergy Parameter Index, which can predict the power generation performance. It was defined the key parameter which considered sensible heat, latent heat and pressure difference in the cycle, and can be predict the power generation performance of the micro ORC. In this study, the EPI could design and operation conditions for the micro ORC.

The micro ORC may not achieve appropriate performance due to variations in environmental conditions such as heat source temperature. These conditions will increase the power generation cost, resulting in negative economic efficiency. After that, the micro ORC should be designed to be sustainable from the perspective of rigorous tecno-economic analysis.

5. Reference

Akimoto, R., T. Yamaki, M. Nakaiwa, K. Matsuda. Applicability study of micro Kalina cycle for regional low grade geothermal heat in Japan. *Case Studies in Thermal Engineering*, **28**, 101506 (2021)

Li, Y. M., Hung, T. C., Wu, C. J., Su, T. Y., Xi, H., Wang, C. C. Experimental investigation of 3-kW organic Rankine cycle (ORC) system subject to heat source conditions: A new appraisal for assessment. *Energy*, **217**, 119342 (2021)

Uusitaro, A., Saaresti, T. T., Honkatukia, J., Dhanasegaran, R. Experimental study of small scale and high expansion ratio ORC for recovering high temperature waste heat. *Energy*, **208**, 118321 (2020)

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50133-0

Scenario Outcomes for Electric Power Generation Expansion Planning considering the State of Indiana as a Case Study

Abdul Ahmed[a], Kayla L. Richardson[b], Yufei Zhao[c], Cornelius M. Masuku[c]*

[a]*Chemical and Biological Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, USA.*
[b]*School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe AZ 85287, USA.*
[c]*Davidson School of Chemical Engineering, Purdue University, West Lafayette, IN 47907, USA.*
cmasuku@purdue.edu

Abstract

This work assesses Indiana's electricity demands through the next few decades to identify an optimal combination. A mixed-integer linear programming mathematical modeling framework in which a combination of sources, generation technology, and capacity of future generation technologies are evaluated to minimize investment, operational and environmental costs is implemented. Our model is a modification of the implementation developed by Lara et al. (2018) with the data from databases such as the U.S. Energy Information Administration and other relevant agencies. Our model is implemented in Pyomo (Python optimization modeling objects), a free and accessible python-based software package for devising optimization models. While there may exist multiple combinations of energy sources that prove to be optimal when considering various sets of constraints, pertinent results are those that maximize the use of renewable energy sources while minimizing operational, environmental, and investment costs. This research aims to inform planning agencies of optimal energy infrastructure configurations for the State of Indiana, allowing for evaluations of the changes required to reach optimal operation with expected growing demand. The model and analysis of results can be applied to any other region (State or Country).

Keywords: Renewable Energy Integration; Mixed-Integer Linear programming; Optimization.

1. Main Text

In the advent of the current climate crisis, a transition to renewable energy systems has become widely apparent as a means to mitigate climate effects. In the past decade, within the United States, the use of renewable energy sources has risen. The desire to maximize the use of renewable energy sources stems from the fact that the alternative non-renewable, or traditional, power sources contribute more to the pollution of the environment (Lopez & Espiritu, 2011). However, energy data collected by the Energy Information Administration (EIA) shows that traditional energy sources, such as fossil fuels (petroleum, natural gas, and coal) and nuclear, continue to generate over 85% of total energy consumed in the U.S (U.S. EIA, 2021). Many factors, such as policy, public sentiment, and scientific advancements currently assist in driving the transition towards

renewable energy systems. Many nations stand at this pivotal moment of restructuring. However, with this transition evolves a problem in terms of, what combinations of renewable and non-renewable technologies are most optimal to meet required energy demand?

Energy optimization models can be used to present this combination of energy sources. These models consider predictions of energy demand over several decades and determine which sources of energy best fulfil future energy needs. These models can then offer insight on creating a more sustainable energy system, presenting ideal evolutions of that energy system over multiple decades, without declaring how probable these evolutions are (Pfenninger et al., 2014). This work aims to formulate a multi-objective optimization modeling framework to evaluate the necessary changes in the energy systems infrastructure of the state of Indiana that meets EIA projected energy demand while minimizing environmental and capital costs. The research focuses on the increased use of renewable energy sources of solar and wind and the non-renewable, natural gas combined cycle. These three technologies are the expected drivers of future energy generation.

2. Research Methods

2.1. Collection of Data

In this research, we assessed the potential of solar and wind generative technologies to fulfil demand imposed by planned retirement, economic growth, and population growth, all while minimizing both environmental and capital costs. To do this, we created a linear program with an objective function of minimizing the implementation cost for a selected set of generative technologies. The model was informed with data retrieved primarily from the NREL (National Renewable Energy Laboratory) and the EIA (Energy Information Administration). The data consists of two primary types, cost, and performance. To drive the cost analysis aspect of the objective function, annual technology baseline (ATB) data for the U.S. of the year 2020 was retrieved from the NREL. This baseline data provided metrics of cost per performance for renewable and conventional generative technologies. The secondary data set consisted of supply curves of solar and wind for the U.S. Supply curve data for each technology consisted of the following: location (longitude and latitude), generative capacity potential, the area available, capacity factor, and distance to interconnect. Excluding transmission, all listed parameters were relevant to inform the model. In addition to these data sets, other data was prevalent in the formulation of the model. For instance, generator data, retrieved from the U.S. EIA, was used to determine the level of demand imposed by planned retirement for the ten-year time horizon (2020-2030). Relevant contributions included listings of planned retirements of coal generators and their respective nameplate capacities. The summations of these nameplate capacities served as the demand imposed by the planned retirement of the ten-year time horizon. For the twenty- and thirty-year time horizons, given the lack of data for planned retirement, linear extrapolation was applied to determine respective demand coefficients. Lastly, data from the EIA 2021 Annual Energy Outlook allowed the determination of time factored, and population factored growth in demand with applications of extrapolation techniques.

2.2. Treatment of Data

In the process of informing the computational model, the data is pre-processed with Python via Jupyter Notebook. As the focus of the research is the state of Indiana, a subset

of supply data of Indiana was devised from the original supply curve data - representative of the whole United States. We employed a rigorous approach to accomplish this. Firstly, border and city location data were retrieved from online sources. This data mapped the border and cities of Indiana. With the Pandas package and data frames, border and city locations were unionized. A grouping of supply data, based on proximity, was then attached to the location data derived by the union of border and city locations. These locations were then mapped as nodes on a geospatial plot, providing an informative supply curve fitted to the state of Indiana. With this approach, we successfully elicited Indiana's supply data from that of the country.

To further refine the data, locations were sorted. The sorting convention was from greatest to least by the following ratio: potential generative capacity to the area of available land. High valued ratios are indicative of richer location quality. Quality, we define as the amount of generative potential per unit area of land. Following this definition, high qualities are desirable as they minimize the amount of land needed for system implementation, which reduces the associated cost of implementation. From this heuristic, a set of fifty nodes (locations) for solar and wind were developed, totaling one hundred decision variables of which selection proportions range from zero to one.

2.3. Mathematical Modeling

A mathematical model was constructed with Pyomo, a python-based optimization package. The optimization problem was solved with glpk as a solver. The data elicited from the model include generator location, generator type, and proportion of capacity selected per selected location. The framework of the mathematical model follows that of a linear program. The objective function is given as Equation (1). The objective function seeks to minimize the total implementation cost of the selected generative technologies and their respective capacities. C^t is the cost of implementing a selected configuration. C^t_{solar} is the average cost of solar per MW generative capacity for the time horizon of consideration. Likewise, C^t_{wind} is the average cost of wind per MW generative capacity for the time horizon of consideration and C^t_{ng} is the average cost of natural gas per MW generative capacity for the time horizon of consideration. The decision variables are x_i, the amount of solar capacity selected, and x_j, the amount of wind capacity selected. S represents solar capacities, W, wind capacities, and NG, natural gas capacities. The respective sets comprising selections for each decision variable are SS and WS. Lastly, t is used to represent the time horizons with subscripted variables dependent on time.

The first constraint is a peak demand constraint, Equation (2). This informs the selection of generative technologies and their respective capacities to ensure that the sum required load (load loss from the retirement and load expected as a result of population and economic growth) is met. In Equation (2), ED_t is the demand imposed by retirement, and the expected growth of population and economy for the planning horizon. Equation (3) confirms that the imposed demand for a given time horizon is equal to the sum of the demand imposed by retirement (R_t) within the horizon, and the contribution to demand from expected economic and population growth (E_t) within the horizon. Equation 4 proposes an intermittency constraint. In order to ensure that a proposed configuration is at least as reliable as the current energy infrastructure of Indiana, Equation 4 was developed to assess whether the realizable load (peak load multiplied by the capacity factor) is greater than or equal to the summation of the realizable load imposed by planned retirement and economic and population growth. $CF^s, CF^w,$ and CF^{ng} are respectively the solar, wind, and natural gas capacity factors. CF^c is the coal capacity factor, and CF^{cc}

is the aggregated averaged capacity factor of solar, wind, and natural gas. The relevant sets are: CS, the Solar Capacity factor set; and CW, the Wind Capacity factor set. The last defined constraint is represented as Equation (5), the emissions constraint. This constraint allows for the assessment of scenarios of varying reliance on natural gas. Natural gas plays a pivotal role in integrative renewable energy systems. However, to limit the environmental effects of natural gas, this constraint controls emission tolerances accordingly. In Equation (5), x serves as the proportion of fulfilled demand occupied by natural gas.

$$\min\ C^t = \left(C^t_{solar} \sum_{i \in SS} S_i x_i\right) + \left(C^t_{wind} \sum_{j \in WS} W_j x_j\right) + \left(C^t_{ng} NG\right), \forall t \in \{10,20,30\} \tag{1}$$

$$\sum_{i \in SS} S_i x_i + \sum_{j \in WS} W_j x_j + NG \geq ED_t \quad \forall\, t \in \{10,20,30\} \tag{2}$$

$$ED_t = R_t + E_t, \quad t \in \{10,20,30\} \tag{3}$$

$$\sum_{i \in SS, a \in CS} S_i x_i CF^s_a + \sum_{j \in WS, a \in CW} W_j x_j CF^W_a + NG * CF^{ng} \geq (CF^c R + CF^{cc} E) \tag{4}$$

$$NG = xED_t \tag{5}$$

3. Results and Discussion

We initially considered the scenario of doubling current renewable generative capacity in the state of Indiana for a 5-year horizon. The objective was to determine the optimal configuration of counties achieving this specification. In our analysis, we observed that in 5 years, doubling renewable energy generative capacity reduces carbon emissions by 20%. Figure 1 compares the relative distribution of current generation technologies to what can be achieved from the considered scenario. The purpose of this illustration is to emphasize the motivation behind this study.

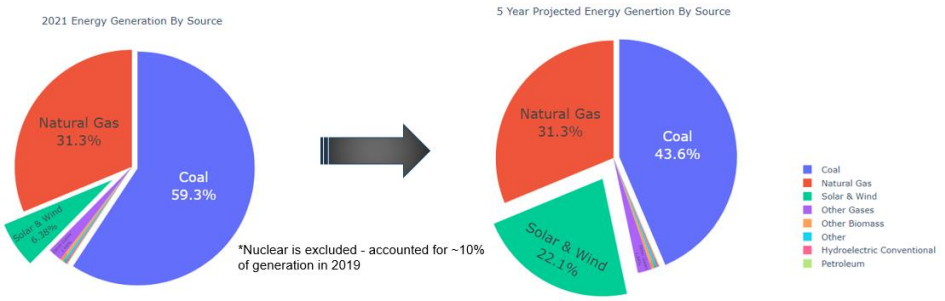


Figure 1: Comparison of relative distribution of current generation technologies to a 5-year projection when considering the scenario of doubling current renewable generative capacity.

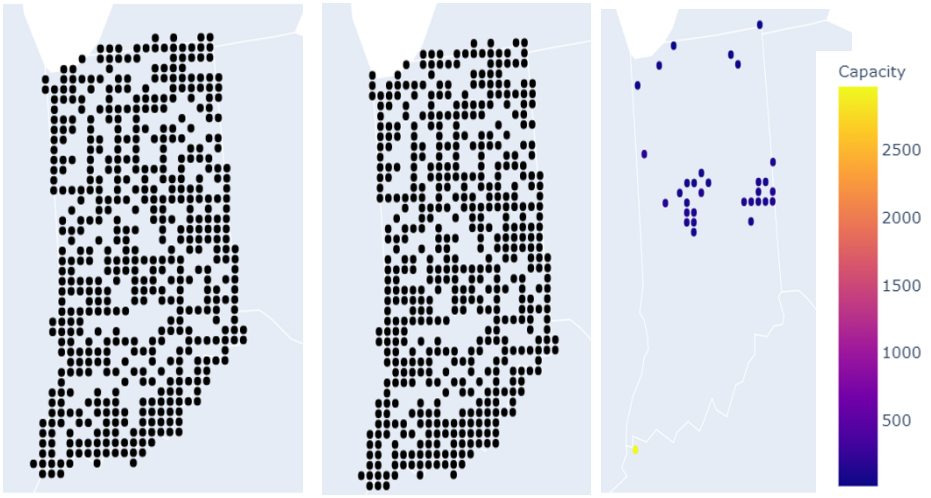
The currently developed mathematical framework aimed to analyze three different time horizons, each incremented by a decade. As expected, the cost of implementing the determined optimal configuration rose each horizon. The driver of this rise was an

increase in imposed demand as dictated by expected increases in planned retirement, economic growth, and population growth. Yet, this was somewhat balanced by the expected decreasing cost of implementing solar and wind generative technologies, resulting from developments in technology. Overall, spanning the three distinctive time horizons, a net increase in implementation cost is observed.

From the ten-year time horizon with a reliance on generation constituting 50 % natural gas, the calculated implementation cost was 28.2 billion USD. The distribution of solar to wind, based on contribution by generative capacity, is depicted in Figure 2. Each node is informative of both longitude and latitude. For the twenty-year time horizon of the same reliance factor, the calculated implementation cost was 84.1 billion USD. However, when the twenty-year model is fitted with data consisting of one hundred of the best (as defined by the heuristic in the treatment of data section) generative wind and solar sites, each, the computed implementation cost is 78.5 billion USD.

(a) (b) (c)

Figure 2: Nodes of selectable generation. (a) solar generation; (b) wind generation; (c) selected configuration of wind and solar for a 10-year horizon with associated capacities. Yellow nodes are representative of solar, blue nodes representative of wind.



The configuration derived from this heuristic is more cost-effective as the higher consideration of wind allows for an overall increase in system reliability. Because of Equation (4), a larger selection of wind generative capacity reduces intermittency. This reduction in intermittency promotes a reduction in the net load of the selected configuration, which lowers the cost of implementing the optimized configuration. To significantly reduce costs, technological developments focused on increasing the reliability of wind and solar generators are warranted. The final analysis was the variation of the proportionality factor of natural gas deployed, x. This variation was carried out for the ten-year horizon. The first considered scenario was x = 0.25, the fulfilment of twenty-five percent of the imposed demand by natural gas. The cost of implementation increased to 38 billion USD. The second scenario was the complete minimization of environmental costs. Such is representative of absolute reliance on solar and wind. The cost of

implementing the optimal configuration for this consideration is 51 billion USD. These sub scenarios are significant in portraying the relationship between capital costs of implementation and the amount of CO_2 emitted, with amounts of CO_2 emitted directly proportional to the demand fulfilled by natural gas.

The model developed for this study is very versatile to improvement. Future research could consider, for example, transmission which would require a more complex design and analysis of cost factors and energy retention. Another favorable consideration is time factorization, which is a means for better controlling intermittency. This consideration could entail the deployment of natural gas only when wind/solar is unavailable during the day. A third consideration is the implementation of storage technologies. Storage can act to negate solar and wind downtime. However, much research is necessary for the effective development and deployment of storage solutions.

4. Conclusions

This research assessed the capabilities of implementing solar, wind, and natural gas generative technologies to replace the retirement of many coal plants and meet the demand imposed by economic and population growth in the state of Indiana. In theory, the most practical way to reduce the effects of environmental degradation is a complete shift to renewable generation technologies. However, many factors impede this shift. Such factors include capital costs and intermittency. Of these two impeding factors, the most significant is that of intermittency. Due to the limited availability of wind and sunlight, energy infrastructures cannot exist solely by wind and solar. Thus, the practical solution has been the integration of renewable energy sources into current energy infrastructures. Such insertions also allow for the displacement of polluting generation sources, such as coal. However, because of intermittency, integrative systems suffer in reliability. Despite the localized nature of this study, the approaches, methods, and findings are well applicable to other regions of the U.S. and the world. The global energy situation is a complex subject, hence works such as this aim to produce informed and accurate conclusions to aid transitions when considering integrative energy infrastructures.

References

S. Akar, P. Beiter, W. Cole, D. Feldman, P. Kurup, E. Lantz, R. Margolis, D. Oladosu, T. Stehly, G. Rhodes, C. Turchi, L. Vimmerstedt (n.d.). *2020 annual technology baseline (ATB) cost and performance data for Electricity Generation Technologies.* NREL Data Catalog. Retrieved October 26, 2021, from https://data.nrel.gov/submissions/145.

C.L. Lara, D.S. Mallapragada, D.J. Papageorgiou, A. Venkatesh, I.E. Grossmann, 2018. Deterministic electric power infrastructure planning: Mixed-integer programming model and nested decomposition algorithm. *European Journal of Operational Research, 271*, 1037–1054.

N. Lopez, J.F. Espiritu, 2011. An approach to hybrid power systems integration considering different renewable energy technologies. *Procedia Computer Science, 6*, 463–468.

S. Pfenninger, A. Hawkes, J. Keirstead, 2014. Energy systems modeling for twenty-first century energy challenges. *Renewable and Sustainable Energy Reviews, 33*, 74–86.

U.S. Energy Information Administration, 2021. Annual Energy Outlook 2021. Accessed at: https://www.eia.gov/outlooks/aeo/pdf/AEO_Narrative_2021.pdf.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50134-2

Requirements for the quality assessment of virtual commissioning models for modular process plants

Isabell Viedt[a*], Jonathan Mädler[a], Julius Lorenz[a], and Leon Urbas[a]

[a]*Chair of Process Control System and Process Systems Engineering Group, Technische Universität Dresden, 01062 Dresden, Germany*
isabell.viedt@tu-dresden.de

Abstract

Fast changing markets and higher demands for higher flexibility in the process industry in general, require modular concepts that decrease the necessary time-to-market. This requires process simulation in earlier engineering phases and make virtual commissioning of the modular process plants besides conventional commissioning essential. Quality constraints of the model for virtual commissioning also have to be considered to confidently trust the results and findings during virtual commissioning. A particular hard problem is the quality assurance of the third-party simulation models. Our currently investigated research hypothesis is that the VC application puts an emphasis on the factors efficiency, maintainability and compatibility as defined in the quality model approach for software quality assessment for a mapping towards process simulation. This paper addresses the necessary requirements for the quality assessment of simulation models for the purpose of virtual commissioning and presents factors, criteria and metrics for the assessment of virtual commissioning models. This paper extends the current framework for functional quality assessment of simulation models in smart equipment to the use case of virtual commissioning models and is a first step to the automated quality assessment for virtual commissioning models. Further, the need for model certification is discussed.

Keywords: Modular Plants; Process Operation; Quality Assessment; Quality Assurance; Virtual Commissioning.

1. Introduction

Fast changing markets and process conditions in the process industry create an increasing demand for flexibility of process plants. One approach for adapting to those demands are modular process plants (VDI, 2020). This standardized modular plant concept strives to be manufacturer independent for the configuration of process modules, which are called Process Equipment Assemblies (PEAs). Automation and orchestration requires standardized interfaces and communication protocols (Süß et al., 2016) which are described within the Module Type Package (MTP) (VDI/VDE/NAMUR, 2019). The orchestration of those plants is realized through the Process Orchestration Layer (POL) that supports the operator in the configuration of the modular process plants and allows service-based control (Bloch et al., 2018) with an emphasis on continuous processes. With modular process plants, new plant configuration and recipes must be tested before the start of production to assess feasibility and to identify possible optimizations (cf. Schenk et al., 2019). One possible strategy is the utilization of Virtual Commissioning (VC) which allows the feasibility and optimization assessment at an early stage of engineering. Since modular process plants will integrate different PEAs and components from different manufacturers, several third-party simulation models will be integrated the

VC framework. One defining aspect is therefore the quality assessment of those models to ensure trust in the quality of VC.

In this paper, section 2 describes current trends for virtual commissioning in modular process plants and further discusses current quality assessment approaches for simulation models. In section 3 the current approach to quality assessment of simulation models for virtual commissioning and the problems with the integration of third-party models is discussed. Quality factors, criteria and metrics specifically relevant for virtual commissioning models are shown. Section 4 presents the concept for the application of the quality assessment for virtual commissioning models to VC schemes. Further, current issues and research needs of applying quality assessment to virtual commissioning are addressed. The obtained requirements aim to be the basis for a future quality assessment framework implementation for virtual commissioning models in smart, modular plants.

2. State of the art

2.1. Virtual Commissioning for modular process plants

As with any process plant, the plant configurations of modular process plants must be validated before the start of production. Virtual commissioning (VC) is a good approach to cut cost, pre-qualify plant configurations and assess feasibility before real commissioning (Puntel-Schmidt et al., 2015). VC includes the testing of individual components, interlocks and functions of the automation system during development via simulation methods and models adapted to the respective purpose (Schenk et al., 2019). Therefore, models must consider actors and sensors in the plant, which are coupled with controllers in a Hardware-in-the-Loop (HIL), Software-in-the-Loop (SIL) or Model-in-the-Loop (MIL) architecture (cf. VDI/VDE 3693). Simulation models are either simple state machines or dynamic simulation models of varying fidelity (cf. VDI/VDE 3693).

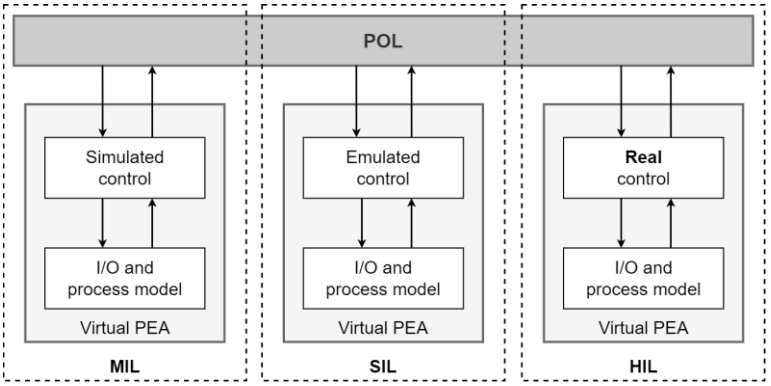


Figure 1: Test configurations for Virtual Commissioning (VC)

In modular plants, virtual commissioning can be used to safely test out new plant configuration and recipes regarding feasibility and optimization (cf. Schenk et al., 2019). Faults can therefore be detected and fixed in an early engineering phase which cuts cost and the overall time-to-market (Klose et al., 2019). Klose et al. (2021) suggest a digital twin applying the MTP for the integration of the simulation models. With this approach, the digital twins (DT) of the PEAs are integrated into a simulation or co-simulation environment and coupled to the POL. This considers the automation system including device and I/O-models, the controller in a SIL or MIL (cf. VDI/VDE 3693) scheme and

a default material system (e.g. water run). To test service sequences, an OPC UA server can be utilized (Schenk et al., 2019). Furthermore, the water run models can be extended with physical property packages to ensure a better understanding of the service parameters and the process timing. Other than the interfaces for property packages, simulation control, and mass and energy flows, the interfaces for the information flow must be reconfigured to operate with the same service commands and data assemblies as the real PEA

2.2. Quality assessment for simulation models

Simulation models that are to be used in virtual commissioning of modular plants will be provided by PEA manufacturer. The partial models needed for VC are then integrated into a virtual commissioning framework, e.g. as co-simulation. To assure the quality of these simulation models assessment strategies are required. Currently used methods for that purpose focus mainly on accuracy of the model (Sargent, 2013). To integrate the quality assessment and control of simulation models into modern digital plants, a framework for what quality of a simulation models actually entails needs to be developed and implemented. A first approach to quality assessment of simulation models with pre-definition of quality attributes was proposed in Mädler et al. (2021).

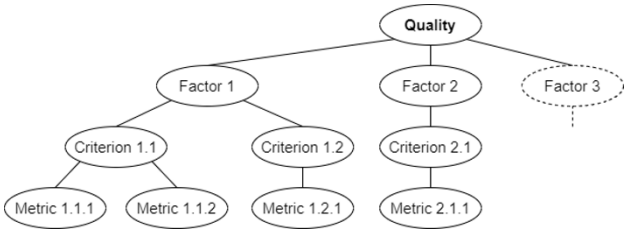


Figure 2: FCM model structure (McCall & Cavano, 1978)

With simulation models becoming an integral part of PEAs and plant equipment in general, simulation models can be considered a part of software. This allows the utilization of quality assessment strategies from software development. Current simulation model validation and verification methods (Sargent, 2013) are coupled with the approach of FCM (factors – criteria – metrics) models (McCall & Cavano, 1978) for structuring and executing quality assessment. One approach to FCM models, as mentioned in Mädler et al. (2021), is the ISO/IEC 25010 for the assessment of software product quality. The exemplary structure of a FCM model is shown in Fig. 2.

3. Requirements for third-party model integration

Virtual commissioning allows a pre-qualification of modular process plant configurations, recipes and interfaces while utilizing simulations models for the process in combination with models for the devices, I/O and controller representations. This means that not all models will be provided by the same manufacturer. Since model for virtual commissioning put an emphasis on the assembly of the VC framework from different partial models which varying depth and purpose, one important aspect to realizing the interchangeability of simulation models in VC and therefore individual PEAs is the integration of third-party models.

The quality model presented in Mädler et al. (2021) focuses on factors like functional suitability and reliability because the quality assessment is in this case intended for

models for process simulation and optimization. For virtual commissioning, probably the most important aspect of models is the ability to combine the necessary partial models to a VC framework (Schenk et al., 2019). This makes the consideration of further quality factors necessary (cf. Mädler et al., 2021). As of now the framework only considers the quality factors functional suitability and reliability but with the extension of the use case to virtual commissioning model, the focus shifts to the factors compatibility, maintainability and performance efficiency. Other metrics like the validity domain of the model, as a metric for functional suitability, that specifically targets the design space of the model loses importance in the context of virtual commissioning.

3.1. Model compatibility

In this case the factor compatibility is defined through the quality criterion interoperability. The concept of interoperable models is not new and can be achieved through a standard for the interface to exchange information and models. They are used to integrate custom physical property packages, exchange mass, energy and information flows, and control the solution process of the simulation model. One approach is the CAPE-OPEN standard mostly for steady-state models (van Baten and Pons, 2014). Since CAPE-OPEN is an interoperability model and not a data model, the interoperability between different process modeling environments is not within the scope of CAPE-OPEN. This makes the CAPE-OPEN interface specification less appealing for dynamic process simulation since it does not initially enable co-simulation between third-party models on different simulation environments. Another standard is the functional mock-up interface (FMI) for dynamic models which allows import and export of both plant and controller models into and out of any simulation packages that support the specification (Blockwitz et al., 2012). Within FMI interfaces for both model exchange and co-simulation are covered. The co-simulation interface is intended for use with models where data is exchanged between subsystems only at discrete communication points and each subsystem is solved independently (Blockwitz et al., 2012). The model exchange interface is intended for use with models described by differential, algebraic and discrete equations with or without discontinuities where the system equations are solved simultaneously (Blockwitz et al., 2012).

3.2. Model maintainability

Besides the compatibility of the simulation for the VC framework, the maintainability (quality factor maintainability) of those models has to also be considered. Especially for virtual commissioning of modular plants different plant configurations have to be considered and tested. This means that the quality criterion reusability also needs to be considered to describe the maintainability of the models and the framework. With this it is important that different partial models can be reuse for different configuration of the MP to allow virtual commissioning. The models must be parameterizable to allow testing of recipes (Schenk et al., 2019) but also allow a reuse through a standardize interface (Fedorova et al., 2015). Furthermore, the transfer of models or model information into different simulation environments through XML allows a wider applicability (Fedorova et al., 2015), which is especially important for model exchange frameworks for virtual commissioning.

3.3. Performance efficiency

Another important quality factor to consider for models for VC applications is performance efficiency. In this context, the quality factor is defined by the quality

criterion time-behaviour. For simulation models or digital twins used for virtual commissioning this means that real time and accelerated, hybrid continuous- and discrete-time simulation must be available to test recipes and interlocks as fast as possible in a virtual environment. To achieve synchronized, accelerated simulation, the POL must be able to operate in an accelerated mode as well.

4. Quality assessment for virtual commissioning models

The addition of three newly considered quality factors and their corresponding quality criteria and metric will make the quality assessment framework more holistic but also complex. This means that new metrics and assessment strategies have to be found to integrate the new factors. The proposed strategy shows that this means a shift from current quantitative assessment methods to more qualitative assessment metrics for the virtual commissioning models. Table 1 shows these additional quality factors, criteria and metrics which must be implemented into the framework in a next step.

Table 1: FCM model for the assessment of virtual commissioning models

Factor	Criterion	Metric
Compatibility	Interoperability	Interface standard adherence
		Interface standard compatibility
Maintainability	Reusability	Parameterizability
		Interface standard adherence
		Modeling environment transferability
Performance efficiency	Time-behaviour	Flexible simulation mode
		POL-compatible simulation mode

While a check for compliance with current interface standards for model exchange is easy to implement into an assessment framework, it is not easy to solve non-compliance of third-party models. Non-compliance with current standards will lead to workarounds and adjustments for successful application of virtual commissioning for modular plants. One strategy to ensure model compatibility could be a certification approach. Independent model certification has long been a topic of interest for modeling and simulation tasks but no concrete certification entities exist yet (Balci, 2010).
As virtual commissioning becomes increasingly important and security aspect will be necessary to consider, the factor of security as described in ISO/IEC 25010 will also need to be considered during systematic quality assessment of models, especially for virtual commissioning. This is especially important in the context of the criterion reusability.

5. Conclusion

With this paper, the first step towards systematic and potentially automated quality assessment for virtual commissioning models is taken. The requirements defined in this paper serve as a basis for the development of a quality assessment framework. The proposed concept outlines that the quality of models for virtual commissioning and the corresponding wish for potential automated assembly of the VC framework depends on additional quality factors in comparison to dynamic models used for process design. Interfaces for physical property packages, flows and simulation control must be developed further to allow tool independent usage and exchange of models. Furthermore, tools that potentially allow the transfers of models, or at least their information, for reuse in different environments also must be addressed. Depending on the use case and complexity of what VC is supposed to do in the future, new requirements to the models

might arise and security concerns must be addressed. This paper extends the in Mädler et al. (2021) proposed framework for automated quality assessment of simulation models. As the use case for the simulation models shifts to virtual commissioning, new quality aspects need to be considered and others loses importance. The proposed extension of the framework now needs to be transferred to real world applications to assess the usability and integration of said concept. Since the proposed requirements and concept are only a first step, the new factors, criteria and metrics must be implemented into the MATLAB/Simulink framework and evaluated via case study.

References

O. Balci, 2010, Golden rules of verification, validation, testing, and certification of modeling and simulation applications, SCS M&S Magazine, 4, 4, 1-7H. Bloch, S. Hensel, M. Hoernicke, K. Stark, A. Menschner (neé Hahn), A. Fay, L. Urbas, T. Knohl, and J. Bernshausen, 2018, State-based control of process services within modular process plants, Procedia CIRP, 72, 1088–1093

T. Blockwitz, M. Otter, J. Akesson, M. Arnold, C. Clauss, H. Elmqvist, ..., and A. Viel, 2012, Functional mockup interface 2.0: The standard for tool independent exchange of simulation models, In ProceedingsISO/IEC, Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models (ISO/IEC 25010:2011) (2011)

M. Fedorova, G. Tolksdorf, S. Fillinger, G. Wozny, M. Sales-Cruz, G. Sin, and R. Gani, 2015, Development of Computer Aided Modelling Templates for Model Re-use in Chemical and Biochemical Process and Product Design: Import and export of models. In Computer Aided Chemical Engineering, 37, 953-958

A. Klose, S. Merkelbach, A. Menschner, S. Hensel, S. Heinze, L. Bittorf, ..., and L. Urbas, 2019, Orchestration requirements for modular process plants in chemical and pharmaceutical industries, Chemical Engineering & Technology, 42, 11, 2282-2291

A. Klose, T. Schenk, R. Rosen, A. Botero, C. Schäfer, P. Da Silva Santos, S. Merkelbach, J. Lorenz, and L. Urbas, 2021, Virtuelle Inbetriebnahme modularer Prozessanlagen: Fallstudien zur Funktionsqualifizierung MTP-basierter Orchestrierung. atp magazin, 63, 04, https://doi.org/10.17560/atp.v63i04.2542

J. Mädler, I. Viedt, and L. Urbas, 2021, Applying quality assurance concepts from software development to simulation model assessment in smart equipment, Computer Aided Chemical Engineering, European Symposium on Computer Aided Process Engineering (ESCAPE), Istanbul

M. Oppelt, and L. Urbas, Otctober 2014, Integrated virtual commissioning an essential activity in the automation engineering process: From virtual commissioning to simulation supported engineering, In IECON 2014-40th Annual Conference of the IEEE Industrial Electronics Society, IEEE, 2564-2570

P. Puntel-Schmidt, and A. Fay, 2015, Levels of detail and appropriate model types for virtual commissioning in manufacturing engineering, IFAC-PapersOnLine, 48, 1, 922-927

T. Schenk, A. Botero Halblaub, R. Rosen, T. Heinzerling, J. Mädler, A. Klose, S. Hensel, and L. Urbas, 2019, Co-Simulation-based virtual Commissioning for modular process plants—Requirements, Framework and Support-Toolchain for a Virtual Automation Testing Environment. Automation 2019, 229–242

S. Süß, D. Hauf, A. Strahilov, and C. Diedrich, 2016, Standardized Classification and Interfaces of Complex Behaviour Models in Virtual Commissioning. Procedia CIRP, 52, 24-29

J. van Baten, and M. Pons, 2014, CAPE-OPEN: Interoperability in Industrial Flowsheet Simulation Software. Chemie Ingenieur Technik, 86, 7, 1052-1064

VDI, 2020, Modular plants - Fundamentals and planning modular plants (2776-1:2020-11)

VDI/VDE/NAMUR, 2019, Automation engineering of modular systems in the process industry – General concept and interface – Part 1 (2658-1:2019-10)

VDI/VDE, 2016, Virtual Commissioning – Model types and glossary (3693:2016-08)

14th INTERNATIONAL SYMPOSIUM ON PROCESS SYSTEMS ENGINEERING

VOLUME 2

COMPUTER-AIDED CHEMICAL ENGINEERING, 49

14th INTERNATIONAL SYMPOSIUM ON PROCESS SYSTEMS ENGINEERING

VOLUME 2

Edited by

Yoshiyuki Yamashita
Professor and Chair of Department of Chemical Engineering
Tokyo University of Agriculture and Technology (TUAT), Tokyo, Japan
yama_pse@cc.tuat.ac.jp

Manabu Kano
Professor, Department of Systems Science
Kyoto University, Kyoto, Japan
manabu@human.sys.i.kyoto-u.ac.jp

ELSEVIER

Amsterdam – Boston – Heidelberg – London – New York – Oxford
Paris – San Diego – San Francisco – Singapore – Sydney – Tokyo

Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, USA

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

ISBN (Volume 2): 978-0-443-18725-4
ISBN (Set) : 978-0-323-85159-6
ISSN: 1570-7946

Publisher: Susan Dennis
Acquisition Editor: Anita Koch
Editorial Project Manager: Lena Sparks
Production Project Manager: Paul Prasad Chandramohan
Designer: Greg Harris

Typeset by STRAIVE

Contents

Contributed Papers: Modeling, Analysis, and Simulation

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50135-4

A combinatorial tool for monitoring flocculation processes: Using non-invasive measurements and hybrid deep learning assisted modelling

Nima Nazemzadeh[a], Josep Serra Olivé[a], Rasmus Fjorbak Nielsen[a], Krist V. Gernaey[a], Martin P. Andersson[b], Seyed Soheil Mansouri[a,*]

[a] *Process and Systems Engineering Centre (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søltofts Plads, Building 228A, 2800 Kgs. Lyngby, Denmark*

[b]*Combustion and Harmful Emission Control Centre (CHEC), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søltofts Plads, Building 228A, 2800 Kgs, Lyngby, Denmark*

seso@kt.dtu.dk

Abstract

Particulate processes have a wide range of applications in many different industries, from wastewater treatment to the pharmaceutical industry. Despite their extensive applications, control and monitoring of chemical and biochemical processes that contain solid particles are challenging due to the lack of fundamental understanding of the process mechanism and the limited availability of real-time process data. In this study, a hybrid multiscale framework is introduced for flocculation processes as a particulate process, and it is validated against experimental data resulting from the flocculation of silica particles. The variations of the particle size distribution are imposed by varying the pH in different experimental batches. In this study, an integrated hybrid deep learning approach combining deep learning with first principles is implemented to predict the future state of the process. The first-principles model combines a population balance model with surface properties of the particles calculated with computational chemistry, while the deep learning model is a deep neural network.

Keywords: Hybrid modelling, flocculation, interactions, multiscale modelling.

1. Introduction

The flocculation process consists of the separation of solid particles from a liquid medium by the aggregation of two or more particles that collide and cluster as an aggregate/agglomerate. In the flocculation process, the stability of the suspension is influenced by parameters, as the particle properties (charge, morphology, characteristics), properties of the flocculant (concentration, type) or properties of the media (pH, ionic strength). Previous studies have investigated the correlation between media conditions and polyelectrolyte (PE) properties. For example, a low charge density is a limiting factor for polyelectrolyte adsorption and retention. Nevertheless, lower ionic strength results in a decreasing surface charge, which can also decrease the flocculation efficiency of the

flocculant. Adsorption can be controlled by the addition of salt (ionic strength) in the PE solution (Scheepers et al., 2021).

In this context, the industry could benefit from the development of a model to predict the optimum flocculant dosage in a particular system. However, specifically in the case of the flocculation process, a hybrid multiscale model can facilitate obtaining more precise predictions of the future state of the process by integrating the physical understanding of the process phenomena with the underlying features of the data. In this case, deep learning models can be used to develop correlations between states of the process that cannot be explained by the available first-principles models.

In this study, the first-principles models are implementations of a population balance model with surface charge properties that are derived from computational methods in chemistry, and the deep learning model is an architecture of deep neural networks (DNNs). These two models are integrated as such that the parameters that are difficult to measure or estimate are being estimated by the DNN and the predictions of the future state of the process are carried out by the first-principles model.

2. Modelling framework

A hybrid multiscale framework is developed based on earlier work (Nazemzadeh et al., 2021a; Nielsen et al., 2020). The inputs for the neural network will be the data collected from the experiments including pH, polymer dosage, particle morphology. Then the neural network will be trained to estimate the kinetic parameters by determining underlying correlations among the process state variables. However, the first-principles model is a discretized population balance model (PBM) with computational and theoretical chemistry models. The PBM is the component that predicts the particle size evolution with the assistance of the parameters estimated from DNN and the surface properties determined from computational and theoretical chemistry models.

The computational chemistry models are employed to calculate the surface charge density, surface potential, and solid-liquid interfacial energy of the particles vs. medium conditions (pH, ionic strength) by using a mean-field model (Andersson et al., 2020). The theoretical chemistry model used in this study is based on the DLVO theory (Yotsumoto and Yoon, 1993). The theory is named after the four scientists (Derjaguin, Landau, Verwey, and Overbeek) that developed the concept. It uses theoretical chemistry approaches to quantify the interaction energy among particles in a system. The implementation of this theory on a silica particle flocculation was demonstrated earlier (Nazemzadeh et al., 2021b). However, in this study, a computational chemistry approach is used rather than a completely theoretical approach to determine the surface interaction among particles. This potentially leads to a more accurate prediction of the surface properties.

In this study, the flocculation process is considered as the combination of aggregation, where particles interact with each other to form bigger particles, and the breakage phenomenon, where flocs break to form smaller sized particles. However, solving the partial integro-differential equation describing the PBM numerically is not trivial, and different methods have been proposed. In this study, the PBM is solved using a discretization method, where the PSD is distributed in size bins where flocs with similar

sizes are grouped, to estimate the variation rate of the aggregates during the process (Kumar and Ramkrishna, 1996). The discretized PBM is represented in Eq. (1):

$$\frac{dN_i}{dt} = \sum_{j,k}^{j\geq k}\left(1-\frac{1}{2}\delta_{j,k}\right)\eta_i\alpha_{j,k}\beta_{j,k}N_jN_k - N_i\sum_k \alpha_{i,k}\beta_{i,k}N_k + \sum_{j\geq i}\gamma_{j,i}S_jN_j - S_iN_i \qquad (1)$$

$$v_{i-1} \leq (v_j + v_k) \leq v_{i+1}$$

Where N_i is the concentration of aggregates in the size bin i, $\delta_{j,k}$ is the Dirac delta function to avoid calculating collision of the same size particles twice, η is a proportional coefficient assigning the fraction of the floc v_i from the aggregate $(v_j + v_k)$, α is the collision efficiency, β is the collision frequency of two aggregates, γ represents the breakage distribution function, and S is the breakage rate of the flocs.

The hybrid model framework is represented in Figure 1. This modelling framework is developed based on the previous work (Nazemzadeh et al., 2021a; Nielsen et al., 2020), which integrates computational chemistry calculations within the population balance model to more accurately estimate the kinetic parameters. The output of the neural network will be the unknown parameters of the population balance model that cannot be explained with the first-principles model. The population balance model evaluates the particle size evolution through time based on the kinetic rates, the data collected from experiments, and complementary non-observable information estimating the interaction energies from computational chemistry approaches. It must be noted that the framework is structured such that density functional theory is carried out in the outer loop of the hybrid model. To evaluate the deviation of the predictions from experimental data an L1norm loss function is defined, which determines the absolute differences between the model-based predictions and the experimental data.

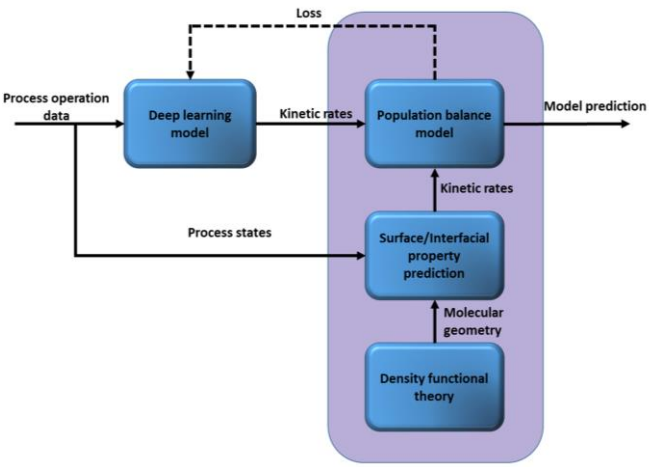


Figure 1: An overview of the hybrid multiscale framework

3. Application example

The proposed hybrid model framework is demonstrated through a laboratory-scale flocculation with silica particles in water, a common model system for flocculation. The particle size distribution is monitored within different batches by monitoring pH as the key process variable. The material and methods are fully described in (Nazemzadeh et al., 2021a). According to previous experiments, pH must be kept between 2 and 3, as this corresponds to the point of zero charge (PZC) (Barisik et al., 2014), overcoming the repulsion forces between particles with the same charge. This is also illustrated by our computational chemistry calculation (mean-field model) in Figure 2. The particles are neutral within the claimed PZC and above that range, the particles are negatively charged. It must be noted that particles with a diameter above 500 nm have extremely close surface (Barisik et al., 2014) charge density as the calculations are only done for nanoparticles with a diameter of 500 nm.

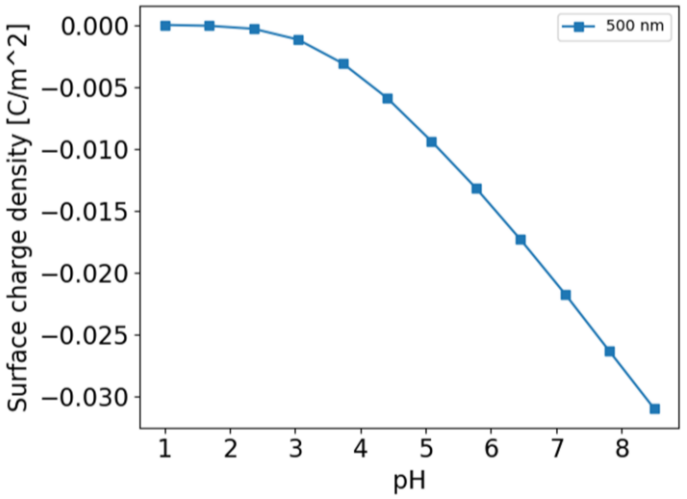


Figure 2: Surface charge density of nanoparticles with a diameter of 500 nm

The experiments have been carried out in 9 batches, two of those show agglomeration of silica and the rest shows breakage. The data is split into three datasets. One training dataset, which consists of one agglomeration batch and four breakage batches, while two batches (different from the training set) are left for validation. Moreover, in order to evaluate the model performance, the second agglomeration batch and one breakage batch are left in the test dataset. The model predictions for agglomeration and breakage batches in the test set are illustrated in Figure 3. The model is trained for 400 epochs and it can roughly predict the dynamics of the flocculation process for both agglomeration and breakage phenomena. The predictions can be improved by training the model for more than 400 epochs. To compare it with the models developed in (Nazemzadeh et al., 2021a; Nielsen et al., 2020) a further uncertainty analysis can be carried out similar to the one developed by (Nielsen et al., 2021) to analyse the performance of each model more in detail.

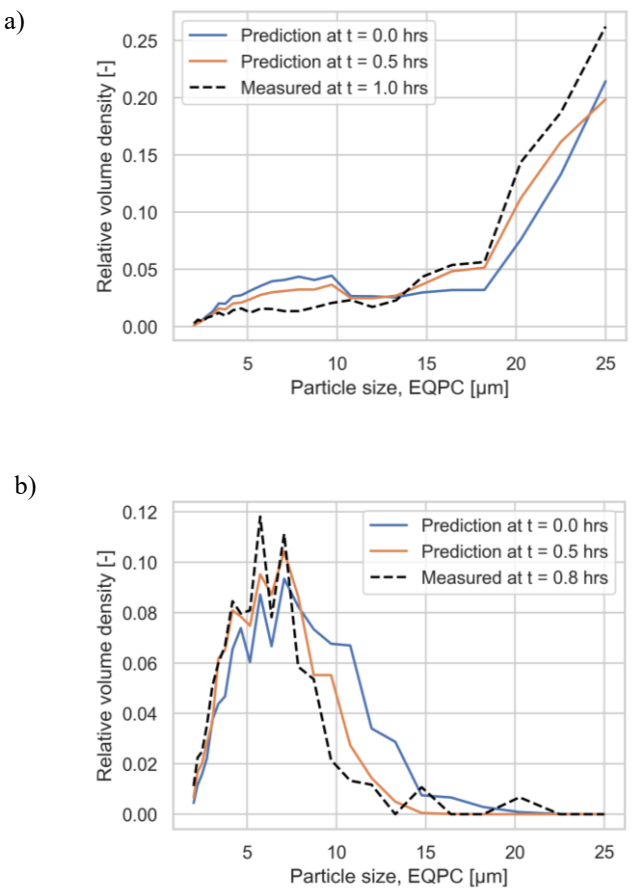


Figure 3: model predictions of a) agglomeration, and b) breakage batches

4. Conclusions

In this study, a hybrid multiscale framework has been developed to predict the dynamics of a laboratory-scale flocculation of silica particles. The hybrid multiscale model has the advantage of employing the data at non-observable scale to assisting the prediction of the dynamics of the process. The framework allows of using computational and theoretical chemistry models to predict surface properties of the colloidal silica particles based on the medium conditions. Hence, these properties are used in the hybrid model to estimate the kinetic parameters of the flocculation process. It is also planned to use this framework for predicting the dynamics of a polymeric flocculation process.

5. Acknowledgements

We would like to thank Novozymes A/S and Greater Copenhagen Food Innovation program (CPH-FOOD) for co-financing this research, and for their support during this project.

References

Andersson, M.P., Hassenkam, T., Matthiesen, J., Nikolajsen, L., Okhrimenko, D. V., Dobberschütz, S., Stipp, S.L.S., 2020. First-principles prediction of surface wetting. Langmuir. https://doi.org/10.1021/acs.langmuir.0c01241

Barisik, M., Atalay, S., Beskok, A., Qian, S., 2014. Size dependent surface charge properties of silica nanoparticles. J. Phys. Chem. C 118, 1836–1842. https://doi.org/10.1021/jp410536n

Kumar, S., Ramkrishna, D., 1996. On the solution of population balance equations by discretization-I. A fixed pivot technique. Chem. Eng. Sci. 51, 1311–1332.

Nazemzadeh, N., Malanca, A.A., Nielsen, R.F., Gernaey, K. V., Andersson, M.P., Mansouri, S.S., 2021a. Integration of first-principle models and machine learning in a modeling framework: An application to flocculation. Chem. Eng. Sci. 245, 116864. https://doi.org/10.1016/j.ces.2021.116864

Nazemzadeh, N., Nielsen, R.F., Gernaey, K. V., Mansouri, S.S., Andersson, M.P., 2021b. Implementation of first-principles surface interactions in a hybrid machine learning assisted modelling of flocculation, Computer Aided Chemical Engineering. Elsevier Masson SAS. https://doi.org/10.1016/B978-0-323-88506-5.50132-7

Nielsen, R.F., Nazemzadeh, N., Andersson, M.P., Gernaey, K. V., Mansouri, S.S., 2021. An uncertainty-aware hybrid modelling approach using probabilistic machine learning, Computer Aided Chemical Engineering. Elsevier Masson SAS. https://doi.org/10.1016/B978-0-323-88506-5.50094-2

Nielsen, R.F., Nazemzadeh, N., Sillesen, L.W., Andersson, M.P., Gernaey, K. V., Mansouri, S.S., 2020. Hybrid machine learning assisted modelling framework for particle processes. Comput. Chem. Eng. 140. https://doi.org/10.1016/j.compchemeng.2020.106916

Scheepers, D., Chatillon, B., Borneman, Z., Nijmeijer, K., 2021. Influence of charge density and ionic strength on diallyldimethylammonium chloride (DADMAC)-based polyelectrolyte multilayer membrane formation. J. Memb. Sci. 617. https://doi.org/10.1016/j.memsci.2020.118619

Yotsumoto, H., Yoon, R.H., 1993. Application of extended DLVO theory. II. Stability of silica suspensions. J. Colloid Interface Sci. https://doi.org/10.1006/jcis.1993.1206

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50136-6

Optimal Design of Experiments Based on Artificial Neural Network Classifiers for Fast Kinetic Model Recognition

Enrico Sangoi[a,b], Marco Quaglio[a], Fabrizio Bezzo[b], Federico Galvanin[a*]

[a]*Department of Chemical Engineering, University College London, Torrington Place, WC1E 7JE London, United Kingdom*
[b]*CAPE-lab (Computer-Aided Process Engineering Laboratory), Department of Industrial Engineering, University of Padova, 35131 Padova PD, Italy*
f.galvanin@ucl.ac.uk

Abstract

Developing mathematical models for the description of reaction kinetics is fundamental for process design, control and optimisation. The problem of model discrimination among a set of candidate models is not trivial, and recently a new and complementary approach based on artificial neural networks (ANNs) for kinetic model recognition was proposed. This paper extends the ANNs-based model identification approach by defining an optimal design of experiment procedure, whose performance is assessed through a simulated case study. The proposed design of experiments method allows to reduce the number of experiments to be conducted while increasing the ability of the artificial neural network in recognising the proper kinetic model structure.

Keywords: optimal design of experiments, kinetic model identification, ANN classifiers.

1. Introduction

The mathematical description of the phenomena occurring in reacting systems is crucial to evaluate the progress of chemical reactions occurring in the reactor, which is a central aspect in reactor design, control and optimisation. The first step in modelling reaction kinetics is the definition of candidate models based on preliminary experimental evidence and hypothesis about the underlying reaction mechanism.
The subsequent step in the modelling procedure is to determine among the candidate models the most suitable one for describing the chemical system, i.e. to perform the model discrimination. Asprey and Macchietto (2000) presented a standard model building approach including model discrimination, which can be summarised in the following steps: i) structural identiability analysis of each model; ii) regression of model parameters on experimental data; iii) evaluation of the quality of fit; iv) application of the chosen model selection criterion; v) if required, design and perform new experiments, then iteration from step (ii). When new experiments must be performed, experimental conditions are typically designed through a mathematical optimisation to maximise their information content. Different approaches can be employed for designing optimally informative experiments, such as model-based design of experiments (MBDoE) (Chen and Asprey, 2003) or probabilistic approaches based on Bayesian statistics or Monte Carlo methods. One limitation of standard MBDoE-driven model building strategies is that these methods may even lead to the rejection of the "exact" model if the identifiability

requirements are not satisfied in step (i) of the procedure. These reasons motivate the research for an alternative approach to the kinetic model discrimination and design of experiments problem. Quaglio et al. (2020) looked at the problem of selecting a kinetic model among a set of candidate ones from a completely different perspective, applying artificial neural network (ANN) classifiers for a fast recognition of the kinetic model structure of reacting systems from the experimental evidence provided. The idea behind the ANN-based approach is to train the ANN classifier (Zou et al., 2009) with in-silico generated data for each model from the set of rival kinetic models, so that the trained ANN would determine the kinetic model based on the input concentration measurements. A key aspect of this approach is that it does not require the fitting of kinetic parameters; parameter uncertainty is included during the dataset generation step. The method proved to be satisfactory, particularly when the noise on the measured values is low, and within this framework the modelling time is expected to decrease. However, the approach was characterised by a fixed design of the experiments, which can lead to poorly informative experiments.

The research here presented extends Quaglio et al. (2020) work by defining a procedure for optimal experimental design coupled to the ANN-based classifier for fast kinetic model recognition. Classical MBDoE approaches are not suitable for the direct integration of the ANN approach, because the kinetic models are not used for experimental data fitting and making predictions; therefore a new criterion for optimally designing experiments has been formulated. The new approach is tested on a simulated case study related to the identification of kinetic models in a batch reaction system.

2. Proposed framework and methods

Chemical reacting system are typically described by nonlinear dynamic models. In this work, the N_m possible models are formulated as systems of differential and algebraic equations in the form

$$\begin{cases} \mathbf{f}_l(\dot{\mathbf{x}}_l(t), \mathbf{x}_l(t), \mathbf{u}(t), \boldsymbol{\theta}_l, t) = 0 \\ \hat{\mathbf{y}}_l(t) = \mathbf{g}_l(\mathbf{x}_l(t)) \end{cases} \quad \forall\, l \in \{1, \dots, N_m\} \tag{1}$$

where $\mathbf{x}$ is the vector of N_x state variables, $\dot{\mathbf{x}}$ is the derivative of the state variables, the vector $\mathbf{u}$ represents the array of N_u input variables, $\boldsymbol{\theta}$ is the N_θ-dimensional vector of model parameters, t is the variable time, $\hat{\mathbf{y}}$ is N_y-dimensional vector of measured output variables that are function of the state variables, and subscript $l = 1, \dots, N_m$ stands for the l-th candidate model. In the ANN model recognition framework, the ANN is trained and then employed to determine among the N_m models the most suitable one for describing the reacting system behaviour.

The aim is to enhance the ANN ability in associating the right model label to the input data, which are affected by the choice of the experimental conditions. To optimise the experimental design, the objective function to be maximised is the ANN accuracy on the testing dataset is defined in Eq. 2 as the percentage of correctly classified model structures where Ψ_{test} is the testing dataset, the dataset elements are the experimental data array $\mathbf{n}_i$ and the respective model label l_i, while $\hat{l}_i$ is the label predicted by the ANN.

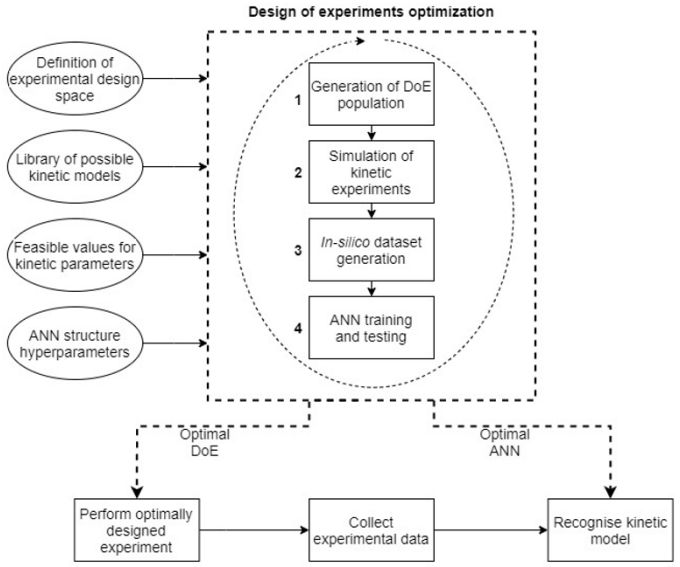


Figure 1 Block diagram representation of the proposed procedure for design of experiments optimisation coupled to ANN-based kinetic model recognition.

$$\mathrm{Acc}_{\Psi_{\mathrm{test}}} = \frac{\left|\left\{i \in \{1,\dots,N_{\Psi_{\mathrm{test}}}\} \,\mathrm{s.t.}\, (\mathbf{n}_i, l_i) \in \Psi_{\mathrm{test}} \wedge \hat{l}_i = l_i\right\}\right|}{\left|\left\{i \in \{1,\dots,N_{\Psi_{\mathrm{test}}}\} \,\mathrm{s.t.}\, (\mathbf{n}_i, l_i) \in \Psi_{\mathrm{test}}\right\}\right|} \cdot 100\% \tag{2}$$

The proposed procedure is illustrated in Figure 1. The procedure starts with the definition of all the inputs that must be fed to the optimiser, which are:

- The library of the rival kinetic model structures.
- A list of feasible values of the kinetic parameters for each candidate model. Parameters values are bounded and are said feasible if they satisfy constraints on the conversion and selectivity of chemical species when simulating the experiments.
- The number of experiments to be optimised and the space of experimental conditions, i.e. which experimental variables can be manipulated and their range of variability.
- The ANN structure: the number of layers, the number of nodes at each layer, the activation functions and all the other ANN hyperparameters.

Afterwards, the optimisation of the experimental design is performed, coupled to the training of ANNs. The discussion about the DoE optimisation step, represented in Figure 1 by the dashed box, addresses two aspects.

i) How the objective function is evaluated for a certain experimental design.

The objective function has to be evaluated many times for each experimental design proposed within the DoE optimiser iterations. The procedure to assess the ANN accuracy for a particular DoE choice involves the following sequential steps:

1. An experimental design is chosen within the experimental design space.
2. The respective experimental campaign is simulated in-silico for each model and for each set of feasible parameter values.

3. Simulated measurements of species concentration $\mathbf{n}_i$ are collected and organised to build a labelled dataset Ψ, where each element of the dataset is associated to the kinetic model used for its generation l_i. An error factor is added to the simulated measurements to mimick real experimental circumstances.
4. The ANN is trained and tested in the classification of the kinetic model structures based on the experimental evidence provided. The ANN accuracy (Eq. 2) is evaluated at this step.

ii) The choice of the algorithm for optimising the DoE.

The accuracy of ANN predictions on the test set is the metric used to evaluate the goodness of the experimental design: the higher the test-accuracy, the better the DoE. Since this metric is affected by intrinsic randomness in the neural network training process, a direct search method has been employed to deal with the optimisation problem. In particular, the differential evolution algorithm (Storn and Price, 1997) is chosen, a population-based algorithm inspired by the evolutionary theory. Once the algorithm converges to a solution, it provides as an output the optimal DoE and the respective ANN, i.e. the most accurate one at recognising the right kinetic model structure. In the final step of the proposed procedure, the optimally determined conditions are applied to the real system. Experimental data are collected and fed to the trained ANN, obtaining as output a measure of how likely each model is to describe the chemical system under investigation.

3. Case study description

The proposed methodology was tested on an in-silico simulated case study, considering an isothermal batch reactor where 3 species (A, B, C) react in the liquid phase.

$$A \xrightarrow{r_1} B\,; \qquad A \xrightarrow{r_2} C\,; \qquad B \xrightarrow{r_3} C \tag{2}$$

The set of equations describing the evolution of the system, assuming constant volume for the liquid mixture in the batch reactor, is given by Eq. 3, where C_i (mol/m^3) indicate the concentration of species i, namely the state variables of the system, r_j (mol/m^{3}s) is the rate of reaction j, and ν_{ij} (-) represents the stoichiometric coefficient of species i in the reaction j.

$$\frac{dC_i}{dt} = \sum_{j=1}^{N_r} \nu_{ij} r_j \quad \forall\, i = A, B, C \tag{3}$$

$$k_j = A_j e^{-E_{a,j}/RT} \qquad \forall j = 1,2,3 \tag{4}$$

Kinetic models are formulated assuming that reactions can occur either in series or parallel, and reactions rates were modelled with first or second order power-law expressions. Moreover, Arrhenius-type kinetic factors are assumed (Eq. 4) for evaluating the reaction rates, characterised by two parameters: pre-exponential factor A_j and activation energy $E_{a,j}$. Therefore, a total of 8 rival kinetic model structures are formulated, as summarised in Table 1.

Table 1 Power-law rate expression for the candidate kinetic models. Models with label l=1,…,4 consider a series mechanism, while models l=5,...,8 consider a parallel mechanism.

	Series				Parallel			
Label:	1	2	3	4	5	6	7	8
$\boldsymbol{r_1}$	$k_1 \cdot C_A$	$k_1 \cdot C_A$	$k_1 \cdot C_A^2$	$k_1 \cdot C_A^2$	$k_1 \cdot C_A$	$k_1 \cdot C_A$	$k_1 \cdot C_A^2$	$k_1 \cdot C_A^2$
$\boldsymbol{r_2}$	$k_2 \cdot 0$	$k_2 \cdot 0$	$k_2 \cdot 0$	$k_2 \cdot 0$	$k_2 \cdot C_A$	$k_2 \cdot C_A^2$	$k_2 \cdot C_A$	$k_2 \cdot C_A^2$
$\boldsymbol{r_3}$	$k_3 \cdot C_B$	$k_3 \cdot C_B^2$	$k_3 \cdot C_B$	$k_3 \cdot C_B^2$	$k_3 \cdot 0$	$k_3 \cdot 0$	$k_3 \cdot 0$	$k_3 \cdot 0$

The dataset Ψ is generated starting from 125 sets of feasible parameters for each model structure, therefore Ψ includes 1000 elements $(\mathbf{n}_i, l_i)$. The dataset is split into training-validation-testing sets on a 60-20-20 basis for ANN training, validation and testing.
In this case study, the experimental error is generated from a normal distribution with zero mean and constant variance σ^2, assuming three different noise levels: σ=1 mol/m^3, σ=4 mol/m^3, σ=10.00 mol/m^3.
The experimental variables manipulated when designing the experiments, and their upper and lower bounds, are: i) initial concentration of reactant A (0-250 mol/m^3), ii) sampling times (50-350 s), and iii) temperature in the isothermal reactor (520-720 K).
As underlined in section 2 of this paper, also the hyperparameters defining the ANN architecture must be provided as an input. Quaglio et al. (2020) work is used as the starting point to retrieve the ANN hyperparameters for this case study:

- one hidden layer with 100 nodes and rectified linear unit activation function,
- 8 nodes in the output layer characterised by softmax activation function,
- the optimiser is adaptive moment estimation (Adam), (Kingma and Ba, 2014).

Before employing the optimisation algorithm on the case study, it is worth conducting some preliminary study to examine whether the experimental design affects the ANN accuracy and to identify potential regions of the design space where to expect optimal experimental conditions. A sensitivity analysis has been carried out to this purpose.
The optimisation is conducted considering three different scenarios of simulated experimental noise on concentration measurements and for each scenario the DoE optimisation is repeated with i) different number of experiments; ii) different number of samples, to determine the minimum amount of experimental effort required from the ANN to reach the highest accuracy.

4. Results

As far as the ANN test-accuracy is concern, the results are extremely promising, as shown in Figure 2a. These results refer to the DoE optimisation of experiments characterised by just one sample per batch, which indeed is even more remarkable. Very high accuracy in the ANN predictions is achieved even in the case of extremely noisy measurements with a limited number of optimally designed experiments. In particular, with four optimally designed batches a 100 % accuracy is reached in the low noise scenario, while in the medium noise case a 95 % accuracy is reached. In the scenario characterised by high noise, still, 90 % accuracy is achieved with just five optimally designed experiments. Figure 2b shows the sensitivity of ANN accuracy to temperature and initial concentration of species A in the high noise scenario, compared to the optimisation result for a single experiment with a single sample collected after t=100 s. Figure 2b suggests that optimal conditions for model discrimination are characterised by high initial concentration of reactant A ($C_{A,0}$) and a medium range temperature around 630K.

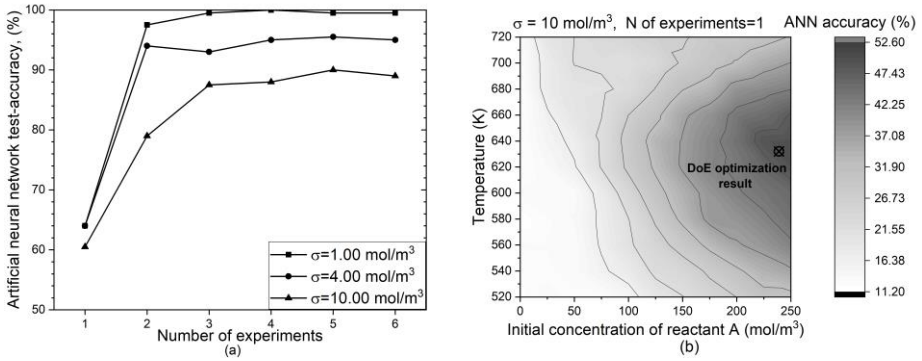


Figure 2 (a) ANN accuracy obtained from the optimal design of experiments for variable number of experiments: comparison between the three experimental noise scenarios. (b) Contour plot of the ANN accuracy as a function of design variables for σ =10 mol/m^3, fixed sampling time t = 100 s.

The high value obtained for $C_{A,0}$ can be due to the hypothesis of constant σ. Under this assumption, the relative error is lower when the concentration measurement is higher. In the scenario described for Figure 2b, the CPU time required for the DoE optimisation was 1376 s (approx. 23 minutes), carried out on an Intel Core i5-7200U processor with 8GB RAM workstation.

5. Conclusions

This study allowed to define a new method for optimal design of experiments that can be coupled with the ANN-based methodology for kinetic model recognition. Optimally designed experiments lead to a significant improvement in terms of ANN accuracy, thus paving the way towards ANN-based DoE methods for the recognition of kinetic models in complex reaction networks. Preliminary results show that kinetic model structures can be identified with limited experimental effort also in the presence of high noise level on the measurements. Future works will include the application to real chemical systems to test the reliability of the methods and the comparison in terms of time required and experimental campaign costs with classical model discrimination approaches.

References

S.P. Asprey and S. Macchietto, 2000, Statistical Tools for Optimal Dynamic Model Building, Computers & Chemical Engineering, 24, 1261-1267

B.H. Chen and S.P. Asprey, 2003, On the Design of Optimally Informative Dynamic Experiments for Model Discrimination in Multiresponse Nonlinear Situations, Industrial & Engineering Chemistry Research, 42, 7, 1379-1390

D.P. Kingma and J. Ba, 2014, Adam: A Method for Stochastic Optimization

M. Quaglio, L. Roberts, M.S. Bin Japaar, E.S. Fraga, V. Dua, F. Galvanin, 2020, An Artificial Neural Network Approach to Recognise Kinetic Models from Experimental Data, Computers & Chemical Engineering, 135, 106759

R. Storn and K. Price, 1997, Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces, Journal of Global Optimization, 11, 4, 341-359

J. Zou, Y. Han and SS. So, 2009, Overview of Artificial Neural Networks, Artificial Neural Networks: Methods and Applications, ed. David J. Livingstone, Totowa NJ, Humana Press, 15-23

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50137-8

Modelling of the rice bran protein extraction using response surface methodology

Mohd Sharizan Md Sarip[a*], Felicia Kang Suet Lyna[a], Mohd Rizuan Mansor[a], Ken-Ichiro Sotowa[b]

[a]*Faculty of Chemical Engineering Technology, Universiti Malaysia Perlis, Kompleks Pusat Pengajian Jejawi 3, Universiti Malaysia Perlis (UniMAP), 02600 Arau, Perlis , Malaysia*
[b]*Department of Chemical Engineering, Kyoto University, Kyotodaigau-Katsura, Nishikyo-ku, Kyoto-shi Kyoto 615-8510,Japan.*

sharizan@unimap.edu.my

Abstract

Rice bran is a natural by-product that is produced by the rice milling process in the rice industry. In this study, the modelling of the protein extraction from rice bran using Soni-auto hybrid method was evaluate via response surface methodology (RSM). Soni-auto hybrid method are the combination of the autoclaving extraction followed by sonication process. The experimental design for this process was performed using Response Surface Methodology (RSM) with Central Composite Design (CCD) model in Design Expert software. Three variables parameters namely solvent to sample ratio from 1:20 to 1:40, sonication temperature from 40 to 60ºC and 20 to 40 minutes was evaluated. The response variable in this study are protein concentration. A highly accurate empirical quadratic response model for protein concentration was developed for three independent variable with adjusted R-squared value for the model was 0.8933. Two parameters; temperature and feed-to-solvent ratio give a significant effect ($p<0.05$) on protein concentration. Meanwhile, extraction times are not given the significant on this response. The model will be used to obtain the optimum condition for the process.

Keywords: Modelling, rice bran protein, response surface methodology.

1. Introduction

In the rice industry, the large amount of rice bran produced through rice milling process is often considered as waste. These rice brans are either being thrown away or sell at a very cheap price as animal feed or fertilizer (Gul et al., 2015). The under-utilization of rice bran is due to the high amount of free fatty acid in the rice bran due to the reaction of the lipase enzyme in the rice bran. Patil et al. (2016) stated that within hours large amount of free fatty acid can be produced in the rice bran and the amount can reach 5 to 7% within the first 24 hours of storage. Nutritionists have considered rice bran with more than 5% of free fatty acid are unsuitable for human consumption (Patil et al., 2016). Moreover, rice bran contains fiber and phytate which are extensively associated or bounded to the proteins makes the separation of protein from these components difficult to achieve. Therefore, various extraction techniques such as alkaline extraction, enzymatic extraction and physical extraction were used to extract the protein from the rice bran fibers and at the same time inhibit the lipase activity in the rice bran (Bhat and Riar, 2017). In this study, Soni-auto hybrid method was utilized for the protein extraction from the rice bran. Three main extraction parameter namely sonicator temperature,

extraction time dan solvent to sample ratio was studied as the independent variable. The modelling for this process was done through response surface methodology using central composite design (CCD) model to obtain the best multivariable equation to represent the proses. Besides that, the significant effect of this parameter toward protein concentration was determine based on analysis of variance (ANOVA) analysis.

2. Methodology

2.1 Raw material

Rice bran was collected from Kilang Beras BERNAS Sdn Bhd, Perlis, Malaysia. It was heated in a microwave oven at 850 W for 3 minutes and stored at 4 to 5ºC prior to each experimental run (Pandey and Shrivastava, 2018).

2.2 Soni-auto hybrid extraction method

Soni-auto hybrid method are the combination of the autoclaving extraction followed by sonication process. 5 g of rice bran was added into a conical flask and followed by 100 ml of distilled water for feed to solvent ratio of 1:20 g:ml. Then, the sample mixture was autoclaved at 121˚C for 20 minutes. After that, the autoclaved rice bran solution was left to cool down to room temperature. Finally, the mixture of rice bran was sonicated by using the sonicator bath (Model: WUC-D22H, Germany). Throughout the sonication process, the ultrasonic frequency was maintained at 70% with temperature range between 40˚C to 60ºC and extraction time between 20 to 40 minutes After sonication, the rice bran mixture was centrifuged at 4000 rpm for 30 minutes at 25˚C. Then, volume of the supernatant was measured. The absorbance of supernatant from each extraction method was measured using the UV-Vis spectrophotometer (Shimadzu, Model: UV-1800, Japan) to determine protein concentration in the rice bran solution. The procedures were repeated three times for all the parameters.

2.3 Response surface methodology (RSM)

In this study, response surface methodology (RSM) with Central Composite Design (CCD) model was utilized to model the extraction of rice bran protein using Soni-auto hybrid. The target response that needed to achieve was the maximum level of protein concentration in the rice bran solution. Therefore, the influence of the three parameters; time, temperature and feed-to-solvent ratio on the protein concentration was investigated. The factors and response as defined in CCD model is shown in Table 1. Design Expert V10.0.7 software was used as statistical tool for this purpose.

Table 1: List of numeric factors and response defined in Central Composite Design (CCD) model

Numeric Factors	**Name**	**Units**	**Minimum**	**Maximum**
A	Temperature	˚C	40	60
B	Extraction time	Minutes	20	40
C	Feed : Solvent Ratio	g:ml	1:20	1: 60
Response				
R1	Protein Concentration	mg/ml		

3. Results and discussion

Based on the response surface methodology (RSM) with Central Composite Design (CCD) model, 20 runs of experiment was done with different operating conditions was evaluated as shown in Table 2. The condition was generated through design of experiment (DOE) procedure using Design Expert V10.0.7 software.

Table 2: Experimental data based on CCD model for protein extraction from rice bran

Run	A Temperature (˚C)	B Extraction Time (minutes)	C Feed: Solvent Ratio (g:ml)	R1 Protein Concentration (mg/ml)
1	50	30	40	17.588
2	56	36	52	14.014
3	50	30	40	17.418
4	50	30	40	17.502
5	50	30	40	17.287
6	44	36	52	14.881
7	44	24	52	14.894
8	56	24	28	15.898
9	50	30	60	13.129
10	50	40	40	16.876
11	56	24	52	13.978
12	50	30	40	17.012
13	40	30	40	15.218
14	50	30	20	14.375
15	56	36	28	15.796
16	50	20	40	16.614
17	50	30	40	16.998
18	60	30	40	15.008
19	44	36	28	16.598
20	44	24	28	16.499

3.1 Model selection

Table 3 shows the model summary statistics for protein concentration, R1. In this section, predicted residual sum of squares (PRESS) value was taken into consideration for the selection of model. PRESS is a measure of how a particular model fits each design point. The coefficients for the model were calculated without the first design point. This model was used to predict the first point and then the new residual was calculated for this point. This was done for each data point and then the squared residuals were summed. In this case, the quadratic polynomial model has the lowest predicted residual sum of squares (PRESS) at 11.42 compared to all other models which are not aliased. In addition, adjusted R^2 values are also used in determining the type of polynomial that is suitable to represent the model. From table 3, other than the cubic model that was aliased, the quadratic polynomial model was observed to have the highest adjusted R^2 value of 0.9046 compared to the linear and 2FI models that have adjusted R^2 value of only 0.0510 and -

0.1671 respectively. Therefore, from the model summary statistics, the quadratic polynomial model with high adjusted R^2 and low PRESS values was chosen to represent the model of study. The predicted R-square value indicates how well a regression model predicts response values; while the adjusted R-square (Adj. R^2) indicates the descriptive power of the regression models while including the diverse numbers of variables (Che Sulaiman et al., 2017).

Table 3 : Model summary statistics for protein concentration

Source	Standard Deviation	R-Squared	Adjusted R-Squared	Predicted R-Squared	PRESS
Linear	1.32	0.2009	0.0510	-0.2031	42.06
2FI	1.47	0.2015	-0.1671	-0.7254	60.32
Quadratic	**0.42**	**0.9498**	**0.9046**	**0.6734**	**11.42**
Cubic	0.29	0.9857	0.9546	-0.2082	42.24

Based on the Table 3, the difference between adjusted R-Squared and predicted R-Squared of 0.2312 are high for quadratic model. The differences of this value more than 0.2 are not considered to be a reasonable agreement (Tesfay et al.,2020). Therefore, the analysis of variance (ANOVA) was performed for the quadratic model as tabulated in Table 4. In general, values of "Prob>F" less than 0.0500 indicate the model terms are significant based on 95 % confidence level. In this case, A, C, A^2 and C^2 are significant model as shown in Table 4. If there are many insignificant model terms, model reduction can be done to improve the model and produce the reasonable agreement model.

Table 4 : Analysis of variance (ANOVA) results for quadratic model

Source	**F Value**	**p-value Prob>F**	**Observation**
Model	21.02	<0.0001	Significant
A- Temperature	5.22	0.0453	Significant
B- Extraction Time	0.088	0.7722	Insignificant
C- Feed: Solvent	34.69	0.0002	Significant
AB	0.016	0.9005	Insignificant
AC	0.10	0.7551	Insignificant
BC	4.813E-4	0.9829	Insignificant
A^2	42.80	<0.0001	Significant
B^2	1.73	0.2180	Insignificant
C^2	118.85	<0.0001	Significant

Based on Table 4, it was found out that B^2, BC, AC and AB are in insignificant but only B^2 model term is removed to improve the model. This is because BC, AC and AB terms are required to support the polynomial equation hierarchy in data fitting based on trial-and-error model reduction process. The results for reduced quadratic model are shown in table 5.

Table 5 : Analysis of variance (ANOVA) results for reduced quadratic model

Source	Sum of Squares	Df*	Mean Square	F Value	p-value Prob>F
Model	32.90	8	4.11	21.97	<0.0001
A- Temperature	0.92	1	0.92	4.90	0.0489
B- Extraction Time	0.016	1	0.016	0.083	0.7786
C- Feed : Solvent	6.09	1	6.09	32.53	0.0001
AB	2.888E-3	1	2.888E-3	0.015	0.9034
AC	0.018	1	0.018	0.096	0.7620
BC	8.450E-5	1	8.450E-5	4.514E-4	0.9834
A^2	7.29	1	7.29	38.94	<0.0001
C^2	20.57	1	20.57	109.90	<0.0001
Residual	2.06	11	0.19		
Standard Deviation	0.43		R-Squared		0.9411
Mean	15.88		Adjusted R-Squared		0.8983
C.V. %	2.72		Predicted R-Squared		0.7049
PRESS	10.32		Adeq Precision		15.651
-2 Log Likelihood	11.29		BIC		38.25

∗degree of freedom

The quadratic model with f value (0, 8) = 21.97 and p <0.0001 implies the model has significant effect on the response. It also indicates that only 0.01% chance that the F-value occurs due to noise. Therefore, the overall model has a significant effect in determining the protein concentration within the experimental ranges.

Based on this result, the feed-to-solvent ratio and temperature had significant effect on protein concentration with f value (1,8) = 32.52, p = 0.0001 and f value (1,8) = 4.90, p = 0.0489, respectively. In contrast, the extraction time for sonication do not exhibit significant effects on the protein concentration as it has a low f value (1,8) = 0.083 with p = 0.7786.

As a conclusion, a final equation of the quadratic polynomial model was established to represent the process as shown in eq. 1. The adjusted R-squared value for the model was 0.8933 while predicted R-squared value was 0.7049. The differences of this value are less than 0.2 and it considered to be a reasonable agreement.

$$R1 = 17.17 - 0.26\mathrm{A} + 0.034\mathrm{B} - 0.67\mathrm{C} - 0.019\mathrm{AB} - 0.047\,\mathrm{AC} + 0.00325\mathrm{BC} - 0.71A^2 - 1.19C^2 \quad (1)$$

Where, R1 = Protein Concentration (mg/ml)
A = Temperature
B = Extraction time
C = Feed: Solvent ratio

Based on the equation, the validation of model is conducted to evaluate the uncertainty of this model by comparing the experiment data with predicted value. The experiment was conducted at temperature of 50˚C, feed to solvent ratio set at 1:40 and extraction time of 30 minutes. The analysis is shown in table 6.

Table 6 : Model validation with experimental data.

Response (R1)	Experimental data		Predicted value	Error
	Average (mg/ml)	AAD[1] (mg/ml)	(mg/ml)	AARD[2] (%)
Protein concentration	16.995	0.137	17.174	1.053

[1] AAD - Absolute average deviation = $\frac{1}{n}\sum_{m=1}^{n}\left|Y_{m-}Y_{average}\right|$, where n= 3, no of experiment repeated.

[2] AARD - Absolute average relative deviation = $\left(\frac{Y_{experiment-}Y_{predicted}}{Y_{experiment}}\right)x\ 100$.

4. Conclusions

As a conclusion, the quadratic model was successfully being developed for protein extraction process from rice bran using Soni-auto hybrid extraction method with adjusted R-squared value of 0.8933 and predicted R-squared value of 0.7049. The model has the f value (0, 8) = 21.97 and p <0.0001 which implies it has significant effect on the response. Two parameters namely feed-to-solvent ratio and temperature are observed to give the significant effect on protein concentration with f value (1,8) = 32.52, p = 0.0001 and f value (1,8) = 4.90, p = 0.0489, respectively. In contrast, the extraction time for sonication not given the significant effects on the protein concentration with f value (1,8) = 0.083 with p = 0.7786. The empirical model develop in this study will be used to obtain optimum condition.

Acknowledgement

The financial support provided by the Ministry of Higher Education Malaysia through Fundamental Research Grant Scheme reference no: FRGS/1/2018/TK02/UNIMAP/02/3 was acknowledged.

References

K.Gul, B.Yousuf, A.K.Singh, P.Singh and A.A.Wani, 2015, Rice bran: Nutritional values and its emerging potential for development of functional food—A review. Bioact. Carbohydr. Diet. Fibre, 6,1, 24-30.

S.S. Patil, A.Kar, and D. Mohapatra, 2016, Stabilization of rice bran using microwave: Process optimization and storage studies, Food Bioprod. Process, 99, 204-211.

F.M. Bhat and C.S. Riar, 2017, Extraction, identification and assessment of antioxidative compounds of bran extracts of traditional rice cultivars: An analytical approach, Food Chem., 237, 264-274.

I.S. Che Sulaiman, M.Basri, H.R. Fard Masoumi, W.J. Chee, S.E. Ashari and M. Ismail, 2017, Effects of temperature, time, and solvent ratio on the extraction of phenolic compounds and the anti-radical activity of Clinacanthus nutans Lindau leaves by response surface methodology, Chem. Cent. J., 11, 1, 54-54.

D.Tesfay, S.Abrha, Z.Yilma, G.Woldu, & F.Molla, 2020, Preparation, Optimization, and Evaluation of Epichlorohydrin Cross-Linked Enset (Ensete ventricosum (Welw.) Cheeseman) Starch as Drug Release Sustaining Excipient in Microsphere Formulation, Biomed Res. Int., 2020, 1-19.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50138-X

Weibull Reliability Regression Model for Prediction of Bearing Remaining Useful Life

Jun-Jie Lai, Shih-Jie Pan, Chong-Wei Ong, Cheng-Liang Chen*
Department of Chemical engineering, National Taiwan University, No. 1, Section 4, Roosevelt Rd, Da'an District, Taipei City, 106319, Taiwan
**CCL@ntu.edu.tw*

Abstract

In this work, a scenario-based approach that uses multiple Weibull Accelerated Failure Time Regression (WAFTR) models is proposed to predict the remaining useful life (RUL) of a benchmark bearing. The external features such as operational load and rotatory speed of the bearing are used to categorize the operational scenarios and a scenario-based WAFTR model is identified for each operational scenario by using the internal features extracted from the sampled horizontal and vertical vibration data. Therein, the Weibull parameters in each WAFTR model are expressed either in exponential or in linear form of these internal and external operational features. By using the mean squared error (MSE) as the performance measure of the prediction model, it is found that the proposed multiple WAFTR models approach can predict the RUL within 20% error.

Keywords: Bearing, Weibull distribution, Remaining Useful Life (RUL), Reliability function

1. Background

Rotation equipment such as motors, compressors are largely used in chemical plants. How to avoid those equipment damage is an importance task for field engineers. Engineers would check some indicators such as the temperature change or vibration amplitude to identify the health condition of the machine. Remaining useful life (RUL) is an estimate of the time that an item or component can function normally. Normally, the health condition of a machine will be worse with time goes by. Thus, maintenance and replacement are needed. In a plant, although earlier maintenance can prevent machines from broken, the more unnecessary cost has to pay. However, if the maintenance is too late, the machine may be broken and cause casualties. The accurate RUL prediction provides a suitable moment for replacement actions and keeps a safe environment for workers.

In 2012, the Institution of Electric, Electrical, and Engineering (IEEE) launched a challenge to the world (Patrick Nectoux et al (2012)). IEEE provided 17 bearing vibration datasets: six of them are the start-to-failure datasets, of which the whole vibration condition is from beginning to failure. The rest of them is the testing dataset, of which the vibration condition is not given fully. The purpose of this challenge is to let the engineers around the world predict the remaining useful life (RUL) for the bearing. There are two sensors detected the vibration acceleration connected with the bearing in the horizontal and vertical directions. These two sensors would take action every 10 seconds. Each time they would collect 2,560 data points in 0.1 seconds. Moreover, this experiment provided three different operating conditions by controlling the rotation speed and the force

conducted on the bearing. Although the experimental data provided includes the horizontal and vertical data, most of the researchers in this field would only use the horizontal data, such as Mao at al. (2018) and Guo et al. (2017), to name a few, the horizontal and vertical data together will affect the prediction accuracy. However, it is believed that the vertical data can still give us important information despite its fluctuation. In this work, both the horizontal and vertical data is used to build an RUL prediction model to enhance the prediction accuracy.

2. Methodology

2.1. Weibull reliability model

Kundu et al. (2019) proposed using the Weibull distribution function to describe the machine failure probability density with time, such as depicted in Eq. (1). The Weilbull reliability function can be derived from the Weilbull distribution function, as shown in equation (2).

$$f(t) = \frac{\beta}{\eta}\left(\frac{t-\gamma}{\eta}\right)^{\beta-1} e^{-(\frac{t-\gamma}{\eta})^{\beta}} \tag{1}$$

$$R(t) = 1 - \int_0^t f(s)ds = \int_t^{\infty} f(s)ds = e^{-(\frac{t}{\eta})^{\beta}} \tag{2}$$

Different types of machine failure can be described by tuning the three parameters in the Weilbull reliability function. Among these three tuning parameters, β is the shape parameter that would determine the type of failure, η is the scale parameter that would determine that machine's life span, and γ is the location parameter that determines at what time the machine starts to fail. It is assumed that the machine starts to break at the beginning, which means that γ can be set as zero. In a normal degradation process, the magnitude of β should greater than 1, and the magnitude of η should be similar to the total life span. When the operating time is closed to the failure time, the reliability of the machine is deemed to be closed to zero. It is assumed that the reliability is 0.01 at the failure time. Eq. (3) shows the relationship of the present time, t, and the RUL.

$$t + \mathrm{RUL}(t) = t_{\mathrm{failure}} \tag{3}$$

By plunging Eq. (3) into Eq. (2), the main prediction model is derived in Eq. (4).

$$\mathrm{RUL}(t) = \eta \times [-\ln(0.01)]^{\frac{1}{\beta}} - t \tag{4}$$

Time *t* is the primary factor that affects the reliability. However, there are some other aging factors contributing to the decreasing reliability. Some parameters can be controlled by the operators, like the applied load, and the rotary speed. The other internal aging parameters are features revealed by the machine like temperature, vibration magnitude. In Kundu's work, the two parameters β and η can be replaced by the bearing's features, where the η is expressed as an exponential expansion such as Eq. (5) shown.

$$\eta = \exp\left(\mathrm{a}_0 + \sum_{n=1}^{N} a_n I_n + \sum_{m=1}^{M} b_m E_m\right) \tag{5}$$

I_n and E_m are the internal and external parameters, respectively, and a_n and b_m are the coefficients of the aging parameters. In this work, β are also a parameter that would change as time goes by. As the result, β can also be expressed as an exponential expansion as shown in Eq. (6). Furthermore, the performance of using linear expansion is investigated for both β and η as shown in Eq. (7) and Eq. (8).

$$\beta = \exp\left(\mathrm{c}_0 + \sum_{n=1}^{N} c_n I_n + \sum_{m=1}^{M} d_m E_m\right) \tag{6}$$

$$\eta = \left(\mathrm{a}_0 + \sum_{n=1}^{N} a_n I_n + \sum_{m=1}^{M} b_m E_m\right) \tag{7}$$

$$\beta = \left(\mathrm{c}_0 + \sum_{n=1}^{N} c_n I_n + \sum_{m=1}^{M} d_m E_m\right) \tag{8}$$

2.2. Features

It is found that only horizontal dataset is used for model building in most of the previous studies. However, it is believed that the vertical datasets still contain important information, so both the horizontal and vertical datasets are took into concern. Moreover, the net vibration magnitude is also an important indicator for the RUL prediction. Thus, the "Scalar dataset" is calculated by Eq. (9):

$$\mathrm{Scalar}(t) = \sqrt{\mathrm{Horizontal}(t)^2 + \mathrm{Vertical}(t)^2} \tag{9}$$

By now, there are three types of datasets: the horizontal, the vertical, and the scalar. Because all of these contain much noise, the frequency-domain features are used as our main inputs for RUL prediction.

The horizontal dataset 1_1 is illustrated as shown in Figure 1. The Discrete Fourier Transform (DFT) is conducted to three types of raw data. Next, the frequency-domain features are normalized to highlight the magnitude ratio of frequency-domain features at each observation. This normalized value is also called the frequency signal energy. Instead of using the specific frequency as our model feature, we prefer to use a zone of the frequency. Thus, the whole frequency span is divided into 16 zones, (0 Hz – 800 Hz), (800 Hz – 1,600 Hz), (1,600 Hz – 2,400 Hz)… (12,000 Hz – 12,800 Hz). The principal component analysis (PCA) is conducted to these 16 frequency-domain features. Three principal components are extracted as the internal parameters. Furthermore, because using the frequency-domain features is possible to lose information, one time-domain internal parameter, the standard deviation, is picked. The standard deviation has high monotonicity. It would be seen as a health indicator for prediction.

$$\mathrm{Std} = \sqrt{\sum_{n=1}^{N} \frac{(x-\mu)^2}{N}} \tag{10}$$

Combined with the standard deviation, 12 features are extracted form three types of raw data (the horizontal, vertical, scalar data) in each dataset. These 12 internal parameters

and the 2 external parameters (force load, rotatory speed) will be plugged into the prediction model.

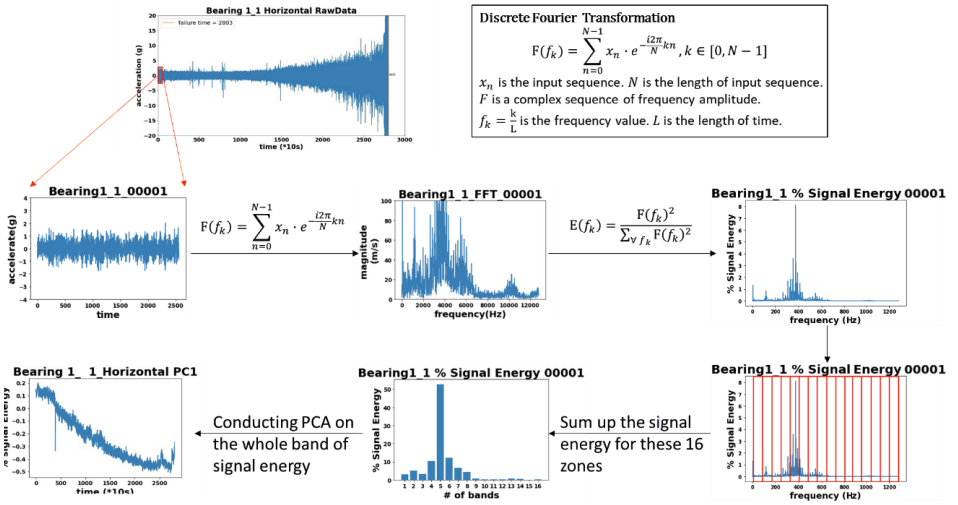


Figure 1 Frequency-domain features extraction with DFT and PCA

2.3. Target

The target of the model, in our case, is accurate prediction of the actual remaining useful life (RUL). The target will be used either in the regression part or in the testing part. It is assumed that the RUL is decreased linearly to zero at the failure time in our case. One can easily use the formula to get our RUL since the IEEE has given all the failure times.

3. Result and Discussion

3.1. Regression pre-process

Before building the prediction model, it is necessary to take a look at the raw data. The data provided by IEEE 2012 challenge is the real experimental data, so some datasets may include too much noise in building the prediction model. The datasets 1_2, 2_3, 2_5, 2_7, and 3_2 are excluded because they contain large fluctuation in the middle of the process. Also, the mean squared error (MSE) of the life percentage as shown in Eq. (11) is introduced. The value of MSE can show the precision of the prediction.

$$\mathrm{MSE} = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{\mathrm{RUL}_{\mathrm{prediction}}(n)}{\text{Total Life}} - \frac{\mathrm{RUL}_{\mathrm{actual}}(n)}{\text{Total Life}}\right)^2 \tag{11}$$

In this way, regardless of any case, the MSE will give the same value, which is 0.04, when the prediction RUL is just at the tolerance boundary. After defining the MSE, one can just use the MSE to describe the accuracy of the prediction model.

3.2. Model selection

In section *2.1.*, two types of expansion of β and η are combined in the model. The adequacy of these models are evaluated by checking the MSE values. The whole datasets are used to regress the model coefficients. All the simulations are done in the Python environment. The Python 3.9.4 script has the package to solve the non-linear regression.

It is found that the model with exponential β and linear η is the best because it brings the smallest MSE value.

During the prediction model process, it is critical to determine the training sets and testing sets. As there are three operating conditions, it is reasonable to pick at least one dataset from each operating condition. Also, the 1_1, 2_1, 2_2, and 3_1 are the start-to-failure datasets, which means that those datasets should have more information than the others. However, the prediction result is not good enough as shown in Table 1. The MSE values of 1_4, 1_5, and 1_7 are larger than 0.04, which means that the predicted RUL is out of the tolerance zone.

Table 1 The MSE testing result by prediction model trained with 1_1, 2_1, 2_2, and 3_1

Testing set	1_3	1_4	1_5	1_6	1_7	2_4	2_6	3_3
MSE	0.032	0.280	1.192	0.021	0.045	0.010	0.003	0.035

It is suspected that regressing the model with different operating conditions would decrease the accuracy. Moreover, the datasets from the same operating condition may not consistent enough. As the result, an analysis in done to check the relationship between each dataset. In this analysis, each dataset itself is used as the regression dataset to build the local model. The MSE value can be used in this analysis since it can represent how similar the datasets are. By test the other datasets with the local model, we can build an MSE map as shown in Figure 2. The blank parts mean that the testing set is the same as the training set. The dot parts mean that the MSE value is lower than 0.04. The other gray parts mean the MSE value is larger than 0.04.

Training \ Testing	1_1	1_3	1_4	1_5	1_6	1_7	2_1	2_2	2_4	2_6	3_1	3_3
1_1	8.01E-19	3.25E-02	1.93E+00	1.91E-02	2.10E-02	5.80E-02	6.24E+00	8.90E+00	1.03E+01	1.23E+01	1.38E+01	2.06E+01
1_3	2.33E-02	4.06E-18	1.05E+00	1.27E-03	8.89E-04	2.64E-03	7.29E+00	1.01E+01	1.14E+01	1.37E+01	4.64E+00	7.28E+00
1_4	3.38E-01	2.56E-01	3.70E-17	2.74E-01	2.71E-01	2.31E-01	2.07E+22	6.19E+21	3.84E+22	3.62E+21	inf	inf
1_5	1.48E-02	1.33E-03	1.21E+00	3.81E-19	3.57E-05	8.15E-03	2.42E+00	3.73E+00	4.48E+00	5.48E+00	6.20E+00	1.13E+01
1_6	1.60E-02	9.42E-04	1.18E+00	3.68E-05	9.81E-19	7.00E-03	2.72E+00	4.12E+00	4.91E+00	5.99E+00	1.38E+01	2.12E+01
1_7	3.06E-02	8.81E-04	1.17E+00	1.06E-04	1.40E-04	4.97E-14	2.55E+00	4.31E+00	4.50E+00	6.08E+00	1.60E+01	2.11E+01
2_1	4.54E-01	3.78E-01	4.83E-02	3.95E-01	3.92E-01	3.54E-01	5.59E-15	2.05E-02	4.54E-02	8.99E-02	6.58E-01	1.36E+00
2_2	5.07E-01	4.34E-01	1.00E-01	4.55E-01	4.53E-01	3.91E-01	1.46E-02	1.60E-13	3.90E-03	1.93E-02	3.39E-01	8.29E-01
2_4	5.34E-01	4.65E-01	1.25E-01	4.79E-01	4.77E-01	4.41E-01	3.04E-02	2.98E-03	9.03E-14	5.49E-03	2.63E-01	7.06E-01
2_6	5.75E-01	5.11E-01	1.79E-01	5.26E-01	5.24E-01	4.89E-01	5.31E-02	1.45E-02	4.43E-03	5.64E-19	1.35E-01	3.88E-01
3_1	7.70E-01	7.33E-01	4.78E-01	7.26E-01	7.34E-01	7.07E-01	2.62E-01	1.90E-01	1.57E-01	1.25E-01	5.96E-19	3.50E-02
3_3	7.06E-01	6.56E-01	3.89E-01	6.73E-01	6.74E-01	6.17E-01	2.69E-01	2.09E-01	1.78E-01	1.44E-01	2.21E-02	3.52E-14

Figure 2 MSE map that represents how similar each dataset is.

As the MSE map shows, except for 1_4, all the MSE values are nice when we test the datasets from the same operating condition. Thus, 1_4 is seen as an abnormal. By checking the MSE map, we aim to pick one dataset from each operating condition. In condition 1, 1_3, 1_5, 1_6, and 1_7 do have suitable MSE values. Taking any one of them as one of the training sets is acceptable. In condition 2, both 2_2 and 2_4 are suitable. In condition 3, both 3_1 and 3_3 are suitable. Finally, we find that taking 1_5, 2_2, 3_3 as the training sets can bring the smallest MSE value when predicting either the whole datasets or the other testing datasets. Thus, these three datasets are set to be our training sets.

3.3. Single model to multiple models

Although the model trained with 1_5, 2_2, 3_3 brings precise prediction, β and η will lose their physical meaning when the value of them is checked. Take 1_1 as an example, the value of β is around 0.34 and the value of η is around 27.7. It is suspected that regressing with the datasets from different operating conditions is more likely to let the parameters optimized in an extreme situation. To solve this problem, it is intended to build one specific model for each specific operating conditions instead of building a single model to be responsible for the RUL prediction over whole operating conditions. 1_5, 2_2, and 3_3 are used to build three prediction models. The MSE results is shown in Table 2. And the magnitude of β and η will be settled in the scale of a normal degradation process.

Table 2 The MSE testing result by prediction model trained with 1_5, 2_2, and 3_3

Testing set	1_1	1_3	1_6	1_7	2_1	2_4	2_6	3_1
MSE	1.47E-02	1.37E-03	3.76E-05	8.15E-03	1.57E-02	3.75E-03	1.88E-02	2.47E-02

4. Conclusions

This work aims at establishing the Weibull Accelerated Failure Time Regression (WAFTR) models to predict the remaining useful life (RUL) of a benchmark bearing. The external features such as bearing load and rotary speed are used to categorize the local WAFTR model and then the internal features extracted from the frequency characteristics of sampled vibration data are used to adjust the scenario-based model parameters. In building the Weibull models, it is found that the parameter β expressed as exponential expansion and the parameter η expressed as linear expansion is the best choice to reflux the complex relationship between the RUL and operational features. By applying the multiple WAFTR models, the RUL prediction error is within 20% for all testing cases.

References

P. Kundu, A. K. Darpe, M. S. Kulkarni, 2019, Weibull accelerated failure time regression model for remaining useful life prediction of bearing working under multiple operating conditions, Mechanical Systems and Signal Processing, Volume 134.

W. Mao, J. He, J. Tang, Y. Li, 2018, Predicting Remaining Useful Life of Rolling Bearings Based on Deep Feature Representation and Long Short-Term Memory Neural Network, Advances in Mechanical Engineering, Volume 10.

P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, C. Varnier, 2012, PRONOSTIA: An experimental platform for bearingsaccelerated degradation tests, IEEE Int.

L. Guo, N. Li, F. Jia, Y. Lei, J. Lin, 2017. A recurrent neural network based health indicator for remaining useful life prediction of bearings, Neurocomputing, Volume 240, Pages 98-109.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50139-1

Supporting Hyperparameter Optimization in Adaptive Sampling Methods

Joschka Winz[a*], Florian Fromme[a], Sebastian Engell[a]

[a] *Process Dynamics and Operations Group, Department of Biochemical and Chemical Engineering, Technische Universität Dortmund, Emil-Figge Str. 70, 44227 Dortmund, Germany*
joschka.winz@tu-dortmund.de

Abstract

Surrogate models can be used to reduce the computational load when a simulation model is computationally costly to evaluate. This is the case if sophisticated thermodynamic models are integrated as e.g. the Perturbed-Chain Statistical Associating Fluid Theory (PC-SAFT) equation of state. When constructing surrogate models, the question of how to choose the training set arises. Recent research showed that promising results were obtained using adaptive or sequential sampling methods. In these approaches, the surrogate model predictions are used to identify additional promising sample locations. The results depend on the structure of the surrogate model, i.e. the choice of the hyperparameters. It is in general a tedious task to choose hyperparameters by trial and error, and a set of hyperparameters that is suitable in the initial phase may not be adequate anymore when the size of the training set increases significantly. Therefore, we here propose a methodology to incorporate hyperparameter optimization (HPO) into the adaptive sampling workflow. As this comes with a significant effort, HPO is only performed when it promises improvements.

Keywords: surrogate modeling, adaptive sampling, gray-box modeling, hyperparameter optimization, machine learning

1. Introduction

For an accurate description of complex phase equilibria in chemical process simulation and optimization, modern thermodynamic models like the Perturbed-Chain Statistical Associating Fluid Theory (PC-SAFT) should be used. When applying such models, the issue of the high computational cost caused by evaluating the phase equilibria arises. To overcome this issue, surrogate models can be applied to approximate the predictions of the complex thermodynamic models. Surrogate models are black-box models that can describe arbitrary relationships while being computationally cheap to evaluate. An overview of different surrogate model types is given in McBride and Sundmacher (2019). Recent applications include the development of biorefineries (Mountraki et al., 2020) and assisting in the optimization of chemical processes (Janus et al., 2020).
The process of creating a training set to fit a surrogate model to data is denoted as sampling. Recent work showed that adaptive sampling can be used to improve the accuracy of the surrogate model, see Nentwich et al. (2019) and Winz et al. (2021).
One important issue in the application of surrogate models is the choice of a suitable surrogate model structure. In a real scenario, there is only limited knowledge about the original functional relationship that the surrogate model must approximate, and determining the structure of the surrogate model structure and the parameters of the

training procedure is a complex task. This issue is further complicated when during the application of adaptive sampling methods additional samples are collected. Some surrogate model types may be suitable for a small initial training set, but not provide enough flexibility for a large training set in later iterations. So there is a need for a methodology to on the one hand automatically determine a suitable model structure for a specific data set and on the other hand detect when during the adaptive sampling it is necessary to adapt the model structure.
In the following, we describe such a methodology. In section 2, we introduce an example to which the methodology is applied. Section 3 entails the description of the developed methodology for simultaneous adaptive sampling and adaptation of the model structure, and in section 4 the results of the application of this method to the use case are presented.

2. Hydroformylation

The considered use case is the design of a process to conduct the hydroformylation of 1-dodecene in a thermomorphic solvent system, see Nentwich et al. (2019) for further details. In this process, a liquid phase reaction is conducted using an expensive homogeneous Rhodium-based catalyst. To minimize the loss of the catalyst, the solvent system is set up in a specific way. At elevated temperatures the mixture is homogeneous and the reaction can be performed with little transport limitations. After cooling down the reaction medium, two liquid phases form, which enables the separation of the mixture using a decanter and the recycling of the catalyst in the polar phase.
The intricate thermodynamic phase behavior of the reaction mixture can be accurately modeled using the equation of state PC-SAFT (Schäfer et al., 2012). But due to the high computational cost involved in solving the PC-SAFT equations, the fugacity coefficients are approximated by a surrogate model as described in the next sections. As a quaternary component system is considered and the temperature has a strong influence on the miscibility, the four fugacity coefficients are a function of four inputs each.

3. Methodology

3.1. Surrogate model types

In this work, different types of surrogate models are considered. Surrogate models can be classified with respect to several aspects. A property that is especially relevant in this context is parametricity. In parametric methods, the functional form of the response surface is fixed before the training. In non-parametric methods, the functional form is determined during the training, which typically leads to inherent scaling of the model complexity with the training data size (James et al., 2013). To cover both categories of models, in this work both Gaussian process regression (GPR) models, as an example of a non-parametric method, and parametric artificial neural network (ANN) models are applied.
For the GPR models, the considered hyperparameters are the types of basis and kernel functions. For the ANN models, these are the numbers of layers and nodes in each layer. The training of the ANN is done using the Levenberg-Marquardt algorithm with early stopping after six iterations of non-decreasing validation errors.

3.2. Adaptive sampling methodology

Adaptive sampling methods are used to select sets of samples such that accurate surrogate models are be obtained when they are fitted to the samples. In this work, the adaptive sampling methodology described in Nentwich et al. (2019) is applied.

A key element of sampling methods is to introduce a measure of how promising it is to evaluate a new sample and to add it to the sample set. This measure is commonly denoted as η_j. Here it is computed as the weighted mean of the scaled Jackknife variance $\hat{\sigma}_j^{\,2}$ and the scaled nearest neighbor distance d_j.

$$\eta_j = \alpha \frac{\hat{\sigma}_j^{\,2}}{\max\limits_j \hat{\sigma}_j^{\,2}} + (1-\alpha)\frac{d_j}{\max\limits_j d_j} \tag{1}$$

The weighting factor α reflects the trade-off between exploration and exploitation. To ensure appropriate surrogate model predictions which are necessary for reliably computing η_j, the hyperparameters of the model have to be chosen carefully.

3.3. Hyperparameter optimization

The task of automatically determining hyperparameters of machine learning models and methods has attracted much interest in the last years. Recently, Bayesian optimization has become popular. In Bayesian optimization, a GPR surrogate of the cost function is constructed and used to define an acquisition function that is utilized to identify promising inputs by considering both the predicted value of the cost function and the estimated uncertainty. A commonly used acquisition function is the expected improvement (EI). A variant, where the EI relative to the expected evaluation time is optimized, is proposed by Snoek et al. (2012). This tends to favor sets of hyperparameters that lead to a low evaluation time. In general, this procedure is time-consuming as the surrogate model has to be trained multiple times. Thus, the next section deals with determining when the HPO is promising.

3.4. Supporting hyperparameter optimization

As the HPO routine is computationally costly to perform, it is impractical to apply it in every iteration of the adaptive sampling process. Therefore, in this section we describe a criterion to determine when the current set of hyperparameters is not likely to lead to sufficient progress. Its combination with adaptive sampling is shown in Algorithm 1.

As can be seen in Algorithm 1, two main elements are added to the adaptive sampling procedure, the hyperparameter optimization itself and a method to evaluate whether the hyperparameter optimization is promising. For this evaluation, we propose to adapt the slope ratio criterion proposed by Nuchitprasittichai and Cremaschi (2013) using the relative progress, RP.

$$\mathrm{RP}_i = \left| \left(\frac{\mathrm{CVE}_{i-1}}{\mathrm{CVE}_i} - 1 \right) (n_i - n_{i-1})^{-1} \right| \tag{2}$$

As can be seen in (2), RP is computed from the cross-validation error in iteration , CVE_i, and the number of samples in iteration i, n_i.This is motivated by the idea that if the errors in one iteration and in the next one are similar, the progress is small, while large progress is observed if the error is reduced by a significant factor. This is different from the criterion proposed in previous work (Nuchitprasittichai and Cremaschi, 2013), where the absolute difference of the errors is considered. This modification is done here to consider a large relative change of the error as significant even if the absolute error is small, as it commonly occurs in later iterations of the algorithm.

As this ratio tends to give noisy results, a smoothening is performed by applying an exponential fit to the data points of $\{n_i\}, \{\mathrm{RP}_i\}$ that were gathered since the last HPO was performed, whenever there are at least $\mathrm{minCVEs} = 4$ error values available. An

optimization of hyperparameters is conducted if there is not sufficient progress, i.e. at the sample size where the exponential fit reaches a minimum threshold ϵ.

Data: Initial training input and output sets X and Y
Conduct hyperparameter optimization using X and Y to obtain an initial set of hyperparameters $\{\lambda_i\}$
Initialize the set of cross-validation errors CVEs $\leftarrow \emptyset$
while stopping criterion not met **do**
  Train surrogate models with hyperparameters $\{\lambda_i\}$
  Update the set of cross-validation errors CVEs $\leftarrow$ CVEs $\cup$ CVE
  if number of elements in CVEs $>$ minCVEs **do**
    Compute $\{\mathrm{RP}_i\}$ from CVEs (see eq. (2)) and perform an exponential fit for $\{\mathrm{n}_i\}, \{\mathrm{RP}_i\}$
    if number of samples where exponential fit meets threshold $\epsilon <$ number of samples in X **do**
      Conduct hyperparameter optimization using X and Y to give a new set of hyperparameters $\{\lambda_i\}$
      Reset the set of cross-validation errors CVEs $\leftarrow$ CVE
      Train surrogate models with the new hyperparameters $\{\lambda_i\}$
    end
  end
  Adaptive sampling using trained surrogate models to update sample sets $X \leftarrow X \cup X_{new}$, $Y \leftarrow Y \cup Y_{new}$
end

Algorithm 1: Inclusion of hyperparameter optimization in adaptive sampling

4. Results

4.1. Hyperparameter dependency

A first investigation is conducted to answer the question of how strongly the prediction capabilities of the considered surrogate models depend on the choice of the hyperparameters. For this, three training sets of varying sizes with 50, 500, and 5000 training examples each were created twice using Latin hypercube sampling. One set was used for the optimization of the hyperparameters, and the other one for the evaluation of the performance of the surrogate models. These sets were created by evaluating the fugacity coefficient of the liquid phase with PC-SAFT for different combinations of temperature and composition (a relationship with 4 inputs and 4 outputs).
A hyperparameter optimization was performed using each of the three sets of different sizes performing 50 iterations using the EI acquisition function. The optimized hyperparameter sets were used to create surrogate models based on different sets also containing 50, 500, and 5000 training examples. The resulting performance is shown in Figure 1 in terms of the normalized RMSE computed on a separate test set of 964120 samples. In this figure, it can be seen that for GPR models all sets of hyperparameters perform similarly on any training set size, as expected for a non-parametric model. On the other hand, for ANN surrogate models there is a strong dependency on the model structure. The model where the hyperparameters were optimized for 50 samples does not perform well when being trained on a set of 5000 samples and vice versa.

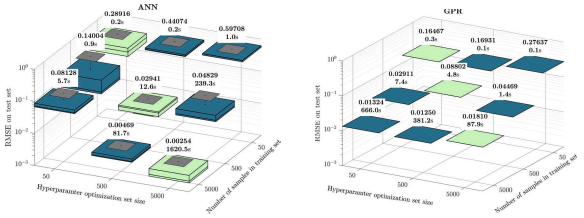


Figure 1: Normalized test set error of different surrogate models for hyperparameters optimized for different sizes of the training sets. The top numbers denote the mean value, the bottom numbers are the mean computation times of training on a standard desktop PC

4.2. Combination of adaptive sampling and hyperparameter optimization

As shown, the ANN surrogate models are sensitive to the choice of the hyperparameters. Thus, the adaptive sampling loop from (Nentwich et al., 2019) was extended with the described supporting hyperparameter optimization. The adaptive sampling was initially conducted only performing an HPO at the first iteration to analyze the chosen criterion for re-optimization as shown below.

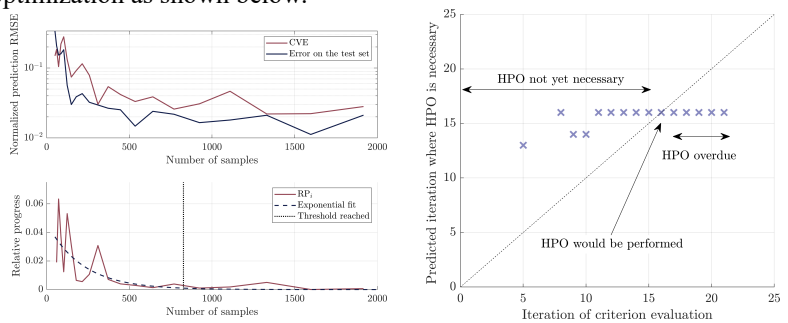


Figure 2: Left, top: Error-values over numbers of samples. Left, bottom: Improvement criterion and exponential fit. Right: Predicted iteration of need for HPO at different adaptive sampling iterations as crosses, identity line as a dashed line

In this figure in the left upper part, the trajectory of the error on the test set and the CVE are presented over the iterations of the adaptive sampling. In the bottom part, the RP criterion values, as well as the exponential fit, are shown. Using the criterion, a HPO should be performed after 924 samples have been collected as shown by the dotted line denoting where the criterion reaches the threshold of 0.001. On the right-hand side, it is shown for which iteration the criterion is fulfilled, with the criterion being evaluated at different stages of the adaptive sampling. It can be seen that after the 11th adaptive sampling iteration the RP criterion consistently predicts the HPO to become necessary in iteration 17. This holds true also for later iterations, where the HPO eventually becomes overdue. This shows the robustness of the methodology and confirms the assumption, that the RP values can be smoothened with an exponential fit.

After validation, the methodology was applied by performing adaptive sampling and HPO, when the RP criterion is met. The results of this procedure are shown in Figure 3.

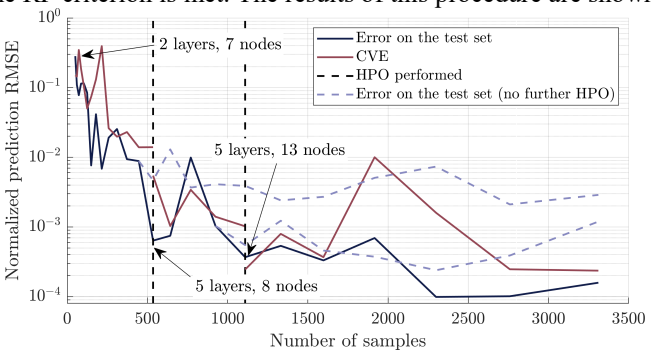


Figure 3: Adaptive sampling in combination with supporting hyperparameter optimization, the hyperparameters are visible as annotations. The dashed lines show the evolution of the error for the test set when the hyperparameters were kept constant.

It can be seen from Figure 3 that when using adaptive sampling with hyperparameter optimization, the model structure is adapted twice, once after 642 samples have been collected and then once again for a sample set of 1109 samples, which leads to a significant decrease in the test set error, compared to the case, where no HPO was performed.

5. Conclusion

In this work, a novel methodology for combining hyperparameter optimization and adaptive sampling is presented and applied to the approximation of the liquid phase fugacity coefficients obtained by PC-SAFT in a quaternary thermomorphic solvent system. The methodology is based on evaluating the slope of the cross-validation error to detect when the accuracy of the surrogate models is not increasing sufficiently anymore. The manual (re-)setting of the hyperparameters, which is a tedious and error-prone task, is avoided and replaced by an automatic procedure. The presented results show the necessity of hyperparameter re-optimization and the successful application of the automatic procedure.

6. Acknowledgment

This research has been supported by the project "KI-Inkubator-Labore in der Prozessindustrie - KEEN", funded by the Bundesministerium für Wirtschaft und Energie (BMWi) under grant number 01MK20014T. This support is gratefully acknowledged.

References

G. James, F. Witten, T. Hastie, R. Tibshirani, 2013, An Introduction to Statistical Learning, Springer New York

T. Janus, A. Luebbers, S. Engell, 2020, Neural Networks for Surrogate-assisted Evolutionary optimization of Chemical Processes, in Proc. 2020 IEEE Congress on Evolutionary Computation (CEC), pp. 1–8

K. McBride, K. Sundmacher, 2019, Overview of Surrogate Modeling in Chemical Process Engineering, Chemie-Ingenieur-Technik, Vol. 91, Issue 3, pp. 228–239

A. D. Mountraki, B. Benjelloun-Mlayah, A. C. Kokossis, 2020, A Surrogate Modeling Approach for the Development of Biorefineries. Frontiers in Chemical Engineering, 2:568196

C. Nentwich, J. Winz, S. Engell, 2019, Surrogate Modeling of Fugacity Coefficients Using Adaptive Sampling. Industrial and Engineering Chemistry Research, Vol. 58, Issue 40, pp. 18703–18716

A. Nuchitprasittichai, S. Cremaschi, 2013, An algorithm to determine sample sizes for optimization with artificial neural networks, AIChE Journal, Vol. 59, Issue 3, pp. 805–812

E. Schäfer, Y. Brunsch, G. Sadowski, A. Behr, 2012, Hydroformylation of 1-dodecene in the thermomorphic solvent system dimethylformamide/decane. Phase behavior-reaction performance-catalyst recycling, Industrial and Engineering Chemistry Research, Vol. 51, Issue 31, pp. 10296–10306

J. Snoek, H. Larochelle, R. P. Adams, 2012, Practical Bayesian optimization of machine learning algorithms, Advances in Neural Information Processing Systems, Vol. 4, pp. 2951–2959

J. Winz, C. Nentwich, S. Engell, 2021, Surrogate Modeling of Thermodynamic Equilibria: Applications, Sampling and Optimization, Chemie-Ingenieur-Technik

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50140-8

Optimal Catalyst-Reactor Design for Load-Flexible CO_2 Methanation by Multi-Period Design Optimization

Ronny T. Zimmermann[a], Jens Bremer[b], Kai Sundmacher[a,b,*]

[a]*Otto von Guericke University Magdeburg, Chair for Process Systems Engineering, Universitätsplatz 2, 39106 Magdeburg, Germany*
[b]*Max Planck Institute Magdeburg, Department Process Systems Engineering, Sandtorstraße 1, 39106 Magdeburg, Germany*
sundmacher@mpi-magdeburg.mpg.de

Abstract

Fixed-bed reactors employed within Power-to-X technologies are expected to be subject to volatile process feeds, in order to minimize intermediate buffer or storage systems. In this context, catalyst particles with an inert shell have proven suitable to prevent reactor runaway and severe temperature excursion at various loads and during load-changes, while still allowing for high reactor space-time yields. The aim of this work is to determine how much active catalyst mass can be saved by multi-period design optimization of methanation reactors with one, two, and infinite fixed-bed segments. Our results show that the amount of active mass, which can be saved with a single fixed-bed segment, is negligible. If more than one fixed-bed segment is considered, up to 45.2 % can be saved at the expense of a more complex dynamic reactor behavior. In the worst case, severe temperature excursions during transient scenarios might occur, which is not the case, if only a single fixed-bed segment is considered.

Keywords: Multi-Period Design Optimization, Load-Flexible Fixed-Bed Reactor, Dynamic Reactor Operation, Carbon Dioxide Methanation

1. Introduction

Fixed-bed reactors are commonly employed to carry out heterogeneously catalysed gas-phase reactions. As they are often the central process unit, their optimal operation and design has been investigated for decades. Special attention is often paid to heat management, as many reactions are highly exothermic. With respect to the increase in volatile process feeds (e.g., due to renewable energies), load-flexible operation of fixed-bed reactors is expected to become increasingly important (Bremer and Sundmacher, 2019). Consequently, the employed reactors must not only operate reliably in one steady state, but in many as well as in the transitions between the steady states. Otherwise, severe temperature excursions could deactivate the catalyst or even destroy the reactor material.
The design of the employed catalyst particles has a significant impact on the dynamic behaviour of fixed-bed reactors. Core-shell catalyst particles consisting of an active core and an inert shell (‘egg-yolk’ catalyst particles) exhibit favourable properties in this context (Zimmermann et al., 2020). In our recent work, we compare such catalyst particles

to ordinary fixed-bed dilution, exemplified for carbon dioxide methanation (Zimmermann et al., 2022). In both cases, industry scale fixed-bed reactors with one, two, and infinitely many segments in axial direction are considered.

We conclude, that fixed-beds consisting of core-shell catalyst particles are favourable for load-flexible reactor operation compared to ordinary fixed-bed dilution. The reason for this can be seen in Fig. 1. Whereas ordinary fixed-bed dilution decreases the effective reaction rate by a constant factor over the whole temperature range, core-shell catalyst particles of the 'egg-yolk' type decrease the effective reaction rate selectively at high temperatures, where the diffusion through the inert shell is rate-determining. In consequence, much higher effective reaction rates are obtained towards the reactor outlet, where no reduction of the heat release rate is necessary. This results in higher space-time yield and lower pressure loss.

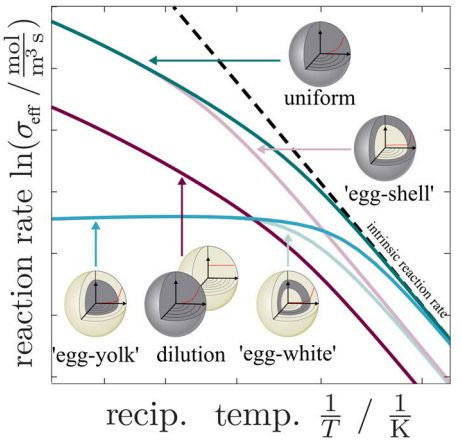


Figure 1: Schematic comparison of the effective reaction rates of different catalyst particle concepts and ordinary fixed-bed dilution in Arrhenius plots.

If the catalyst particles are operating in the regime of diffusion limitation by the inert shell, the active mass in the centre of the catalyst particles does not contribute to the conversion of reactants. However, as the active mass often consists of precious noble metals, it should be located within the particles only where needed, in order to save investment costs. In this work, we extend the discussed procedure to include catalyst particles, which consist of an inert core and an inert shell with a layer of catalytically active mass in between (a.k.a. 'egg-white' catalyst particles), as seen in Fig. 1. We aim to minimize the amount of active mass needed, while ensuring a high space-time yield and the compliance of conversion and temperature bounds within the entire reactor.

2. Fixed-Bed Reactor and Catalyst Particle Model

A pseudo-homogeneous fixed-bed reactor model is employed. In addition to the carbon dioxide conversion X and radially averaged temperature T, the fixed-bed's centre temperature $\widehat{\mathrm{T}}_{\mathrm{C}}$ is approximated by the correlation of Dixon (1996). The reaction kinetic model of Koschany et al. (2016), which considers no side reactions, is used to describe the methanation reaction kinetics. Further details on the remaining constitutive equations are given by Zimmermann et al. (2022).

$$\varepsilon_R \frac{\partial X}{\partial t} = -u \frac{\partial X}{\partial z} + \frac{M_{\mathrm{CO_2}}}{\rho\, y_{\mathrm{CO_2,in}}} (1-\varepsilon_R)\, r_{\mathrm{eff}} \tag{1}$$

$$(\rho c_p)_{\mathrm{eff}} \frac{\partial T}{\partial t} = -u_{\mathrm{in}}\, \rho_{\mathrm{in}} c_p \frac{\partial T}{\partial z} + \frac{\partial}{\partial z}\left[\Lambda_{\mathrm{ax}} \frac{\partial T}{\partial z}\right] - \frac{4U}{D}(T - T_{\mathrm{cool}}) - H_{\mathrm{r}}\,(1-\varepsilon_R)\, r_{\mathrm{eff}} \tag{2}$$

The effective reaction rate of the catalyst particles is calculated according to

$$\frac{1}{\eta} = (1+\delta)^{n+1}\left[\frac{\Phi}{\tanh(\gamma\Phi)} + \frac{\Phi^2}{\mathrm{Bi}_{\mathrm{ext}}(1+\delta)^n} + \frac{\Phi^2}{\mathrm{Bi}_{\mathrm{int}}}\right] \tag{3}$$

$$\eta = \frac{r_{\mathrm{eff}}}{r_{\mathrm{int}}}, \qquad \Phi = \frac{R_{\mathrm{core}}}{n+1}\sqrt{\frac{\sigma(c_{\mathrm{bulk}})}{\mathfrak{D}_{\mathrm{core}}(c_{\mathrm{bulk}} - c_{\mathrm{bulk,eq}})}}, \qquad \mathrm{Bi}_{\mathrm{ext}} = \frac{\mathfrak{D}_{\mathrm{core}}(n+1)}{\beta R_{\mathrm{core}}}$$
$$\mathrm{Bi}_{\mathrm{int}} = \frac{\mathfrak{D}_{\mathrm{core}} R_{\mathrm{core}}^{n-1}\Psi(n+1)}{\mathfrak{D}_{\mathrm{shell}}}, \qquad \gamma = \frac{R_{\mathrm{core}}^{n+1} - R_0^{n+1}}{R_{\mathrm{core}}^{n+1}}, \qquad \delta = \frac{R_{\mathrm{shell}} - R_{\mathrm{core}}}{R_{\mathrm{core}}} \tag{4}$$

This equation is derived by extending the procedure given by Zimmermann et al. (2022) and is an explicit analytical approximation of the effectiveness factor for a single reaction. It considers the influence of an inert core with radius R_0 in the dimensionless group γ, and the inert shell with radius R_{shell} in the dimensionless group δ.

3. Multi-Period Design Optimization

Multi-period design optimizations, as proposed by Grossmann and Sargent (1979), are performed by employing the pseudo-homogeneous reactor model. Three periods of equal temporal length and carbon dioxide loads of 0.5 t/h, 1.0 t/h, and 1.5 t/h are considered. The operating variables consist of the inlet pressure, velocity, and temperature as well as the coolant temperature and wall heat transfer coefficient. The inlet feed is always a stoichiometric mixture of pure carbon dioxide and hydrogen. The design variables are the reactor's tube length, diameter, and number, together with the catalyst particle diameter, shell porosity, pore diameter, and thickness, as well as radius of the inert core. In each state, the temperature in the centre of the fixed-bed must never exceed 750 K, and the conversion at the outlet of the reactor must be at least 95 %. The remaining constraints are given by Zimmermann et al. (2022).

The amount of catalyst, which can be saved by employing catalyst particles with an inert core is determined in two steps. First, the time-average methane space-time yield is maximized, considering a fully active core. In the second step, the amount of active mass in the reactor is minimized, while at least 99 % of the first steps space-time yield has to be achieved. Each of the steps is performed with one, two, and infinite segments, which can consist of different catalyst loadings. Details on the numerical solution strategy and the used solvers are given by Zimmermann et al. (2022).

Subsequently, dynamic load-change simulations are performed, by switching between the optimal operating parameters of the respective cases. Additionally, a start-up and a shut-down are investigated. The initial conditions for the start-up are given as X_0 = 0 % and T_0 = 300 K. For the shut-down, T_{in} and T_{cool} are reduced to 300 K.

4. Results

4.1. Multi-Period Design Optimization

The results of the multi-period design optimization are shown in Fig. 2 and in Tab. 1. The optimal steady states of the first step are identical to those discussed in detail by Zimmermann et al. (2022) and serve as reference. In summary, the optimal reactor is designed such that the maximum heat transfer to the coolant is realized and the length is determined by the residence time required to achieve 95 % reactant conversion in each case. The latter is the main factor for the differing time-average space-time yields.

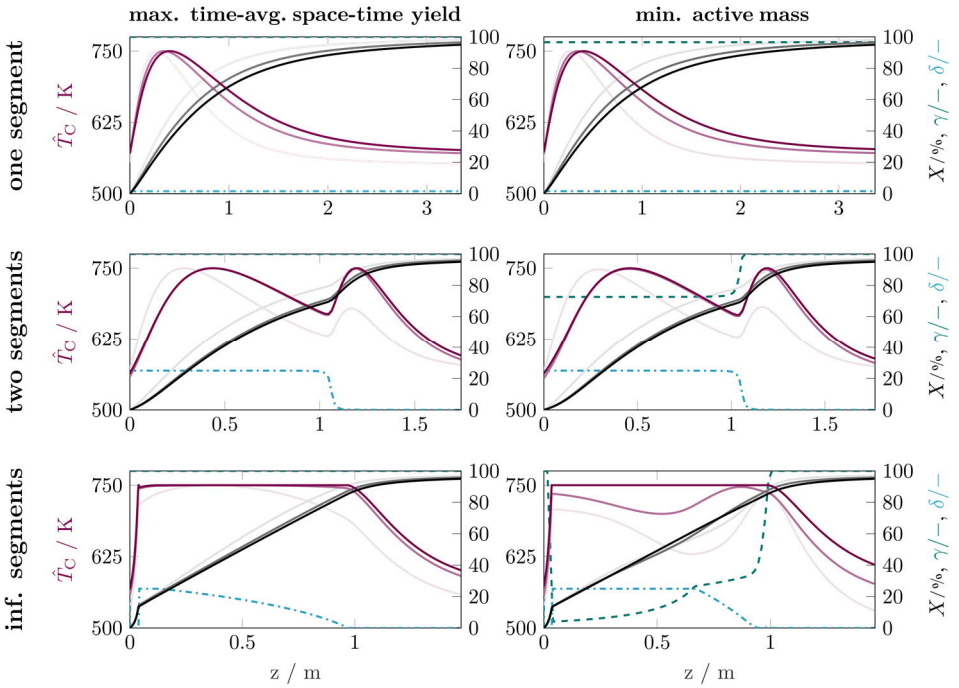


Figure 2: Multi-period design optimization results, darker colours correspond to higher loads (0.5 t/h, 1.0 t/h, and 1.5 t/h).

Table 1: Results of all multi-period design optimizations, respective objective function values in bold.

objective func.	max. time-average STY (step 1, reference cases)			min. active mass (step 2)		
fixed-bed segments	1	2	Inf.	1	2	Inf.
STY / kg m^{-3} s^{-1}	**0.3850**	**0.7440**	**0.8773**	0.3811 (-1%)	0.7366 (-1%)	0.8771 (-0%)
m_{act} / kg	215.1	82.2	77.3	**211.1** (-2%)	**73.95** (-10%)	**42.3** (-45%)

Compared to these results, reactors with minimized active mass show very similar trends, especially if one or two segments are considered. In the case of a single fixed-bed segment, almost no potential for removing active mass from the centre of the catalyst particles is available. This is because it is only suitable to save active mass in the particle cores in the regime of mass transfer limitation, since no reactants enter the interior of the particles in this case. However, this happens only at high temperatures near the reactor inlet. At low temperatures near the reactor outlet, the inert core acts as fixed-bed dilution (Fig. 1) and consequently longer reactors would be necessary to achieve the lower conversion constraint of 95%, which would lead to a significant drop of reactor space-time compared to the reactor with maximum time-average space-time yield.
Consequently, more potential to save active mass exists for the two-segment reactor. Only in first segment close to the reactor inlet, an inert shell is used to limit the effective reaction rate in order to stay within the temperature limit. Consequently, the core of the catalyst particles is chosen to be inert. As a large part of the reactants has been consumed in the first segment, no restriction of the effective reaction rate is necessary in the second segment. Thus, the optimizer prefers uniform catalyst particles, as these exhibit the highest effective reaction rate of all particle designs considered in this work. This results in a reduction of the required active mass by 10.0 %.
The results of the reactor with infinite segments confirm this trend in shape of singular arcs for the inert core and inert shell radii. Additionally, a segment of uniform catalyst is introduced directly at the reactor inlet, to increase the reactor temperature and thus the reaction rate. This case yields the maximum potential of active mass reduction, which is 45.2 %.

4.2. Dynamic Load-Change Simulations

Although all steady-states shown in the previous section convey to the upper temperature limit of 750 K, severe temperature excursions might still occur during dynamic transitions. Thus, dynamic load-change simulations have been performed as shown in Fig. 3.
All cases show a fast start-up behaviour and the ignited steady-states are achieved in less than two minutes. The one- and two-segment reactors with maximized space-time yields also show smooth transitions between the optimal sets of operating parameters. The same is true for their counterparts with minimized active mass, although slight oscillations can be observed after load-

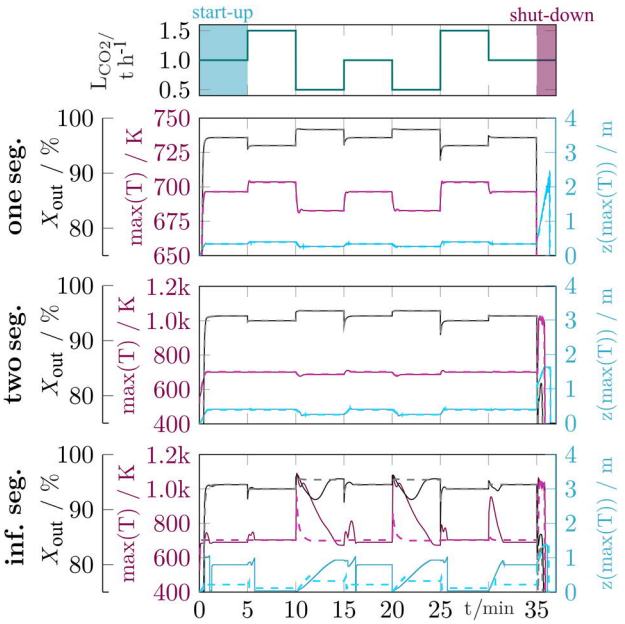


Figure 3: Dynamic load-change simulations, dashed lines show the results of reactors with maximized time-average space-time yield and solid lines show the results of reactors with minimized active mass. Please note the differing scales.

changes. In the cases with infinite segments, the reactors with maximized time-averaged space-time yield exhibit a travelling hot-spot, whenever the load is reduced. This is even more pronounced for the reactor with minimized active mass. In addition, such reactors also show temperature excursions when the load is increased.
During the shut-down, all reactors with more than one fixed-bed segment show pronounced temperature excursions. The reason for this is the high amount of inert mass of the segments close to the reactor inlet, which cool down quickly. Consequently, high amounts of reactants encounter the hot segments with high amounts of active mass towards the reactor outlet. As the catalyst particles in these segments also do not exhibit an inert shell, a steep increase in temperature is observed. The reactors with just one segment contain catalyst particles with an inert shell along the whole length, and thus no temperature excursions are observed.

5. Conclusions

The introduction of catalyst particles with an inert core offers the possibility to reduce expensive active mass in fixed-bed reactors on the one hand, but also reduces the effective reaction rate of the catalyst particles in the kinetically limited regime on the other. Thus, the reactor space-time yield might drop, such that an inert particle core is only purposeful at fixed-bed locations, where the reactants are not present in the particle core, which is the case in the regime of mass transport limitation. In consequence, only minimal potential is available for reducing the active mass if one fixed-bed segment and high space-time yield are aimed for, but this case offers fast and save transition during load-changes. Up to 45.2 % of active mass can be saved if more than one segment is considered. However, in these cases active mass is in direct contact with the surrounding gas phase, leading to the risk of severe temperature excursions during load-changes.

6. Acknowledgment

This research work was conducted within the DFG Priority Program SPP2080 'Catalysts and reactors under dynamic conditions for energy storage and conversion' and was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 406914011 (Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) - 406914011.). Ronny Zimmermann is also affiliated with the International Max Planck Research School (IMPRS) for Advanced Methods in Process and Systems Engineering, Magdeburg, Germany.

References

J. Bremer, K. Sundmacher, 2019. Operation range extension via hot-spot control for catalytic CO_2 methanation reactors. Reaction Chemistry & Engineering 4 (6), 1019–1037.
A. G. Dixon, 1996. An improved equation for the overall heat transfer coefficient in packed beds. Chemical Engineering and Processing: Process Intensification 35 (5), 323–331.
I. E. Grossmann, R. W. H. Sargent, 1979. Optimum design of multipurpose chemical plants. Industrial & Engineering Chemistry Process Design and Development 18 (2), 343–348.
F. Koschany, D. Schlereth, O. Hinrichsen, 2016. On the kinetics of the methanation of carbon dioxide on coprecipitated $NiAl(O)_x$. Applied Catalysis B: Environmental 181, 504–516.
R. T. Zimmermann, J. Bremer, K. Sundmacher, 2020. Optimal catalyst particle design for flexible fixed-bed CO_2 methanation reactors. Chemical Engineering Journal 387, 123704.
R. T. Zimmermann, J. Bremer, K. Sundmacher, 2022. Load-flexible fixed-bed reactors by multi-period design optimization. Chemical Engineering Journal 425, 130771.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50141-X

Data-Driven Modeling of Long-Term Performance Degradation in Solid Oxide Electrolyzer Cell System

Mina Naeini[a], James S. Cotton[b], Thomas A. Adams II[a*]

[a]*Department of Chemical Engineering, McMaster University, Hamilton, ON, L8S 4L7, Canada*
[b]*Department of Mechanical Engineering, McMaster University, Hamilton, ON, L8S 4L7, Canada*

tadams@mcmaster.ca

Abstract

One key challenge of Solid Oxide Electrolyzer Cell (SOEC) systems is degradation over long periods of time. Degradation decreases efficiency by increasing the electrical energy required for H_2 production. This paper presents the first step in managing long-term degradation in SOEC systems. In this work, the first data-driven dynamic model for the prediction of performance degradation in SOECs as a function of humidity, operating temperature, and current density was developed. The model was trained using experimental data from multiple data sets in the literature under various conditions. The model showed good agreement with validation data over 7000 h operation. One key finding is that the data show there are three distinct time regimes in which degradation behaviour is qualitatively different. This is likely due to different degradation phenomena, although the specific phenomena have not yet been isolated. This is significant for PSE applications because operators can choose to vary the operating conditions over time in order to predict, account for, or minimize the effects of long-term degradation.

Keywords: SOEC; Performance Degradation; Data-Driven Model; Operating Condition.

1. Background

When baseload power facilities or renewable sources of intermittent power are used for electricity generation, excess electricity may be produced during low demand hours. Producing surplus electricity can impose additional costs to electricity systems as it should be either curtailed or sold at a low price. Therefore, there is a great need for reliable energy storage systems at low demand hours. SOEC is a promising technology for energy storage that can utilize the excess electricity from the grid to electrolyze steam and generate hydrogen. H_2 is a clean, useful source of energy with high energy content per weight unit. The produced H_2 can be compressed and stored in cylinders for various purposes, supplied to stationary power generators such as fuel cells to produce electricity, or injected into the existing natural gas network to lower carbon intensity. SOECs are Solid Oxide Fuel Cells (SOFCs) which are operated in reverse. Indeed, the same technology can either generate electricity via some exothermic electrochemical reactions (SOFC mode) when supplied with fuel and air, or it can electrolyze water and produce H_2 through endothermic electrochemical reactions (SOEC mode), when supplied with steam and electricity. SOFCs and SOECs feature several advantages over some existing power generation technologies and electrolyzer systems such as high efficiency, low noise, and reduced greenhouse gas (GHG) emissions. However, this technology suffers the

drawback of significant long-term degradation under normal operating conditions. This issue has been a barrier for commercialization of SOFCs and SOECs and has limited their application in the industry (Zaccaria et al., 2015). In our prior work, we were able to successfully overcome this problem for SOFCs specifically by constructing a mathematical model for dynamic simulation of the long-term performance drop in SOFC systems (Naeini et al., 2021). Using this model in a techno-economic analysis (TEA) enabled us to determine the optimal capacity and corresponding optimal dynamic operational trajectories of SOFCs that make them more cost-competitive with traditional power generation, by managing their long-term degradation and increasing their useful lifetime (Naeini et al., 2021). This was possible because the individual mechanisms that cause SOFC degradation are known, and thus first principles could be considered directly in the model.

Since the electrochemical reactions and degradation mechanisms in SOFCs are different from those in SOECs, the model developed for degradation in SOFCs cannot be used to predict degradation in SOECs. As such, a dynamic model should be constructed for simulation of long-term performance degradation in SOECs. This is essential for identifying optimal sizing and operating strategies of SOEC systems for a cost-effective energy storage over an extended lifetime, with controlled degradation. Our broad literature survey revealed that a dynamic degradation-based model for SOECs is lacking in the open literature. Also, the mechanisms of degradation phenomena in electrolysis mode are not well known, which prevents us from developing a first-principles model. However, there have been number of experiments investigating impacts of operating conditions on SOEC performance degradation. Hoerlein et al.'s (2018) experiments in particular showed the impacts of current density, humidity, and operating temperature on voltage and ohmic resistance of 20 SOECs over 1000 h of operation. This is a good, representative dataset that contains acceptable ranges of operating parameters as it includes current density ranging from 0 to 1.5 A/cm^2, humidity varying from 40 % to 80 %, and temperature from 750 to 850 °C. This dataset is used in the current work to develop a data-driven model for degradation of SOEC.

2. Modeling strategy

The following information shows that the time evolution of SOEC's ohmic resistance can be used to calculate SOEC degradation. SOECs produce H_2 at a constant rate as long as the current density supplied to the cell is constant (Eq.(1)).

$$r_{H_2} = \frac{i}{nF} \tag{1}$$

where r_{H_2} is the rate of H_2 production, i, n, and F are the current density, moles of electrons involved in the electrolysis electrochemical reactions, and Faraday constant, respectively. The ohmic resistance of SOECs increases over time due to degradation. According to Ohm's law (Eq.(2)), the voltage of the SOEC and therefore the energy required to produce a constant amount of H_2 increase with its degradation (Eq.(3)).

$$V = R\, i \tag{2}$$

$$P = R\, i^2 = V\, i \tag{3}$$

In these equations R, V, and P represent the SOEC's ohmic resistance, voltage, and power, respectively. The magnitude of resistance increase depends on the operating conditions. Therefore, a model that predicts time evolution of resistance as a function of current density, humidity, and temperature will allow us to quantify performance degradation in the SOEC. Data from figure 1 is used in the present work to develop a model for SOEC degradation. Given in the y-axis of these plots is ΔR_{ohm} which shows the ohmic resistance increase with respect to the initial ohmic resistance of the cells.

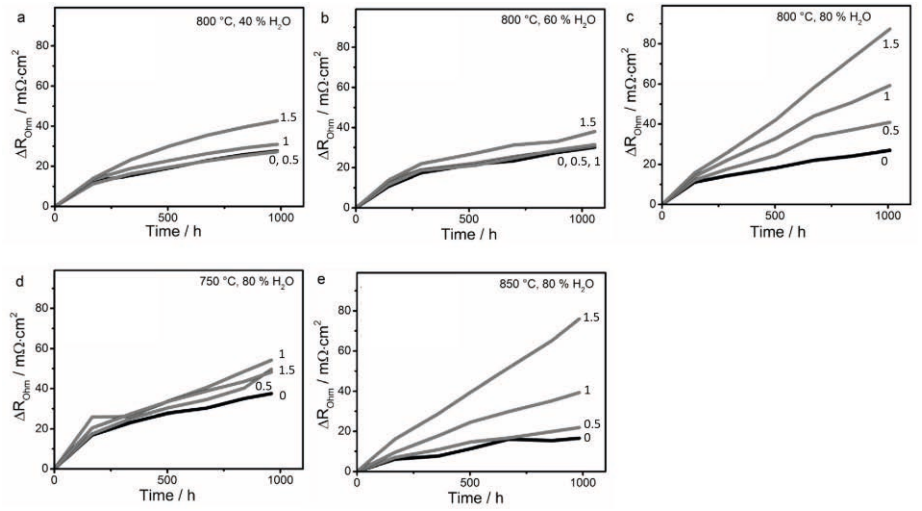


Figure 1. Time evolution of ohmic resistance of SOECs at various current densities from 0 to 1.5 A/cm^2 and a) 800°C and 40% humidity, b) 800°C and 60% humidity, c) 800°C and 80% humidity, d) 750°C and 80% humidity, and e) 850°C and 80% humidity. Original data from (Hoerlein et al.).

The data show an unstable behavior in some SOECs during the first 200 h of operation, where the resistance, voltage, and other parameters do not follow a reproducible or consistent trajectory from run to run. This behavior is known as cell conditioning or transition period and cannot be modelled since it does not conform to any known pattern (Sohal; Hubert). However, the data show that the cumulative degradation of the cell at the 200h mark does correlate well with temperature, humidity, and current density, even though the trajectory in getting there does not. We removed the first 200 h from the training data and as a result the developed model applies only beyond 200 h. It should also be noted that even though Hoerlein et al.'s (2018) data includes voltage trajectories of SOECs, voltage is not used for building the SOEC model in this study. The reason is that when there is not any current in the external circuit, i.e. the open circuit voltage condition (OCV), the voltage remains constant and as a result it is not a good state variable for predicting degradation of SOECs.

ALAMO (Automated Learning of Algebraic Models) was used to develop the data-driven model. This software employs a machine learning approach to learn accurate and simple algebraic models from the training dataset (Wilson and Sahinidis). ALAMO models are typically linear combinations of nonlinear transformations of the input variables. The software uses an optimization approach to find not only the best fit of model parameters, but the selection of the basis functions themselves. This avoids overfitting and helps choose basis functions that best characterize the data.

2.1. Development of the linear model from 200h to 2500h

First, a partial model was developed based on Hoerlein et al's (2018) dataset (which has data out to 1000h), shown in Eq. (4).

$$R_{t'} = R_0 + \underline{0.019\,TH + 7.290\,Hi} + [0.033\,Ht' + 0.017\,it'] \quad (4)$$

where t' is time since 200h (h), $R_{t'}$ is ohmic resistance at t' (mΩ.cm^2), R_0 is resistance of virginal SOEC at time 0h (mΩ.cm^2), T is temperature (°C), and H is humidity of the fuel feed expressed as mole fraction of water. Eq.(4) includes two time-independent (underlined) and two time-dependent (enclosed in square brackets) terms. Underlined terms show the change in resistance from R0 within the first 200h. While time-dependent terms show the increase in resistance from 200h and onward. Since the data for the first 200h were removed from training data, the model cannot indicate time evolution of resistance within this period. Instead, it quantifies the total change in resistance from R0 within the first 200h. As can be seen, given a constant molar humidity and current density, ohmic resistance of SOEC increases linearly with time for at least 1000h, and so we refer to this model as the "linear model". This is in agreement with number of other studies that reported linear trend in time for SOEC degradation at fixed operating conditions (Hauch; Hubert; Trofimenko et al.). These studies all consider operation for 2500h or less. Eq.(4) was fit with R^2=0.94 across the training datasets considered up to 1000h. Rapid transients may have impacts which are not considered in the model.

2.2. Validation of the linear model out to 2500 h

In order to ensure validity of the developed model for predicting performance degradation in SOECs under given operating conditions, Tietz et al.'s (2013) experimental data – which was not considered in building the model – was used to compare with the model. In Tietz et al. (2013) an SOEC was operated at 778±6°C out to 7600h, supplied with current density of 1A/cm^2, and 80% humidity. We calculated the SOEC's ohmic resistance trajectory using Eq. (4) and then its voltage trajectory using Eq. (2) under the given condition. The results, given in Figure 2, indicate capability of this linear model to simulate performance deterioration in SOECs for the first 2500 h of operation (R^2=0.87). After 2500h, it becomes more apparent that the SOEC degrades sub-linearly and the degradation rate decreases. We note that most studies that conclude that degradation rates are linear, such as Hauch et al. (2006), Hubert et al. (2018), and Trofimenko et al. (2017) do not run the experiments long enough to experience this inflection point around 2500h.

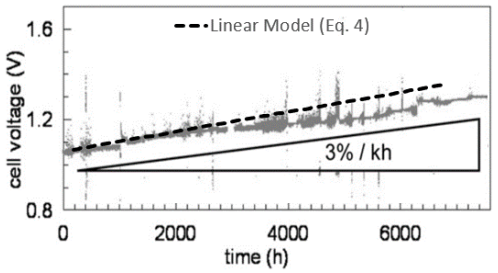


Figure 2. Results of the developed model for SOEC performance degradation compared to experimental data from (Tietz et al.).The dashed line is our model, drawn on top of a modified version of the original figure reproduced from that work.

2.3. Sublinear model out to 7000 h

The model was modified to reflect the sublinear nature of the degradation by adding a power term to the time component, and fitting the exponent to Tietz et al.'s data. The basis functions and their parameters were unmodified from Eq. 4. The sublinear model is given in Eq.(5), in which degradation is sublinear to the order of 0.97, and R^2=0.95 for Tietz et al.'s data out to 7000h.

$$R_{t'} = R_0 + 0.019\,T + 7.290Hi + (0.033H + 0.017i)(t')^{0.97} \quad (5)$$

A comparison between the linear model, sublinear model, and experimental data is provided in Figure 3. As can be seen, the original and modified models are almost equally good for simulation of short-term operation. But for the long-term operation, sublinear model is significantly better than the linear one.

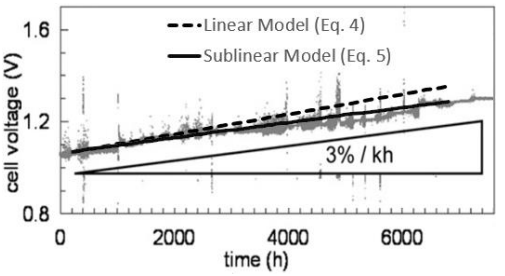


Figure 3. Results of the linear and sublinear models for long-term performance degradation in SOEC compared to experimental data from (Tietz et al.). Figure modified from the original.

3. Discussion and Future Work

One key finding is the sublinear nature of the degradation rate, but this is supported only by one data set since most experiments in the open literature are not conducted for long enough in order to see this trend. Additional research is needed to validate this sublinear characteristic. However, sublinear behavior makes sense, as in practice SOECs can be used for many years. If the degradation rate was truly linear such that long-term degradation continued at the same rate as in the first 1000h, the amount of degradation would be much higher than what is actually experienced in practice. For example, comparing extrapolating the sublinear (Eq. (5)) and linear (Eq. (4)) models out to 10 years, the linear model would predict a required voltage of 30% higher than the sublinear model. This will make a significant difference in design and operation of large-scale systems. Using either the linear or sublinear model, as well as Eq.(2) and Eq.(3), one can calculate voltage of SOECs and the required power supply for producing a specific amount of H_2 at different points of operating time considering the degradation. This will enable system designers to find the optimal capacity and operation strategy of SOEC system for an economical energy storage with reduced degradation rate. As such, a TEA

will be performed on SOEC systems as the future work. The findings will improve cost-effectiveness of this technology and will help its commercialization.

4. Conclusions

This paper presented a data-driven model constructed from experimental data derived from the literature for prediction of the performance deterioration in SOEC. To the best of authors' knowledge, this is the first dynamic model for simulation of the SOEC degradation under different operating conditions with high accuracy across a range of cell conditions. This model, developed using ALAMO, is a composition of nonlinear functions of operating parameters namely humidity, temperature, and current density. The model was well-validated using many training and testing sets for the first 2500h of operation, but the extension to 7000h requires additional validation. Both the linear (Eq 4) and sublinear (Eq 5) models work equally well for those first 2500h but the linear model breaks down afterward. The results indicated that the rate of SOEC degradation depends on the humidity and current density, and decays sublinearly in the long-term.

References

Hauch, Anne. "Performance and Durability of Solid Oxide Electrolysis Cells." *ECS Meeting Abstracts*, vol. 37, no. October, 2006, doi:10.1149/ma2006-01/24/842.

Hoerlein, M. P., et al. "A Parameter Study of Solid Oxide Electrolysis Cell Degradation: Microstructural Changes of the Fuel Electrode." *Electrochimica Acta Journal*, vol. 276, 2018, pp. 162–75, doi:10.1016/j.electacta.2018.04.170.

Hubert, M. "Durability of Solid Oxide Cells: An Experimental and Modelling Investigation Based on Synchrotron X-Ray Nano-Tomography Characterization." *PhD Thesis*, 2018.

Naeini, Mina, Haoxiang Lai, et al. "A Mathematical Model for Prediction of Long-Term Degradation Effects in Solid Oxide Fuel Cells." *Industrial and Engineering Chemistry Research*, vol. 60, no. 3, 2021, pp. 1326–40, doi:10.1021/acs.iecr.0c05302.

Naeini, Mina, James S. Cotton, et al. *Economically Optimal Sizing and Operation Strategy for Solid Oxide Fuel Cells to Effectively Manage Long Term Degradation*. 2021.

Sohal, Manohar S. *Degradation in Solid Oxide Cells During High Temperature Electrolysis*. 2009.

Tietz, F., et al. "Degradation Phenomena in a Solid Oxide Electrolysis Cell after 9000 h of Operation." *Journal of Power Sources*, vol. 223, Elsevier B.V, 2013, pp. 129–35, doi:10.1016/j.jpowsour.2012.09.061.

Trofimenko, Nikolai, et al. "Optimization of ESC Performance for Co-Electrolysis Operation." *ECS Transactions*, vol. 78, no. 1, 2017, pp. 3025–37, doi:10.1149/07801.3025ecst.

Wilson, Zachary T., and Nikolaos V. Sahinidis. "The ALAMO Approach to Machine Learning." *Computers & Chemical Engineering*, vol. 106, Pergamon, Nov. 2017, pp. 785–95, doi:10.1016/J.COMPCHEMENG.2017.02.010.

Zaccaria, Valentina, et al. "A Real-Time Degradation Model for Hardware in the Loop." *Proceedings of ASME Turbo Expo 2015: Turbine Technical Conference and Exposition GT2015*, 2015, pp. 1–8.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50142-1

Modeling and Optimal Design of Pressure Swing Adsorber for Carbon Dioxide and Hydrogen Separation from Industrial Waste Gas

Toji Kakiuchi[a] , Tomoyuki Yajima[a] ,Nobuyuki Shigaki[b], Yoshiaki Kawajiri[a*]

[a]*Department of Materials Process Engineering, Nagoya University, Aichi 464-8603, JAPAN*
[b]*Steel Research Laboratory, JFE Steel Corporation, 1 Kokan-cho, Fukuyama, JAPAN*
kawajiri@nagoya-u.jp

Abstract

Recently, carbon dioxide capture utilization and storage (CCUS) is gaining attention as a way to reduce carbon dioxide emissions. Blast furnace gas (BFG) in the steel industry is one of the major sources of carbon dioxide (CO_2) emissions, which has the potential to be a hydrogen (H_2) source by water gas shift reaction to. In this work, we pursue simultaneous purification of H_2 and CO_2 from reformed BFG by pressure swing adsorption (PSA). A mathematical model of newly designed PSA for simultaneous separation of CO_2 and H_2 is developed from experimental results. After fitting to the experimental results to the model, we were able to reproduce the purity and recovery of the recovered gas with an error of about 5 %. The model was also used to perform multi-objective optimization of the PSA process. From the Pareto solutions, the relationship between energy consumption and throughput is analyzed. In addition, it was confirmed that there is an operation condition that can increase the purity of carbon dioxide and throughput while suppressing the decrease of purity and recovery of hydrogen.

Keywords: Pressure Swing Adsorption; Modeling; Dynamics; Optimization.

1. Introduction

Due to rising concentration of carbon dioxide in the atmosphere, global warming is posing a serious threat. To deal with this issue, Carbon Dioxide Capture, Storage and Utilization (CCUS), a technological concept to capture carbon dioxide from flue gas from power plants and steel mills, etc. to use it as a resource, or to inject it into the ground and seawater, has been attracting attention. It is estimated that CCUS could reduce 20% of worldwide carbon dioxide (CO_2) emission in 2008, which has substantial potential[1]. In particular, blast furnace gas in the steel industry is one of the major sources of carbon dioxide emissions, and the application of CCUS to this source is expected[2].

One of the promising techniques for capturing CO_2 is Pressure Swing Adsorption (PSA). PSA has been used in many applications of large-scale gas separation[3]. However, power consumption and energy cost must be reduced substantially for successful CCUS implementation[4]. A study reported that the estimated cost is $72-114 to capture and store a ton of carbon dioxide, most of which is spent to capture CO_2[5]. It is expected that this cost should be reduced to approximately $18 [3].

Blast furnace gas (BFG) contains not only CO_2 at high concentration, but also carbon monoxide and a small amount of hydrogen. BFG has the potential to be a hydrogen source by steam reforming. The production of methane and methanol, promising products of CO_2 utilization in CCUS, can be produced from carbon dioxide and hydrogen (H_2) included in the reformed BFG. To realize this CCUS approach, simultaneous capture of

CO_2 and H_2 from BFG after reforming is expected. However, the process of capturing both gases at the same time remains a challenge. For example, multiple PSA units designed by Air Products and Chemicals, Inc. may increase the total capital cost[6]. Other PSAs that separate these two components at the same time also tend to require high hydrogen concentrations in the feed gas, and only a few studies have considered dilute hydrogen [7].

This work pursues simultaneous purification of H_2 and CO_2 from reformed BFG by a novel PSA process. A mathematical model is developed from experimental results. Using the model, multi-objective optimization of the PSA process is performed, where the Pareto solutions provides insights into the relationship between energy consumption and throughput.

2. Process description

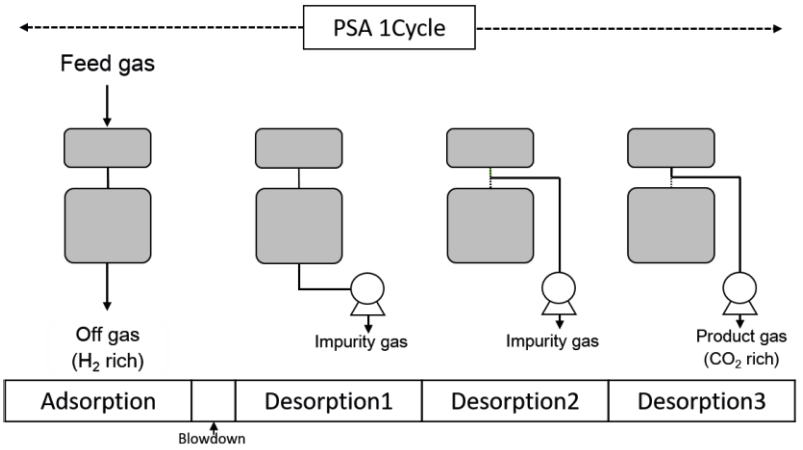


Figure 1 Operation of PSA for multiple separation

Figure 1 shows the operation of the newly designed PSA for CO_2 and H_2 multiple gas separation. The process consists of 2 tandem columns, which allows separation of two gas components within one cycle. Reformed BFG from a water gas shift reactor is supplied as the feed gas. In the Adsorption step, the off-gas is collected, which is rich in the component that adsorbs onto the adsorbent most weakly. In the following two steps, Desorption 1 and Desorption 2, a vacuum pump withdraws impurity gas, mainly N_2, from the upper and lower column, respectively. Finally, CO_2 is withdrawn from the upper column in the last desorption step through the gas fraction line, fractionated at the delivery side of the vacuum pump.

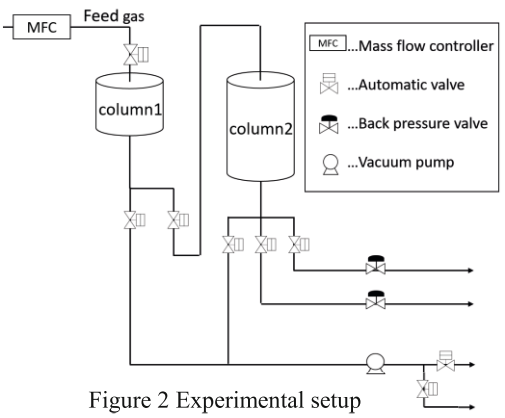


Figure 2 Experimental setup

To demonstrate this novel operation, a laboratory-scale PSA experiment was conducted. Figure 2 and Table 1 show the experimental conditions and setup

of this 2-column PSA. Zeolite 13X was employed as adsorbent in both columns. In this setup, the flow rate of the feed gas was controlled by mass flow controller. The adsorption pressure was controlled by a back pressure valve, and the desorption pressure was manipulated by a vacuum pump.

Table 1 Experimental condition

Condition	Value	Condition	Value
Adsorbent	Zeolite13X	Feed gas composition [%]	N_2:40,CO_2:37 H_2:23
Pellet diameter [mm]	1.5	Feed gas temperature [K]	298.15
Height of column [mm]	Upper:60 Lower:200	Cycle time [s/cycle]	100
Weight of adsorbent [g]	Upper:57 Lower:190	Pressure at the bottom of column [kPa]	Adsorption: 151 Desorption: 6
Column inner diameter [mm]	42	Feed gas flow rate [NL/min]	3.0

3. Mathematical model

A model was constructed that consists of mass and energy balance equations, Langmuir isotherms, and mass transfer rate assuming linear driving force[8]. Boundary conditions for each operation stage were also employed from the ones in a previous study[8].

3.1. Model fitting

To fit the model to the experimental data, we employ the formulation of Tikhonov regularization[9]. The model parameters were estimated using equation (1) as the objective function.

$$\min_{\theta} \sum_{j=1}^{3} \sum_{i=1}^{N_{Comp}} \left(Flow_{i,j,Model} - Flow_{i,j,exp}\right)^2 + \rho \sum_{i=1}^{N_{Comp}} \left(\frac{\theta_{opt,i} - \theta_{lit,i}}{\theta_{lit,i}}\right)^2 \quad (1)$$

N_{Comp} is the number of component, with N_{Comp} = 3; θ is the vector of parameters to be estimated; *Flow* [NL/min] is the flow rate from the column; ρ [-] is the regularization coefficient; subscripts i and j are for gas components and steps shown in Figure 1, respectively; *Model* and *exp* denote calculated and experimental value; and *opt* and *lit* denote optimized and literature parameter values, respectively. The isotherm equation is given by the Langmuir model, and the mass transfer between the gas and adsorbent phases is described by the linear driving force model:

$$q_i^* = \frac{q_{eq,i} K_i P_i}{1 + \sum_{i=1}^{n} K_i P_i} \tag{2}$$

$$K_i = K_{1_i} \exp\left(1000 K_{2_i}\left(\frac{1}{T} - \frac{1}{250}\right)\right) \tag{3}$$

$$q_i = \frac{15 D_e}{{R_p}^2}(q^* - q_i) \tag{4}$$

where $q_{eq,i}$ [mol/kg], q_i^* [mol/kg] and q_i [mol/kg] are saturation adsorption capacity, equilibrium adsorption amount, and adsorption amount, respectively; K_1 [1/Pa] and K_2 [K] are affinity constants; P_i [kPa] is partial pressure; De [m^2/s] is the diffusion coefficient; R_p [m] is the particle radius of adsorbent; and T [K] is temperature. The parameter vector θ is defined as defined as $\theta = [K_1, K_2, De]^T$.

3.2. Process optimization

Feed gas inflow rate and energy consumption are strongly associated with efficiency. The objective function for process optimization is formulated as in Equation (5) and (6).

$$\max_u \frac{\sum_{i=1}^{N_{Comp}} Feed_i}{t_{cy}(L_1 + L_2)} - M\varepsilon_1 - M\varepsilon_2 \tag{5}$$

where $Feed$ [mol/m^2] is the total molar volume of gas that enters the PSA; t_{cy} [s] is the cycle time; L_1 and L_2 [m] are the height of column1 and column2; ε_1 and ε_2 [-] are tolerance variables to enforce a cyclic steady state; M is the penalty constant set to 5000; u is a vector of decision variables, defined as $u = [P_{ad}, P_{de}, tcy, L_1, L_2, F]^T$; P_{ad} and P_{de} [kPa] are the pressure of adsorption and desorption step; F [NL/min] is the feed flow rate.

$$\min_u E = \frac{\frac{work_{ad}}{\eta_{blower}} + \frac{work_{de}}{\eta_{pump}}}{Product} \tag{6}$$

E [kJ/mol] is the energy consumption for a unit mole of recovered gas; $work_{ad}$ and $work_{de}$ [kJ] are the work of adsorption and desorption step; η_{blower} and η_{pump} [-] are the efficiency of blower and pump; $Product$ [mol] is the total molar volume of recovered CO_2 and H_2. Finally, the product purity and recovery must be at least $Pur_{min,i}$ and $Rec_{min,i}$, respectively:

$$Purity_i \geqq Pur_{min,i} \tag{7}$$

$$Recovery_i \geqq Rec_{min,i} \tag{8}$$

where i = H_2, CO_2. In this study, the optimization formulation and solution approach in Ko et al. [8] was employed.

4. Results and Discussion

4.1. Model fitting

Table 2 Experimental values and model fitting of purity and recovery

	$Purity_{CO2}$	$Recovery_{CO2}$	$Purity_{H2}$	$Recovery_{H2}$
Experiment	0.894	0.718	0.534	0.752
Model	0.823	0.758	0.532	0.758

Table 2 shows the model fitting for the experiment. It can be seen that both of the purity and recovery of CO_2 have an error of about 5 % compared to the experimental results, while hydrogen reproduces the experimental results well. While further reducing the value of ρ in Equation (1) would allow the parameter values to deviate from literature and reduce the model error, overfitting and parameter values that are physically inconsistent must be avoided. A potential reason for the model mismatch is the isotherm model in Equation (2), which has substantial influence on the desorption gas flow rate in Desorption1, 2 and 3.

4.2. Optimization Results

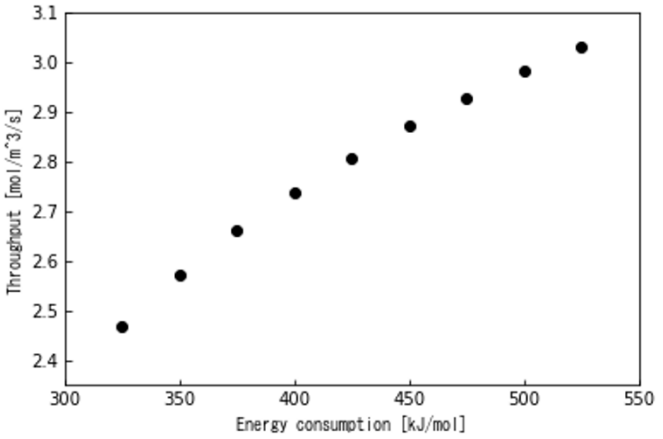


Figure 3 Pareto optimal solutions for energy consumption and throughput with $Pur_{min,CO2}$ = 0.99, $Rec_{min,CO2}$ = 0.90, $Pur_{min,H2}$ = 0.60, $Rec_{min,H2}$ = 0.70 in Equation (7) and (8).

Figure 3 shows the energy consumption and throughput, showing the trade-off between throughput and energy consumption for a unit mole of recovered CO_2 and H_2. All of the optimal solutions reach the lower bound of the constraints shown in Chapter 3. It can be confirmed that as the energy consumption increases, the impact on the throughput decreases. According to this graph, the higher energy consumption is, the smaller increment of throughput.

Comparing the experimental and optimized performance in Table 3, it is confirmed that better operating conditions can be found by the optimization, where the product constraints, Equation (7) and (8), are adjusted to the experimental data. In the optimal solution, the throughput is increased by approximately 20 %, and CO_2 purity and CO_2 recovery are both higher, while H_2 recovery and H_2 purity are same and the energy consumption is slightly lower.

Table 3 Comparison of experimental and optimized performance. The optimization was performed with the following conditions to match the experimental results: $Pur_{min,CO2} = 0.894$, $Rec_{min,CO2} = 0.718$, $Pur_{min,H2} = 0.534$, $Rec_{min,H2} = 0.752$ in Equation (7) and (8).

	experiment	optimization		experiment	optimization
Energy Consumption E [kJ/mol]	468	461	Throughput [mol/m^3/s]	3.064	3.65
$Purity_{CO2}$	0.894	0.983	$Recovery_{CO2}$	0.718	0.877
$Purity_{H2}$	0.534	0.534	$Recovery_{H2}$	0.752	0.752

5. Conclusions

In this study, we developed a mathematical model of the newly designed PSA for simultaneous gas separation of CO_2 and H_2, and optimized the process to improve the performance. The model was fitted to experimental results by Tikhonov regularization to consider literature data. Furthermore, the analysis by multi-objective optimization enabled trade-off analysis of process performance and identification of new operating conditions. Future study aims to reduce model error, and analyze the influence of constraints for product purity and recovery.

Acknowledgement

This article is based on results obtained from a project, JPNP16002, commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

References

1. International Energy Agency. (2008) Energy technology perspectives. Paris: International Energy Agency
2. Shigaki, N., Mogi, Y., Haraoka, T., and Sumi, I. (2018). Reduction of electric power consumption in CO2-PSA with zeolite 13X adsorbent. *Energies*, *11*(4), 1–21.
3. Saima H., Mogi Y., and Haraoka T.: JFE Tech. Rep. 2014, 19, 133.
4. Saima, H., Mogi, Y., and Haraoka, T. (2013). Development of PSA system for the recovery of carbon dioxide and carbon monoxide from blast furnace gas in steel works. *Energy Procedia*, *37*(19), 7152–7159.
5. Hasan, M. M. F., First, E. L., Boukouvala, F., and Floudas, C. A. (2015). A multi-scale framework for CO_2 capture, utilization, and sequestration: CCUS and CCU. *Computers and Chemical Engineering*, *81*, 2–21.
6. Sircar, S. (1993). Recent trends in pressure swing adsorption: Production of multiple products from a multicomponent feed gas. Gas Separation and Purification, 7(2), 69–73.
7. Streb, A., Hefti, M., Gazzani, M., and Mazzotti, M. (2019). Novel Adsorption Process for Co-Production of Hydrogen and CO_2 from a Multicomponent Stream. *Industrial and Engineering Chemistry Research*, *58*(37), 17489–17506.
8. Ko, D., Siriwardane, R., and Biegler, L. T. (2005). Optimization of pressure swing adsorption and fractionated vacuum pressure swing adsorption processes for CO_2 capture. *Industrial and Engineering Chemistry Research*, *44*(21), 8084–8094.
9. Suzuki, K., Harada, H., Sato, K., Okada, K., Tsuruta, M., Yajima, T., and Kawajiri, Y. (2021). Utilization of operation data for parameter estimation of simulated moving bed chromatography. *Journal of Advanced Manufacturing and Processing*, e10103

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50143-3

Membrane Characterization with Model-Based Design of Experiments

Xinhong Liu[a], Jialu Wang[a], Jonathan A. Ouimet[a], William A. Phillip[a], Alexander W. Dowling[a*]

[a] *Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN 46556, USA*
adowling@nd.edu

Abstract

Membrane characterization provides essential information for the scale-up, design, and optimization of new separation systems. We recently proposed the diafiltration apparatus for high-throughput analysis (DATA), which enables a 5-times reduction in the time, energy, and the number of experiments necessary to characterize membrane transport properties. This paper applies formal model-based design of experiments (MBDoE) techniques to further analyse and optimize DATA. For example, the eigenvalues and eigenvectors of the Fisher Information Matrix (FIM) show dynamic diafiltration experiments improve parameter identifiability by 3 orders of magnitude compared to traditional filtration experiments. Moreover, continuous retentate conductivity measurements in DATA improve A-, D-, E-, and ME-optimal MBDoE criteria by between 6 % and 32 %. Using these criteria, we identify pressure and initial concentrations conditions that maximize parameter precision and remove correlations.

Keywords: Membranes, Design of experiments, Parameter Estimation, Dynamic Modelling, Diafiltration

1. Introduction

Membrane processes have shown promise for addressing the critical needs for sustainability and energy efficiency. Recent material design to achieve separations of similar-sized molecules has evolved in the directions of precisely controlling the nanostructure of membranes and identifying chemical functionalities which accentuate desired transport properties (Hoffman and Phillip, 2020; Sadeghi et al., 2018). A detailed understanding of the underlying thermodynamic and transport phenomena can elucidate the molecular interactions and mechanisms that affect the macroscopic transport properties of the membrane (Geise et al., 2014; Yaroshchuk et al., 2018). Motivated by this need, the development of membrane characterization techniques that explore the dependency of membrane performance on feed conditions can greatly accelerate the development of materials (Ghosh et al., 2000). In addition, membrane characterization that elucidates underlying mechanisms provides essential information for scale-up, design, and optimization, facilitating the development of separations.

Design of Experiments (DoE) methods optimize computational and physical experiments to maximize the information gain and to minimize time and resource costs. Classical ‘black-box’ (a.k.a. factorial, response surface) DoE approaches, which decide the best design by the input-output relationship, does not (directly) incorporate membrane science knowledge; in contrast, model-based DoE (MBDoE) leverages high-fidelity models

constructed from underlying physical principles that describe the experimental system (Franceschini and Macchietto, 2008). The information collected from experiments can be applied to discriminate between scientific hypotheses, posed as mathematical models, and to improve the precision of parameter estimation. However, to date, MBDoE has not been applied to membrane characterization techniques.

Guided by data analytics, Ouimet et al. (2021) developed a diafiltration apparatus for high-throughput analysis (DATA) to address the limitations of current membrane characterization methods, e.g., time-consuming experimental campaigns and parameter non-identifiability. In this paper, we use MBDOE and FIM-based analysis to mathematically quantify the improvements reported by Ouimet et al. (2021) and further refine the experimental conditions needed in DATA to characterize membrane transport properties and discriminate between possible transport mechanisms.

2. Mathematical model, materials, and methods

In the dynamic diafiltration experiments described by Ouimet et al. (2021), a concentrated diafiltrate is continuously injected into a stirred cell under applied pressure, permeate is collected in several scintillation vials with the mass of the sample vial, m_v, permeate concentration, c_v, and retentate concentration in the stirred cell, c_f, measured. Using these measurements, three model parameters - hydraulic permeability, L_p, the solute permeability coefficient, B that correspond to the membrane transport properties, and the reflection coefficient, σ, that depends on the thermodynamics of the membrane-solution interface - are estimated via weighted least-square nonlinear regression (Eq. (1) where $\boldsymbol{\theta} = \{L_p, B, \sigma\}$). These parameters are related to the volumetric flux of water, J_w, and the molar flux of the solute, J_s, across the membrane in Eq. (2).

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_i w_{m_{v,i}} \left(m_{v,i} - \widehat{m}_{v,i}\right)^2 + \sum_j w_{c_{v,j}} \left(c_{v,j} - \hat{c}_{v,j}\right)^2 + \sum_k w_{c_{f,k}} \left(c_{f,k} - \hat{c}_{f,k}\right)^2 \tag{1}$$

$$J_w = L_p(\Delta P - \sigma \Delta \pi), \;\; J_s = B \Delta c \tag{2}$$

The diafiltration apparatus, the differential-algebraic equations (DAEs) model, the data, and the regressed parameters values, i.e., $L_p = 3.90$ L·m^{-2}·h^{-1}·bar^{-1}, $B = 0.29$ μm·s^{-1} and $\sigma = 1$ are described by Ouimet et al. (2021). Three key design decisions, the diafiltrate concentration, c_d, the initial feed concentration, $c_f(0)$, and the applied pressure, ΔP may be optimized to maximize the precision of the estimated parameters from dynamic diafiltration experiments.

3. Fisher Information Matrix (FIM)

The Fisher Information Matrix (FIM), $\mathbf{M}$, measures the information content of measurements and is defined as the inverse of the posterior covariance matrix $\mathbf{V}$, Eq. (4), ignoring the prior information (Franceschini and Macchietto, 2008). Here, $v_{m,rs}$ is the rsth element of the $N_y \times N_y$ inverse matrix of measurements error. $\mathbf{J}_r$ is the sensitivity matrix of output y_r sampled at times t_s and evaluated at nominal parameters values $\widehat{\boldsymbol{\theta}}$ and specified experimental design conditions $\boldsymbol{\phi}$.

$$\mathbf{M} = \left[\mathbf{V}\left(\hat{\boldsymbol{\theta}}, \boldsymbol{\phi}\right)\right]^{-1} = \sum_{r=1}^{N_y}\sum_{s=1}^{N_y} v_{m,rs}{}^{-1}\mathbf{J}_r^T\mathbf{J}_s\,, \qquad \mathbf{J}_r = \begin{bmatrix} \left.\frac{\partial y_r}{\partial \hat{\theta}_1}\right|_{t_1} & \cdots & \left.\frac{\partial y_r}{\partial \hat{\theta}_m}\right|_{t_1} \\ \vdots & \ddots & \vdots \\ \left.\frac{\partial y_r}{\partial \hat{\theta}_1}\right|_{t_n} & \cdots & \left.\frac{\partial y_r}{\partial \hat{\theta}_m}\right|_{t_n} \end{bmatrix} \tag{3}$$

MBDoE techniques increases parameter precision by minimising a metric of $\mathbf{V}$ or equivalently maximizing a metric of $\mathbf{M}$. A-, D-, E-optimal experimental designs correspond to minimising the trace, the determinant, and the maximum eigenvalue of $\mathbf{V}$, respectively (or maximizing the trace, the determinant, and the maximum eigenvalue of $\mathbf{M}$). The determinant and trace of the covariance matrix $\mathbf{V}$ can be interpreted as the volume of the covariance ellipsoid under feasible experimental conditions, while the maximum eigenvalue represents the size of the major axis, minimizing them reduces model parameter uncertainty. Additionally, the modified E-optimal (ME-) criterion minimizing the condition number of $\mathbf{M}$ which is defined as the ratio of the largest to the smallest eigenvalues, removes the correlation of parameters.

4. Results and discussions

4.1. Diafiltration experiment enables identification of all model parameters

Table 1 compares the FIMs and their eigen decompositions for experiments in both filtration (F) and diafiltration (D) modes as reported by Ouimet et al. (2021). The analysis of each mode considers one experiment with continuous data collection from the inline conductivity probe (M1) and one experiment encompassing only the initial and final retentate measurements (M2). The elements of the FIMs are one order of magnitude larger for diafiltration (D) than filtration (F) experiments. This shows diafiltration experiments contains more useful information to infer the model parameters. Moreover, analysing the eigenvalues and eigenvectors indicates which parameter can be precisely estimated through experimental design. For example, the minimum eigenvalue of filtration (F) M1 is 4.93E+05; the corresponding eigenvector is predominantly in the direction of model parameter σ. Under the same mode, the largest eigenvalue, 4.71E+09, corresponds the eigenvector in the direction of L_p. This difference, 4 orders of magnitude, indicates that a filtration experiment alone is unable to precisely estimate σ. In contrast, for diafiltration (D) mode, the eigenvalues whose corresponding eigenvectors in the direction of σ, 8.53E+10 in M1 and 8.18E+10 in M2, become the largest ones. Moreover, the smallest eigenvalues for diafiltration mode are 2.17E+08 (M1) and 1.96E+08 (M2), which are 3 orders of magnitude larger than the smallest eigenvalues for filtration mode. This difference indicates that diafiltration experiments are better suited to precisely estimate all three model parameters. Both findings are consistent with the sensitivity analysis results from Ouimet et al. (2021).

4.2. Additional retentate measurements improve parameter precision

Ouimet et al. (2021) show that measuring the retentate concentration is necessary to identify a converging set of parameters. We now use MBDoE to quantify the information content of the additional measurements. Recall M1 in Table 1 considers inline conductivity probe measurements for the retentate whereas M2 omits these measurements

and only considers initial and final retentate measurements. Table 1 shows elements and eigenvalues of FIMs of M1 are always larger than M2, which shows the additional data increases the precision of the estimated parameters for both modes. Furthermore, for diafiltration, Table 2 shows 6 % and 32 % reduction in terms of the volume of the confidence ellipsoid from A-, D-optimal criteria, respectively, 11 % reduction in terms of the uncertainty of the least confident parameter (B) from E-optimal, and 6% improvement in the ME-optimal criterion which measures parameter correlation. Similarly, Table 2 also shows 8 %, 17 %, and 2 % improvements from A-, D-, E-optimal criteria, respectively, in filtration experiments. However, the 6 % worsening of the ME-optimal criterion, indicates collecting additional data in filtration mode increases the correlation of the estimated parameters.

Table 1. FIM, eigenvalues and eigenvectors of FIM are calculated in both filtration (F) mode and diafiltration (D) mode. Model M1 includes inline conductivity probe measurements while models M2 includes only the initial and final retentate measurements.

Mode	Model	FIM (×1e9)			Eigenvalues	Eigenvectors		
		L_p	B	σ		L_p	B	σ
F	M1	4.67	-0.01	-0.40	4.93E+05	-8.57E-02	-8.50E-03	-9.96E-01
		-0.01	0.02	0.00	1.74E+07	1.80E-03	1.00E+00	-8.70E-03
		-0.40	0.00	0.04	4.71E+09	-9.96E-01	2.60E-03	8.57E-02
	M2	4.34	0.01	-0.37	4.85E+05	8.60E-02	3.30E-03	9.96E-01
		0.01	0.02	0.00	1.63E+07	2.00E-03	-1.00E+00	3.20E-03
		-0.37	0.00	0.03	4.37E+09	-9.96E-01	-1.70E-03	8.60E-02
D	M1	20.85	3.09	-20.62	2.17E+08	-1.46E-01	-9.47E-01	-2.88E-01
		3.09	5.56	-19.14	1.45E+10	9.41E-01	-2.23E-01	2.56E-01
		-20.62	-19.14	73.61	8.53E+10	-3.07E-01	-2.33E-01	9.23E-01
	M2	17.78	3.00	-18.36	1.96E+08	-1.41E-01	-9.47E-01	-2.89E-01
		3.00	5.57	-19.03	1.27E+10	9.50E-01	-2.11E-01	2.29E-01
		-18.36	-19.03	71.38	8.18E+10	-2.78E-01	-2.43E-01	9.29E-01

Table 2. DoE optimality criteria for models M1 and improvement of using M1 instead of M2.

Mode	Model	Functions of FIM			
		A-optimal	D-optimal	E-optimal	ME-optimal
		Trace	Determinant	Minimal eigenvalue	Condition number
F	M1	4.73E+09	4.04E+22	4.93E+05	9551
	Improvement	8%↑	17%↑	2%↑	6%↑
	M2	4.38E+09	3.45E+22	4.85E+05	9002
D	M1	1.00E+11	2.69E+29	2.17E+08	393
	Improvement	6%↑	32%↑	11%↑	6%↓
	M2	9.47E+10	2.03E+29	1.96E+08	418

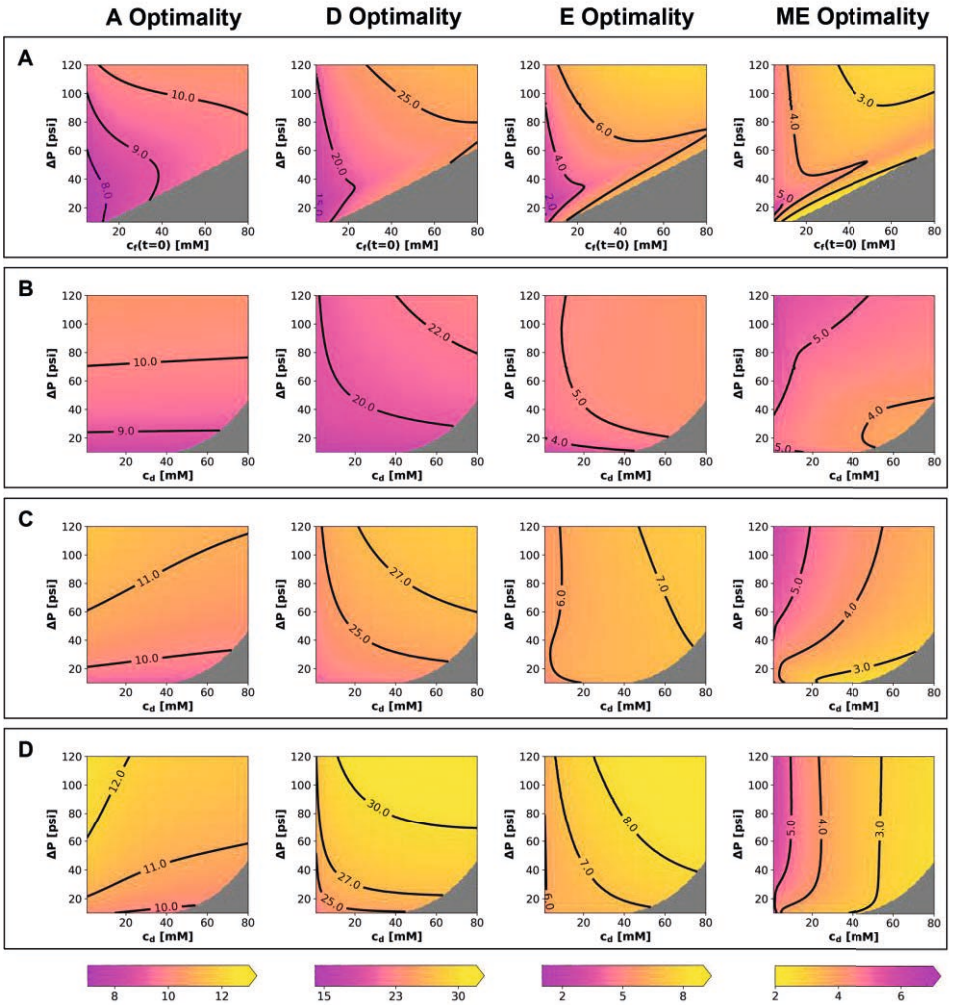


Fig. 1. A-, D-, E-, ME-optimal criteria evaluated under varying experimental conditions. Panel A shows filtration experiment (with 8 vials collected) predictions for varying initial feed concentration and applied pressure. Panels B, C, D examine the diafiltration experiment with 1, 5 and 10 vials collected, respectively, for the diafiltrate concentration and applied pressure.

4.3. MBDoE optimizes DATA system

We now use A-, D-, E-, and ME-optimality criteria to inform the applied pressure, initial retentate or diafiltrate concentrations (experimental design decisions) necessary to identify all parameters in filtration and diafiltration. Fig. 2A. examines filtration experiment at varying initial feed concentration $c_f(t=0)$ and applied pressure ΔP with 8 vials collected. Fig. 2B, 2C, and 2D examine diafiltration experiments at varying diafiltrate concentration c_d with 1, 5, and 10 vial collected, respectively. The gray regions correspond to physically impossible operating conditions where the water flux is equal to or less than zero. The contour lines show the $\log_{10}$-transformed values of every criterion. Comparing Fig. 2C to 2A, the lighter color and larger contour values for A-, D-, E-optimality metrics indicates that the diafiltration experiments with 5 vial collections

contains more information than the filtration experiment with 8 vial collections. Moreover, higher applied pressures maximize A-, D-, and E-optimal metrics. However, based on ME-optimality, low applied pressure is desired in diafiltration experiment with 5 or fewer vial collections to remove the correlation among parameters. Increasing to 10 vial collections in diafiltration, shown in Fig. 2D, resolves the trade-off between parameter precision (A and D) and removing correlations (ME). Thus, with 10 vial collections, diafiltration experiments with a feed concentration of 5 mM KCl should be performed with a diafiltrate concentration greater than 50 mM KCl and an applied pressure at least 45 psi to identify all parameters with an order of magnitude of improvement in precision over filtration experiments.

5. Conclusions

In this paper, we apply MBDoE analyses to quantify the information gain in a recently proposed diafiltration apparatus for high-throughput analysis (DATA) for membrane characterization. In the future, MBDoE can be used to discriminate possible phenomena and mechanisms within complex multi-component systems and optimize diafiltration experiments with more degrees of freedom (e.g., time-varying applied pressure).

Acknowledgements

We appreciatively acknowledge support from the National Science Foundation (NSF) through the Advanced Manufacturing Program (Award Number: 1932206) and the CAREER Program (Award Number: CBET-1941596). J.W. gratefully acknowledges support from the Carbon Capture Simulation for Industry Impact ($CCSI^2$), funded through the U.S. DOE office of Fossil Energy by the Lawrence Berkeley National Laboratory through contract# DE-AC02-05CH11231.

References

G. Franceschini and S. Macchietto, 2008, Model-based design of experiments for parameter precision: State of the art, Chemical engineering science, 63, 19, 4846-4872, 10.1016/j.ces.2007.11.034

G.M. Geise, D.R. Paul and B.D. Freeman, 2014, Fundamental water and salt transport properties of polymeric materials, Progress in polymer science, 39, 1, 1-42, 10.1016/j.progpolymsci.2013.07.001

R. Ghosh and Z. Cui, 2000, Analysis of protein transport and polarization through membranes using pulsed sample injection technique, Journal of membrane science, 175, 1, 75-84, 10.1016/s0376-7388(00)00397-5

J.R. Hoffman and W.A. Phillip, 2020, 100th anniversary of macromolecular science viewpoint: Integrated membrane systems, ACS Macro Lett, 9, 9, 1267-1279, 10.1021/acsmacrolett.0c00482

J.A. Ouimet, X. Liu, D.J. Brown, E.A. Eugene, T. Popps, Z.W. Muetzel, A.W. Dowling and W.A. Phillip, Data: Diafiltration apparatus for high-throughput analysis, Journal of membrane science, 10.1016/j.memsci.2021.119743

I. Sadeghi, P. Kaner and A. Asatekin, 2018, Controlling and expanding the selectivity of filtration membranes, Chem. Mater, 30, 21, 7328-7354, 10.1021/acs.chemmater.8b03334

A. Yaroshchuk, M.L. Bruening and E. Zholkovskiy, 2019, Modelling nanofiltration of electrolyte solutions, Adv Colloid Interface Sci, 268, C, 39-63, 10.1016/j.cis.2019.03.004

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50144-5

Systematic Modelling of Distillation Columns based on Topologies and Ontologies

Robert Pujan[a,b*], Philipp Sengupta[a], Heinz A. Preisig[b]

[a]*DBFZ Deutsches Biomasseforschungszentrum gemeinnützige GmbH, Torgauer Straße 116, 04347 Leipzig, Germany*
[b]*NTNU Norwegian University of Science and Technology, Høgskoleringen 5, 7491 Trondheim, Norway*
robert.pujan@dbfz.de

Abstract

This contribution proposes a systematic modelling approach for tray distillation columns. Abstracting the distillation column into a network of the underlying, essential processes introduces a model topology. The topology is equipped with a centrally-established ontology that is a collection of fundamental principles and definitions describing the processes' nature. This systematic approach minimises modelling errors, enables rapid model design, and significantly lowers the entry threshold for non-expert modellers.

This study provides a holistic model that is, in contrast to common distillation models, applicable to steady-state as well as dynamic process conditions. The model's design and mathematical formalism are described, followed by a discussion of the model capabilities including model fitting and evaluation of the simulation results against experimental data as well as a commercial flowsheeting package.

Keywords: process modelling, simulation, topology, ontology, methodology

1. Motivation

Distillation is probably the most widespread industrial process for the separation of liquid mixtures. Accordingly, design, optimisation and process control are required for the exploration and testing of a myriad of various, differing process scales, operational domains and mixtures. Experimental studies may have been the conventional approach in process engineering for the better part of the 20th century, but they are expensive, time-consuming and limited in their viability. Process simulations, on the other hand, are executed in computers and thus not limited by physical constraints. The fast development of computing rapidly enhanced the capabilities of numerical simulation, thus moving the centre of process engineering progressively away from lab facilities towards modelling suites. The increasing need for model implementations places a high demand on the modellers, thereby prompting a growing interest in more effective model generation methods that incorporate the expertise from different scientific domains like physics, chemistry, biology and engineering. Providing the modeller with customisable compound models equipped with multi-disciplinary expert knowledge is thus a critical undertaking for process engineering in general and distillation applications in particular.

Applying the modelling method established by Preisig (2014), this study proposes a systematic modelling approach for distillation model design. First, an exemplary tray distillation process is graphically abstracted through a topology. Due to the established links between the topology and a physics ontology, the modeller-specified capacities

inherently provide the model's fundamental mathematical abstraction. This study showcases the capabilities and reliability of the resulting compound model by validating the simulation results against experimental data as well as the results of commercially available flowsheeting software.

2. Structural model design

The processes we observe are intrinsically composed of various sub-processes on multiple scales. Modelling has thus to include both, the identification of the physical processes involved, and the adaptation of the mathematical formalisms to the systems' nature and application.

Topologies appear to be particularly convenient for model design since they are easily translated into matrix algebra and applied to multiple scales. A topology is the graphical abstraction of a process as a graph network of nodes representing finite-volume capacities and arcs that are flows of extensive quantities driven by gradients in intensive properties. Our methodology limits the topology design to basic entities, provides a maths description for each entity, and associates the entities to the graphical objects used to establish the topologies. This approach enables the rapid design of rigorous, holistic models that are also easily customisable and minimised in the amount of potential, modeller-caused errors. Furthermore, the topology structure enables the process depiction in both, steady-state as well as dynamic operation.

Figure 1 presents the topology for tray distillation columns that was applied in this study. For an in-depth description of topology symbols and terminology, please see Pujan & Preisig (2022). A column outfitted with n trays is fed a fluid mixture by the reservoir F on feed tray n_F. On every tray, part of the stage's liquid phase L_i is dripping down to tray $i+1$ while other parts evaporate into the gas phase G_i, subsequently rising up to tray $i-1$. While ascending, the vapour passes both an empty column space S_n as well as the cavities of the tray T_i. The condenser C at the column's head cools down the arriving vapour G_0, forming the condensate L_0 that is, according to the specified reflux ratio, partially dripping back to the column's first tray and partially drained as distillate D. The column bottom is heated by H, thus evaporating parts of the boiler's liquid L_{n+1}, while the bottom product B is continuously drained at a certain rate. In steady-state (on the left of Figure 1), each tray's liquid and vapour capacities are in thermodynamic equilibrium and without any time-dependencies, thus depicted as point capacities (black dots). Under dynamic conditions, for example during start-up and shut-down, these volumes change over time, both in extensive quantities as well as intensive properties. If considering distributional effects in them to be negligible, the phases appear as lumped capacities (circles) in terms of mass and heat distribution (central topology of Figure 1).

As described in detail by Pujan & Preisig (2020), the pressure distribution in dynamic processes can only be assessed employing a model split. This split results in a dynamic model for the observable mass and heat flows and an event-dynamic model (on the right of Figure 1) for the supposedly immediate pressure distribution.

The graphical topology structure is accessed by mathematical modelling through the so-called incidence matrix $\underline{\mathrm{F}}$ that is a structural matrix that can be sub-divided into $\underline{\underline{\mathrm{F}}}_m$, $\underline{\underline{\mathrm{F}}}_q$, and $\underline{\underline{\mathrm{F}}}_w$ specific for mass, heat and work transport (Pujan & Preisig, 2022).

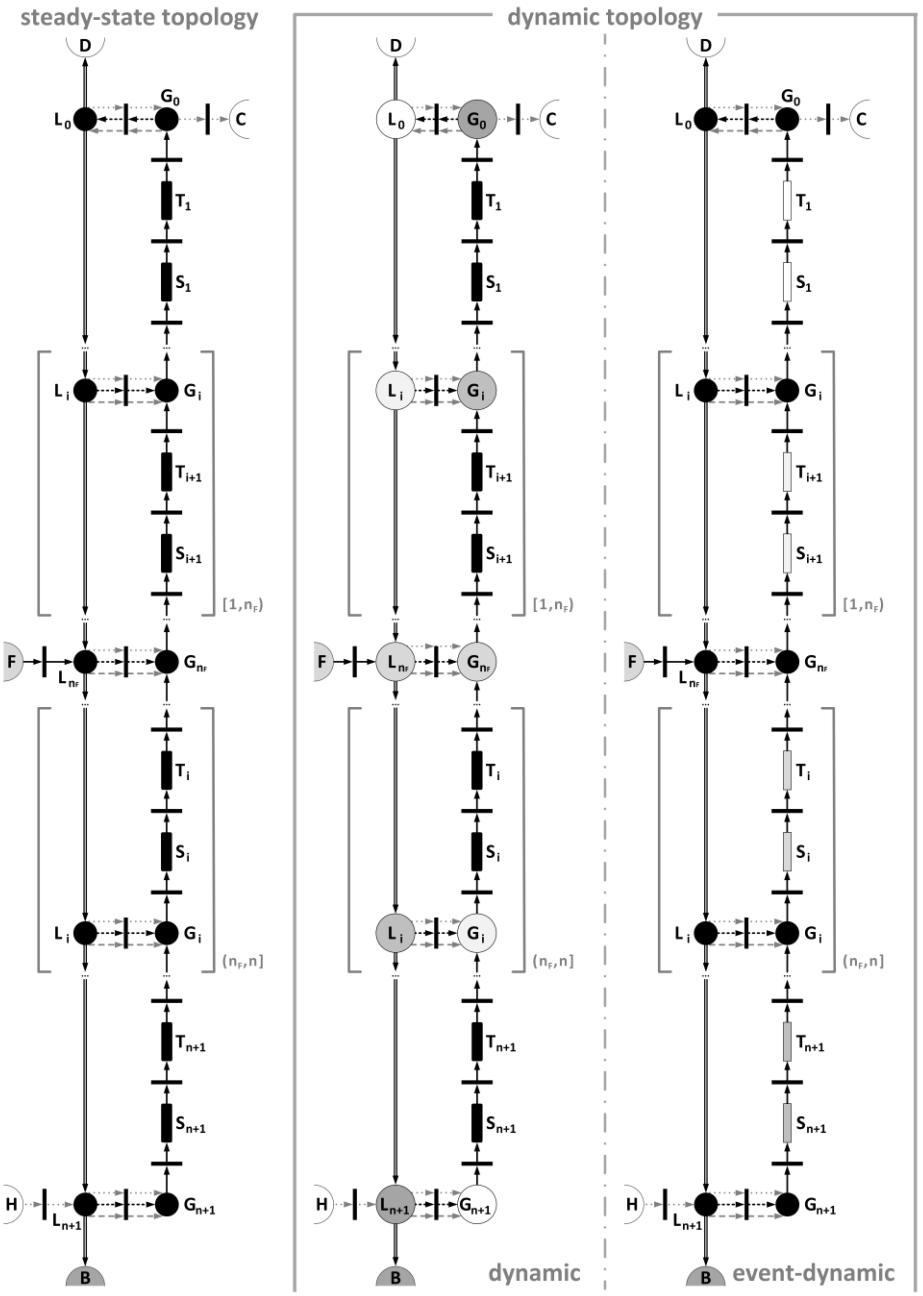


Figure 1: Topologies for a tray distillation column in steady-state (left) and dynamic process conditions (middle: dynamic mass and heat distribution; right: event-dynamic pressure distribution)

3. Mathematical model design

The model's mathematical behaviour is extracted from a centrally established equation ontology. The term ontology encompasses the entirety of fundamental concepts, definitions and relations taken from the application-specific scientific roots. Ontologies are a means of abstraction and concentrating information, promoting a more systematic approach to model design (Preisig, 2012).

Table 1 shows the fundamental equation ontology for the physical domain of tray distillation columns. Vectors and matrices are depicted as $\underline{x}$ and $\underline{\underline{x}}$, respectively. The decorator $\dot{x}$ represents an accumulation and $\hat{x}$ a flow. The grey-marked variables have to be specified further for the model initiation. These specifications typically are:

- assumptions and preset values, like the nullification of the system's change in kinetic and potential energy $\dot{K}$ and $\dot{P}$, and the gravitational acceleration g,
- initial conditions x_0,
- plant dimensions such as liquid and weir heights h_L and h_W, diameters d, weir length s_W, and weir drain coefficient C_W,
- thermodynamic values from databases or external estimators, like the specific enthalpy h_i, density ρ_i, molar mass M_i, and transport coefficients for heat k_i^q and mass k_i^d for substance i,
- and further ontology equations, e.g. for temperature T, volume-specific interphase area a, and frictional pressure drop Δp^f.

4. Results and discussion

The model was initiated for the distillation of the binary mixture water-methanol and the steady-state results were validated against the experimental data by Kazameas et al. (2015) and Morinaga & Yao (1965), as well as simulation results derived from the flowsheeting package Aspen Plus®. Unfortunately, at the time of this publication, no sufficient data set was available for the validation of the model's dynamic capabilities.

Below the feed tray (feed stage marked with a circle), the initial simulation results in Figure 2 show a notably increased deviation from the experimentally determined molar fractions. Whereas the flowsheeting results seem to better match the experiments below the feed, our model seems more appropriate above. This performance was significantly improved through a simple model fitting, including only thermodynamic estimation parameters and the Murphree tray efficiency.

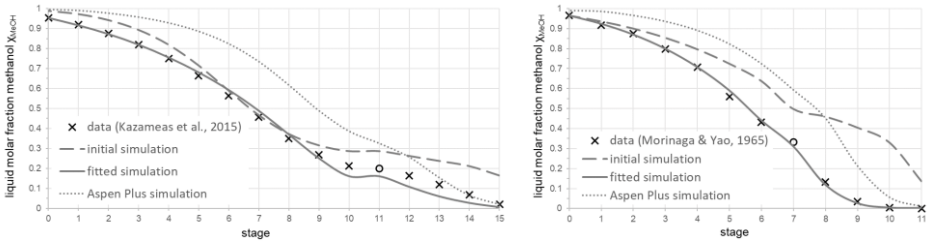


Figure 2: Comparing simulation results to Kazameas et al. (2015) and Morinaga & Yao (1965)

Table 1: Physics equation ontology for tray distillation columns

Integrals		
Substance	$\underline{n} = \int_0^{\tau} \underline{\dot{n}}_t \mathrm{dt} + \underline{n}_0$	(1)
Energy	$\underline{E} = \int_0^{\tau} \underline{\dot{E}}_t \mathrm{dt} + \underline{E}_0$	(2)
Balances		
Substance conservation	$\underline{\dot{n}} = \underline{\underline{\mathrm{F}_m}}\, \underline{\hat{n}}$	(3)
Energy conservation	$\underline{\dot{E}} = \underline{\dot{U}} + \underline{\dot{K}} + \underline{\dot{P}}$	(4)
Internal energy conservation	$\underline{\dot{U}} = \underline{\underline{\mathrm{F}_m}}\left[\underline{\hat{U}} + \underline{\hat{K}} + \underline{\hat{P}} + \underline{\hat{w}}^v + \underline{\hat{w}}^f\right] + \underline{\underline{\mathrm{F}_w}}\, \underline{\hat{w}} + \underline{\underline{\mathrm{F}_q}}\, \underline{\hat{q}}$	(5)
Transport terms		
Mass flow	$\underline{\hat{m}} = \sum \underline{\hat{m}}_i$	(6)
Mass component flow	$\underline{\hat{m}}_i = \underline{\underline{M_i}}\, \underline{\hat{n}}_i$	(7)
Substance flow	$\underline{\hat{n}} = \sum \underline{\hat{n}}_i$	(8)
Molar flow of i (convection)	$\underline{\hat{n}}_i = \underline{c_i}\, \underline{\hat{V}}$	(9)
Molar flow of i (diffusion)	$\underline{\hat{n}}_i = -\underline{k_i^d}\, \underline{a}\, \underline{\underline{\mathrm{F}_m^{\mathrm{T}}}}\, \underline{c_i}$	(10)
Internal energy flow	$\underline{\hat{U}} = \underline{\hat{H}} - \underline{\hat{w}}^v$	(11)
Kinetic energy flow	$\underline{\hat{K}} = {}^1/_2\, \underline{\underline{\hat{m}}}\, \underline{v}^2$	(12)
Potential energy flow	$\underline{\hat{P}} = \underline{h}\, \underline{\underline{\hat{m}}}\, g$	(13)
Volume work flow	$\underline{\hat{w}}^v = \underline{p}\, \underline{\hat{V}}$	(14)
Friction work flow	$\underline{\hat{w}}^f = \underline{\underline{\Delta p^f}}\, \underline{\hat{V}}$	(15)
System volumetric work	$\underline{\hat{w}} = \underline{p}\, \underline{\dot{V}}$	(16)
Heat flow (conduction)	$\underline{\hat{q}} = -\underline{\underline{k^q}}\, \underline{\underline{A}}\, \underline{\underline{s}}^{-1}\, \underline{\underline{\mathrm{F}_q^{\mathrm{T}}}}\, \underline{T}$	(17)
Enthalpy flow	$\underline{\hat{H}} = \sum \underline{\hat{H}}_i$	(18)
Enthalpy component flow	$\underline{\hat{H}}_i = \underline{h}_i\, \underline{\hat{m}}_i$	(19)
Volume flow	$\underline{\hat{V}} = \underline{\underline{A}}\, \underline{v}$	(20)
Volume flow (weir)	$\underline{\hat{V}} = {}^2/_3\, \underline{C_W}\, \underline{s_W} \sqrt{2g}\, \left(\underline{h_L} - \underline{h_W}\right)^{1.5}$	(21)
State variable transformations		
Cylindrical cross-section	$\underline{A} = {}^{\pi}/_4\, \underline{d}^2$	(22)
Molar concentration	$\underline{c_i} = \underline{\underline{n_i}}\, \underline{V}^{-1}$	(23)
Substance component	$\underline{\chi_i} = \underline{\underline{n_i}}\, \underline{n}^{-1}$	(24)
Mass fraction	$\underline{\omega_i} = \underline{\underline{m_i}}\, \underline{m}^{-1}$	(25)

Figure 3 compares the simulation results to the temperature profile recorded by Kazameas et al. (2015). The model shows a better (unfitted) performance than Aspen Plus®, however, the significant different shapes of experimental and simulation results hypothesise measurement errors in the experiment. As the dashed curve indicates, assuming constant pressure does not perceptibly alter the model's temperature profile under steady-state conditions.

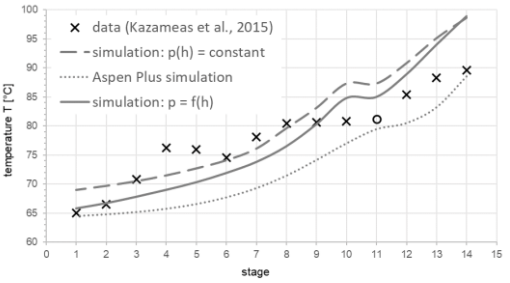


Figure 3: Stage temperatures T compared to Kazameas et al. (2015)

5. Conclusions

As this paper shows, ontology-imposed topology model designs are a fast and reliable method for modelling distillation columns. The presented model will be implemented in the compound model library of the modelling suite ProMo, which is currently in development at the NTNU (Elve & Preisig, 2019; Preisig, 2020). The software automatically selects the required ontology equations and generates executable program code, thus enabling rapid custom distillation model design. Future work will add models and ontology equations specific for packed and divided-wall distillation columns as well.

Acknowledgements

This study was created as part of the research projects Bio4Fuels (Norwegian Centre for Environment-friendly Energy Research (FME), project 257622), MarketPlace (Horizon 2020, project 760173), and VIPCOAT (Horizon 2020, project 952903).

References

A. Elve & H.A. Preisig, 2019, From ontology to executable program code, Computers & Chemical Engineering, 122, 383-394

C.G. Kazameas, K.N. Keller & W.L. Luyben, 2015, A Comprehensive Real-World Distillation Experiment, Chemical Engineering Education, 49 (3), 131-140

T. Morinaga & H. Yao, 1965, Regelung einer Destillationskolonne für ein binäres Gemisch, at-Automatisierungstechnik, 13 (1-12), 543-546

H.A. Preisig, 2012, Thinking Ontologies, Computer Aided Chemical Engineering, 31, 1682-1686

H.A. Preisig, 2014, Visual Modelling, Computer Aided Chemical Engineering, 34, 729-734

H.A. Preisig, 2020, ProMo – A multi-disciplinary process modelling suite, Computer Aided Chemical Engineering, 48, 571-576

R. Pujan & H.A. Preisig, 2020, Systematic Modelling of Flow and Pressure Distribution in a Complex Tank, Computer Aided Chemical Engineering, 48, 1945-1950

R. Pujan & H.A. Preisig, 2022, Systematic Modelling of Flow and Pressure Distribution in a Complex Tank, Computers & Chemical Engineering, 157, 107608

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50145-7

Sensitivity Analysis of an Electrospray Dehumidification System

David Young[a], Yasuhiro Shoji[a], Maliha Yel Mahi[b], Mike Ellis[c], Selen Cremaschi[a*], Lorenzo Cremaschi[b]

[a]*Department of Chemical Engineering, Auburn University, Auburn, AL 36849, USA*
[b]*Department of Mechanical Engineering, Auburn University, Auburn, AL 36849, USA*
[c]*Advanced Cooling Technologies, Inc. (ACT), Lancaster, PA 17601, USA*
selen-cremaschi@auburn.edu

Abstract

Electrically enhanced condensation achieved by an electrospray system can enhance dehumidification in HVAC applications if successfully scaled up. The first step towards scale-up is identifying critical design and operating parameters that significantly impact dehumidification for the electrospray system. Sensitivity analysis is one approach for identifying critical parameters. This paper presents a methodology that can efficiently perform quantitative sensitivity analysis on a computationally expensive high fidelity simulation model, which is a computational fluid dynamics (CFD) model of the electrospray system. The methodology employs surrogate modeling and Sobol sensitivity analysis. We estimate the main and total effect Sobol sensitivity indices of four operating parameters of the electrospray system. The results reveal that the most important parameter is the size of the sprayed droplets, followed by the volumetric flow rate of the bulk air and injected water in the system. The water injection velocity of the spray was found to have little to no impact on dehumidification.

Keywords: CFD; Sensitivity Analysis; Surrogate-modelling; ALAMO; PRESTO

1. Introduction

Dielectrophoresis is a well-documented phenomenon, with the first study coining the name in 1951 (Pohl, 1951). The phenomenon refers to the force imparted on a dielectric particle from a nonhomogenous electric field. Dielectrophoresis has been used in various applications, e.g., from cell separation (Henslee et al., 2011) to nanowire assembly (Collet et al., 2015). One more promising application is in heating, ventilation, and air conditioning (HVAC) systems. For an HVAC application, water condensation in the air can be enhanced by dielectrophoresis by spraying electrically charged droplets into the air, reducing the energy consumption of conventional HVAC systems.

Electrically enhanced condensation uses highly charged water droplets injected in the moist air. The droplets become electric seeds that attract polar water vapor molecules to their surfaces and promote condensation. The growth of the charged droplets depletes the vapor phase near a droplet, while dielectrophoresis flow and diffusion compensate for this depletion. Dielectrophoresis flow involves surrounding vapor at a distance of about 10 to 100 nm for droplets charged by an electrospray compared to ~2 nm for a single electron charge in a droplet. As vapor molecules collapse on the surface of the droplets, their initial electrical charge decreases due to the neutralization of the ions.

An electrically enhanced dehumidification system thus far has only been demonstrated in laboratory-scale studies (Higashiyama and Kamada, 2017). One challenge that hinders its large-scale application is the lack of knowledge on critical design and operating parameters that significantly impact the dehumidification rate. This work addresses this challenge by efficiently developing an approach to identify the critical parameters. The approach utilizes a CFD model of the electrospray system within a duct, surrogate modeling, and Sobol sensitivity analysis. The CFD model includes a particle-laden flow, where the dielectrophoresis flow is modeled within the mass transfer between the two phases (Section 2). We develop a surrogate model that links the dehumidification rate to select CFD model parameters (Section 3) and perform Sobol sensitivity analysis (Section 3) on the surrogate model to identify the critical parameters (Section 4).

2. Electrospray System Computational Fluid Dynamics (CFD) Model

The CFD model was built using open-source CFD simulation software OpenFOAM v8. We modified the sprayFoam solver to add an electro-condensation model, which includes the thermal and transport equations for water vapor dielectrophoresis condensation at the interface of electrically charged liquid droplets. SprayFoam is a numerical solver that takes the Eulerian-Lagrangian approach to simulate dispersed particle-laden flow. The Euler method was used to model the continuous phase, bulk airflow as dry air and water vapor mixture. The Lagrangian approach was used to model the dispersed phase, the electrically charged water droplets from the spray. The water vapor was modeled as a component of the bulk air, transported by both diffusion and convection. The governing equations for the moist airflow were the conservation of continuity, momentum, species, energy, and the equation of state in the laminar flow regime. Droplets in the spray were subjected to several physical phenomena, including coalescence and drag effects. Two-way coupling was introduced in the numerical study to establish the interaction between the two phases (air and water droplet). Source terms calculated in the sub-models using the Lagrangian approach were introduced into gas phase equations in the Eulerian approach to represent the interaction between Euler and Lagrangian methods. The water molecules placed in a gradient electric field experienced a force that moved them toward the charged droplets.

2.1. *Particle-Laden Gas Using Eulerian Method*

Eqn. 1 reflects the conservation of mass for the gas phase,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho U) = \dot{\rho}_s \tag{1}$$

where ρ is the density of air, t is time, U is the velocity, and $\dot{\rho}_s$ is a source term defined only for the water vapor. It is calculated using the water vapor condensation into liquid droplets. Mass transport for individual species in the gas phase is given in Eqs. 2 - 4,

$$\frac{\partial \rho Y_i}{\partial t} = \nabla \cdot (\mu_L \nabla Y_i) = \dot{\rho}_s^i \tag{2}$$

$$\dot{\rho}_s = \sum_i \dot{\rho}_s^i \tag{3}$$

where, Y_i is the mass fraction for the i-th gas species ($i = H_2O\ , O_2, N_2$) in the bulk air, and μ_L is the laminar viscosity.

$$\frac{\partial \rho U}{\partial t} + \nabla \cdot (\rho UU) = -\nabla P + \nabla \cdot (\mu_L \nabla \mathrm{U}) + \nabla \cdot [\mathrm{dev}(\mu_L (\nabla U)^T)] + \rho g + F_S \quad (4)$$

In Eq. 4, ∇P is the pressure gradient, g is the acceleration due to gravity, F_S is the the resultant force induced by the charged water droplets due to drag force and gravity force.

2.2. *Spray Charged Droplets Using Lagrangian Method*

The dispersed charged spray droplets were solved using the Lagrangian particle tracking method. The approach assumes that spray injection consists of different droplet parcels, and each parcel consists of many spherical droplets sharing the same location, diameter, velocity, and temperature. Newton's second law describes the charged droplet motion. The driving forces for the charged droplets are drag force from the surrounding air, the electrostatic force, and gravity force. These forces link droplet mass (m_D) and velocity (U_{inj}) to droplet density (ρ_d), and diameter (d_{inj}).

2.3. *Droplet Electro-Condensation Model*

With the dielectrophoresis effects, the saturated vapor pressure can be evaluated by the modified Kelvin-Thomson (MKT) model as described in Eqn. (1) of Morcelli and Cremaschi (2021). This equation provides the oversaturation pressure ratio, (P_{sat}/P_{sat}^{flat}) where, P_{sat} is the saturated vapor pressure near the surface of the droplet and P_{sat}^{flat} is the saturated water vapor pressure above a flat surface at temperature T. The oversaturation ratio is calculated, as the electric charge on the droplet increases the differences between the vapor pressure of the water droplet and bulk air. If the bulk air vapor pressure is higher, droplets grow. For the present study, the temperature T was constant at $20°C$ for both moist airflow and droplets. For the MKT model, the droplets were kept fixed to the maximum electric charge (q_e), dictated by the Rayleigh limit, Eq.5,

$$q_e = 8\pi\sqrt{\varepsilon_0 \sigma R^3} \quad (5)$$

Where R is the instantaneous water droplet radius. The oversaturation ratio, was determined from MKT model and a mass transfer from the water vapor to the droplet surface occurred. A prerequisite condition of oversaturation ratio less than 1 was introduced to limit the droplet growth via diffusion for realistic cases. The dielectrophoretic force sustain the vapor flow towards the droplet surface even if the diffusion transport diminishes due to the equality of the pressures near the droplet surface and far away from the droplet. The mass transfer due to electrically enhanced condensation is calculated using Eq. 6.

$$\frac{dm_D}{\mathrm{dt}} = -4\pi R^2 \frac{dR}{dt} \rho_d \quad (6)$$

Here $\frac{dR}{dt}$ is the constant droplet growth rate, estimated using data from Abe and Kameda (2003) and Morcelli and Cremaschi (2021).

3. Surrogate-based Sobol Sensitivity Analysis

Sensitivity analysis (SA) identifies parameters that significantly impact the output of interest for a model. We employed Sobol SA (Sobol, 2001) to investigate the sensitivity of the dehumidification rate in the duct, estimated by the CFD model, to select operating

parameters of the electrospray system. Sobol SA is a global method that quantifies the effect of each variable and its interactions with other variables on the model output by decomposing the output variance. These effects are quantified using main-effect and total Sobol indices and show the true sensitivity of the outputs with a high degree of accuracy. However, obtaining stable Sobol index estimates may be computationally expensive as it may require a large number of model evaluations (Tian, 2013). A surrogate model-based Sobol SA is proposed to overcome this issue, considering a single run of the CFD simulation for electrospray system in the duct takes, on average, seven days (wall time) on an Intel E5-2660, 20 core 2.6 GHz node.

The surrogate-based Sobol SA is performed to determine the impact of four operating parameters on dehumidification. The parameters are injected water flowrate ($\dot{V}_{inj}$), the injected water droplet size (d_{inj}), the injected water velocity (U_{inj}), and the bulk air flowrate ($\dot{V}_{air}$). The potential operating ranges for these parameters are listed in Table 1.

Table 1 Potential operating ranges for the four parameters considered for sensitivity analysis

Parameter	Upper bound	Lower bound	Unit
U_{inj}	2.73×10^{-2}	2.73×10^{-1}	[m/s]
d_{inj}	15	100	[m]
$\dot{V}_{inj}$	0.1	50	[μL/min]
$\dot{V}_{air}$	0.5	5	[ft^3/min]

The surrogate-based SA starts by running the CFD simulation at selected parameter values to compute the corresponding output, dehumidification. The parameter values for the CFD runs are selected according to Saltelli's (2002) method. We collected 78 unique input/output pairs by running the CFD simulation.

Next, an appropriate surrogate modeling technique is selected to represent the input-output relationship. Previous analysis revealed that the data characteristics and the application (surface approximation or optimization) are essential for selecting the correct surrogate modeling technique (Williams and Cremaschi, 2021). To determine the correct modeling technique, we employed PRESTO, Predictive REcommendations of Surrogate models To Optimize (Williams et al., 2022), which recommended automated learning of algebraic models (ALAMO) (Cozad et al., 2014) as the technique to best represent the data out of eight potential surrogate modeling approaches. We trained a surrogate model using ALAMO and the input-output pairs and employed it to estimate the Sobol indices following Saltelli's (2002) method.

4. Results

The maximum number of terms allowed during training a surrogate model should be determined *a priori* when using ALAMO. The allowed basis functions were $sin(x)$, $cos(x)$, $exp(x)$, $\ln(x)$, Gaussian radial basis function, polynomials up to third power, and pairwise combination of variables, both linear and quadratic. We employed 4-fold cross-validation (Burman, 1989) with 30 Monte Carlo replications using a 75 % to 25 % training to testing split to determine the maximum number of terms. Fig. 1 (A) includes a plot of the root mean square error (RMSE) for the training and test data sets vs. the maximum number of terms. We used 20 terms to train the ALAMO model used in the SA based on the beginning of an upward trend of the testing RMSE. The dehumidification

rate predictions of the ALAMO model are plotted against the values calculated by the CDF model in Fig. 1(B). The plot also includes the RMSE of the trained model with the model output normalized to the range of the data.

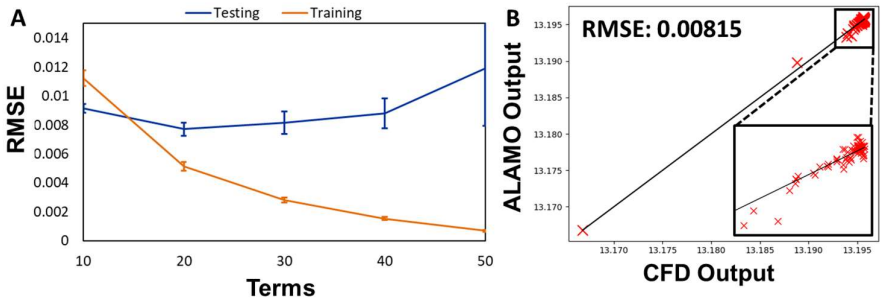


Figure 1 A) RMSE of training and testing data for term number determination of ALAMO model and B) parity plot of the trained ALAMO model versus the CFD output.

The main-effect and total Sobol indices are estimated using the ALAMO model. A plot of how the index values and their confidence intervals change as the number of model evaluations increases is given in Fig. 2. The number of model evaluations, N, are increased according to $N = n(2D + 2)$ where D is the input dimension, and $n = 2^x$ ($x \in \mathbb{Z}^+$) is the number of Sobol sequence samples to satisfy the convergence properties of the Sobol sequence. For this study, $x \in [1,15]$.

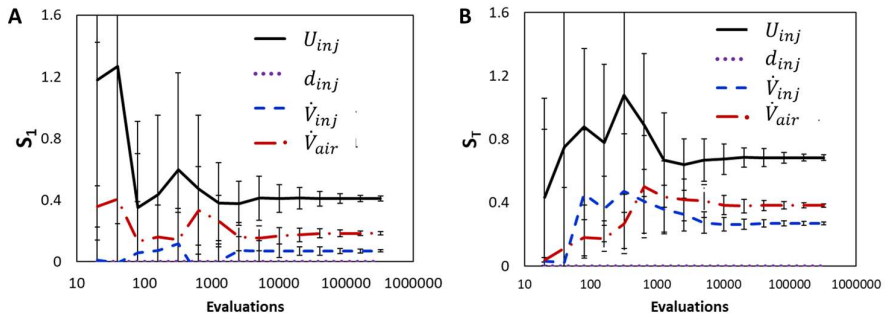


Figure 2 A) Main-effect and B) Total Sobol indices estimated using the ALAMO model

Fig. 2A is a plot of the main-effect indices, and Fig. 2B total effect indices. As can be seen from Fig. 2, the droplet diameter, d_{inj}, is the most impactful parameter and the injection velocity, U_{inj}, the least with a negligible contribution to the variation of outlet humidity for the system considering the main-effect and total Sobol indices. The bulk air flowrate, $\dot{V}_{air}$, has the second-highest contribution for both the main effect and the total Sobol indicies, while the injected water flowrate, $\dot{V}_{inj}$, shows to have the third most contribution. The droplet diameter impacts the number of droplets for condensation to occur within the system, driving the overall reduction in humidity. The air flowrate determines the droplet residence time in the duct, controlling the time allowed for condensation to occur on each droplet. When the injected water flowrate is considered on its own, it has little impact on dehumidification (Fig. 2A). However, when its interactions

with the other parameters are considered, its impact increases (Fig. 2B). The primary interaction contributing to its increased impact is with the droplet diameter. For higher injection rates with smaller droplet diameters, many droplets are injected, enhancing condensation.

5. Conclusions and Future Work

We introduce a computationally-efficient surrogate-model-based sensitivity analysis (SA) approach to identify the critical parameters of a CFD model. The approach combines accurate surrogate modeling and Sobol SA. We applied the approach to a CFD model of an electrospray system designed for enhanced dehumidification of moist air in a duct. The results indicate that the dehumidification rate was most sensitive to the diameter of the charged water droplets injected into the duct. Our analysis also revealed that the number of CFD model evaluations to estimate stable Sobol indices is temporally intractable without employing the proposed approach for this system. The SA results will be used to construct an experimental matrix for a scale-up study of the electrospray system for maximizing dehumidification.

6. Acknowledgments

This work is supported by the U.S. DOE Office of Energy Efficiency and Renewable Energy (EERE) under Building Technologies Program Award Number DE-EE0009161.

References

Abe, K., Kameda, M., 2003. Condensation on droplets in moist air by heterogeneous nucleation. Comput. FLUID Dyn. J. 12, 295–308.

Burman, P., 1989. A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. Biometrika 76, 503–514.

Collet, M., Salomon, S., Klein, N.Y., Seichepine, F., Vieu, C., Nicu, L., Larrieu, G., 2015. Large-scale assembly of single nanowires through capillary-assisted dielectrophoresis. Adv. Mater. 27, 1268–1273.

Cozad, A., Sahinidis, N. V, Miller, D.C., 2014. Learning surrogate models for simulation-based optimization. AIChE J. 60, 2211–2227. https://doi.org/https://doi.org/10.1002/aic.14418

Henslee, E.A., Sano, M.B., Rojas, A.D., Schmelz, E.M., Davalos, R. V, 2011. Selective concentration of human cancer cells using contactless dielectrophoresis. Electrophoresis 32, 2523–2529.

Higashiyama, Y., Kamada, M., 2017. Dehumidification using negative corona discharge from a water droplet, in: Proceedings of the Annual Meeting of the Electrostatics of America, Ottawa, ON, Canada. pp. 13–15.

Morcelli, S., Cremaschi, L., 2021. Analysis of Electro-static Assisted Air Dehumidification Processes.

Pohl, H.A., 1951. The Motion and Precipitation of Suspensoids in Divergent Electric Fields. J. Appl. Phys. 22, 869–871. https://doi.org/10.1063/1.1700065

Sobol, I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Math. Comput. Simul. 55, 271–280.

Tian, W., 2013. A review of sensitivity analysis methods in building energy analysis. Renew. Sustain. energy Rev. 20, 411–419.

Williams, B., Cremaschi, S., 2021. Selection of surrogate modeling techniques for surface approximation and surrogate-based optimization. Chem. Eng. Res. Des. 170, 76–89.

Williams, B., Otashu, J., Leyland, S., Eden, M.R., Cremaschi, S., 2022. PRESTO: Predictive REcommendation of Surrogate models To approximate and Optimize. Chem. Eng. Sci. 249, 117360. https://doi.org/https://doi.org/10.1016/j.ces.2021.117360

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50146-9

Rigorous modelling for comparing batch and flow syntheses of a drug substance using heterogeneous hydrogenation

Junu Kim[a], Hironori Yonekura[b], Takeaki Watanabe[b], Satoshi Yoshikawa[b], Hayao Nakanishi[b], Sara Badr[a], Hirokazu Sugiyama[a*]

[a]*Department of Chemical System Engineering, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-8656, JAPAN*
[b]*Production Technology Department, Shionogi Pharma Co., Ltd., 2-5-1, Mishima, Settsu-Shi, Osaka, 566-0022, JAPAN*
sugiyama@chemsys.t.u-tokyo.ac.jp

Abstract

This work presents the simulation-based comparison of batch and flow syntheses of a drug substance. The target reaction is heterogeneous hydrogenation which is the most widely used reduction method in pharmaceutical synthesis. We developed rigorous physical models of batch and flow syntheses considering the decrease in catalyst activity due to poisoning. Sensitivity analysis of reaction rate constant and viscosity was performed. The reaction rate constant had an effect on the conversion of both batch and flow syntheses. On the other hand, the impact of viscosity on conversion was greater in flow synthesis than in batch synthesis. This result suggests the importance of careful selection of operational conditions when changing from batch to flow synthesis as optimal conditions might differ between them.

Keywords: Flow chemistry; Drug Substance; Physical modelling; Viscosity; Sensitivity

1. Introduction

There have been significant advances in the application of flow synthesis in drug substance production. Bogdan et al. (2009) reported the synthesis of ibuprofen, a widely used painkiller. Other studies include an anti-malaria drug (Lévesque and Seeberger, 2012); olanzapine, an atypical antipsychotic (Hartwig et al., 2013) and rolipram, an anti-inflammatory drug (Tsubogo et al., 2015).

With progress in flow synthesis research, the question remains whether the novel technology proves to be more beneficial than the conventional batch technology. Few studies report specifically on the comparison of drug substance synthesis. Pedersen et al. (2018) redesigned the synthetic routes of melitracen, an antidepressant using flow synthesis, and compared it with the conventional synthetic routes in batch. However, simulation-based investigation in this field is still in infancy despite the need for it due to expensive experimental costs. Therefore, this work presents the simulation-based comparison of batch and flow syntheses for an actual drug substance using heterogeneous hydrogenation. Multi-objective comparison of batch and flow syntheses were performed in the previous work (Kim et al., 2022). Here, the impact of reaction rate and viscosity on batch and flow syntheses are investigated.

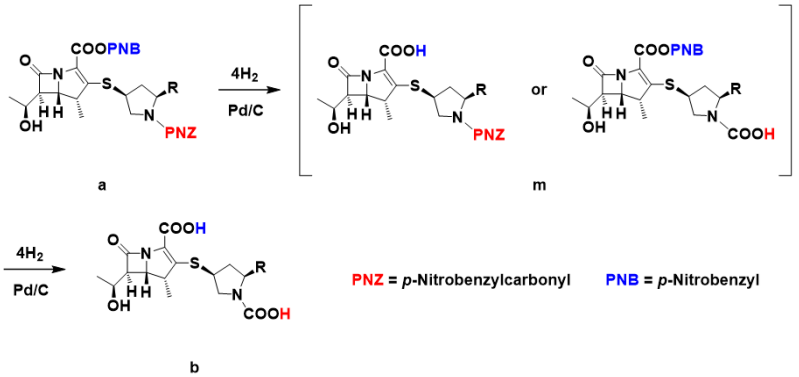


Figure 1. Target reaction scheme (modified from Kim et al. (2022))

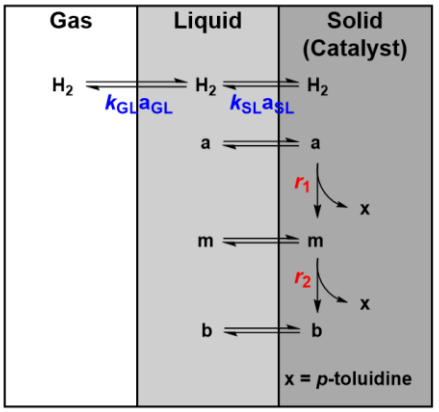


Figure 2. Mass transfer and reactions in gas/liquid/solid phases (modified from Kim et al., (2022))

2. Materials and methods

2.1. Target reaction

Figure 1 shows the target reaction, heterogeneous hydrogenation. The figure was modified based on the previous work (Kim, et al., 2022). The reaction scheme is the part of the synthetic route of an antibiotic active pharmaceutical ingredient, doripenem. In the hydrogenation, compound **a** is reduced to give intermediate **b** in the presence of the hydrogen gas and Pd/C catalyst. During this process, *p*-toluidine is produced as byproduct **x**, which causes the poisoning of the catalyst. In the model, **m** was assumed as intermediate of the hydrogenation reaction, which refers to compounds having either of the two protecting groups remaining. In the original reaction scheme (Kim, et al., 2022), compound **b** undergoes decarboxylation reaction, where the product decomposes into various byproducts. Here, only the hydrogenation part was considered as the comparison of the conversion of **a** is the focus of this work.

2.2. Model assumptions

Figure 2 shows the overview of the mass transfer and reactions. The figure was modified based on the previous study (Kim, et al., 2022). In the model, the same

reaction mechanisms and parameters are assumed in batch and flow syntheses. The mass balance in the liquid and solid phases are considered. The energy balance was neglected in this work because batch experiments did not show significant temperature changes. The hydrogenation reaction was modelled based on the Langmuir–Hinshelwood (LH) mechanism.

2.3. Model equations for batch synthesis

The mass balance equations of liquid and solid phases in batch synthesis were defined as:

$$\frac{\mathrm{d}C_{i,\mathrm{L}}}{\mathrm{d}t} = k_{\mathrm{GL}}a_{\mathrm{GL}}\left(C^{*}_{i,\mathrm{L}} - C_{i,\mathrm{L}}\right) - k_{\mathrm{SL}}a_{\mathrm{SL}}\left(C_{i,\mathrm{L}} - C_{i,\mathrm{S}}\right) \tag{1}$$

$$\frac{\mathrm{d}C_{i,\mathrm{L}}}{\mathrm{d}t} = k_{\mathrm{SL}}a_{\mathrm{SL}}\left(C_{i,\mathrm{L}} - C_{i,\mathrm{S}}\right) + \upsilon_{i,j}r_1{\alpha_{\mathrm{act}}}^2 + \upsilon_{i,j}r_2{\alpha_{\mathrm{act}}}^2 \tag{2}$$

where $C_{i,\mathrm{L}}$ [mol m^{-3}] and $C_{i,\mathrm{S}}$ are the concentrations of compound i (H_2, **a**, **m**, **b**, **x**) in the liquid and solid phases, respectively, t [s] is time, k_{GL} [m s^{-1}] is the gas-liquid mass transfer coefficient, a_{GL} [m^{-1}] is the specific gas-liquid contact area, $C^{*}_{i,\mathrm{L}}$ [mol m^{-3}] is the saturation concentration of i, k_{SL} [m s^{-1}] is the solid-liquid mass transfer coefficient, a_{SL} [m^{-1}] is the specific solid-liquid contact area, $\upsilon_{i,j}$ [–] is the stoichiometric coefficient, r_1 [mol m^{-3} s^{-1}] and r_2 [mol m^{-3} s^{-1}] are the reaction rates for the first and second steps of hydrogenation, respectively, and α_{act} [–] is the activity coefficient of the catalyst. The values for mass transfer between the gas and liquid phases in batch synthesis ($k_{\mathrm{GL}}a_{\mathrm{GL}}$) were estimated by catalogue data (e.g., Satake MultiMix Corporation, 2021) for the actual batch reactor in use. The values for mass transfer between the solid and liquid phases (k_{SL}, a_{SL}) were estimated using the following equations (Sano et al., 1974):

$$k_{\mathrm{SL}} = \frac{ShD_i}{d_{\mathrm{p}}} \tag{3}$$

$$Sh = 2 + 0.4Re_{\mathrm{L}}^{1/4}Sc^{1/3} \tag{4}$$

$$a_{\mathrm{SL}} = \frac{6m_{\mathrm{S}}\rho_{\mathrm{S}}}{d_{\mathrm{p}}m_{\mathrm{L}}\rho_{\mathrm{L}}} \tag{5}$$

where Sh [–] is the Sherwood number, D_i [m^2 s^{-1}] is the diffusion coefficient of compound i in the liquid phase, Re_{L} [–] is the particle Reynolds number in the liquid phase, Sc [–] is the Schmidt number, m_{S} [kg] is the catalyst mass, ρ_{S} [kg m^{-3}] is the density of the catalyst, d_{p} [m] is the particle diameter, m_{L} [kg] is the liquid mass and ρ_{L} [kg m^{-3}] is the density of the liquid. The reaction rates of hydrogenation were modelled by using the following equations:

$$r_1 = \rho_{\mathrm{cat}}k_1\frac{K_{\mathrm{H_2}}K_{\mathbf{a}}C_{\mathrm{H_2,S}}C_{\mathbf{a},\mathrm{S}}}{\left(1 + K_{\mathrm{H_2}}C_{\mathrm{H_2,S}} + K_{\mathbf{a}}C_{\mathbf{a},\mathrm{S}} + K_{\mathbf{m}}C_{\mathbf{m},\mathrm{S}} + K_{\mathbf{b}}C_{\mathbf{b},\mathrm{S}} + K_{\mathbf{x}}C_{\mathbf{x},\mathrm{S}} + K_{\mathbf{c}}C_{\mathbf{c},\mathrm{S}}\right)^2} \tag{6}$$

$$r_2 = \rho_{\mathrm{cat}}k_1\frac{K_{\mathrm{H_2}}K_{\mathbf{m}}C_{\mathrm{H_2,S}}C_{\mathbf{m},\mathrm{S}}}{\left(1 + K_{\mathrm{H_2}}C_{\mathrm{H_2,S}} + K_{\mathbf{a}}C_{\mathbf{a},\mathrm{S}} + K_{\mathbf{m}}C_{\mathbf{m},\mathrm{S}} + K_{\mathbf{b}}C_{\mathbf{b},\mathrm{S}} + K_{\mathbf{x}}C_{\mathbf{x},\mathrm{S}} + K_{\mathbf{c}}C_{\mathbf{c},\mathrm{S}}\right)^2} \tag{7}$$

where ρ_{cat} [kg m^{-3}] is the bulk density of the catalyst, k_1 [mol kg^{-1} s^{-1}] is the reaction rate constant of hydrogenation and K_i [mol^{-1}] is the adsorption coefficient of compound

i. The change in catalyst activity was calculated by the following equation (Kilpiö et al., 2012):

$$\frac{\mathrm{d}\alpha_{\mathrm{act}}}{\mathrm{d}t} = \frac{-\alpha_{\mathrm{act}} k_{\mathrm{p}} K_{\mathbf{x}} C_{\mathbf{x},\mathrm{S}}}{\left(1 + K_{\mathrm{H_2}} C_{\mathrm{H_2,S}} + K_{\mathbf{a}} C_{\mathbf{a},\mathrm{S}} + K_{\mathbf{m}} C_{\mathbf{m},\mathrm{S}} + K_{\mathbf{b}} C_{\mathbf{b},\mathrm{S}} + K_{\mathbf{x}} C_{\mathbf{x},\mathrm{S}} + K_{\mathbf{c}} C_{\mathbf{c},\mathrm{S}}\right)} \tag{8}$$

where k_{p} [s^{-1}] is the poisoning rate constant. Eqs. (6)–(8) were used in the flow synthesis as well.

2.4. Model equations for flow synthesis

The mass balance of liquid and solid phases in flow synthesis are defined as:

$$\frac{\partial C_{i,\mathrm{L}}}{\partial t} = -\frac{u_{\mathrm{L}}}{\varepsilon_{\mathrm{L}}}\frac{\partial C_{i,\mathrm{L}}}{\partial l} + D_i \frac{\partial^2 C_{i,\mathrm{L}}}{\partial l^2} + \frac{1}{\varepsilon_{\mathrm{L}}} k_{\mathrm{GL}} a_{\mathrm{GL}} \left(C^*_{i,\mathrm{L}} - C_{i,\mathrm{L}}\right) - \frac{1}{\varepsilon_{\mathrm{L}}} \mathrm{k_{SL}} a_{\mathrm{SL}} (C_{i,\mathrm{L}} - C_{i,\mathrm{S}}) \tag{9}$$

$$\frac{\partial C_{i,\mathrm{S}}}{\partial t} = \frac{1}{\varepsilon_{\mathrm{s}}} \mathrm{k_{SL}} a_{\mathrm{SL}} \left(C_{i,\mathrm{L}} - C_{i,\mathrm{S}}\right) + \upsilon_{i,j} r_1 {\alpha_{\mathrm{act}}}^2 + \upsilon_{i,j} r_2 {\alpha_{\mathrm{act}}}^2 \tag{10}$$

where u_{L} [m s^{-1}] is the superficial liquid velocity, ε_{L} [–] is the liquid holdup and l [m] represents the location in the reactor, and ε_{s} [–] is the solid holdup. The values for mass transfer between the gas and liquid phases ($k_{\mathrm{GL}} a_{\mathrm{GL}}$) were estimated by the following equation (Kilpiö et al., 2012):

$$\frac{k_{\mathrm{GL}} a_{\mathrm{GL}} d_{\mathrm{p}}^2}{D_{\mathrm{H_2}} (1 - \frac{\varepsilon_{\mathrm{L}}}{\varepsilon_{\mathrm{P}}})} = 2\theta^{0.2} {Re_{\mathrm{L}}}^{0.73} {Re_{\mathrm{G}}}^{0.2} Sc^{0.5} \left(\frac{d_{\mathrm{p}}}{d_{\mathrm{pipe}}}\right)^{0.2} \tag{11}$$

where ε_{P} [–] is the porosity of the reactor, θ [–] is the surface shape factor for the particle (specified as 2.0 for a sphere according to Kilpiö et al., 2012), Re_{G} [–] is the particle Reynolds number in the gas phase and d_{pipe} [m] is the diameter of the reactor. The values for mass transfer between the liquid and solid phases (k_{SL}, a_{SL}) were estimated by the following equations (Burghardt et al., 1995).

$$Sh = 2 + 1.1 Re_{\mathrm{L}}^{0.6} Sc^{1/3} \tag{12}$$

$$a_{\mathrm{SL}} = \frac{6}{d_{\mathrm{p}}} (1 - \varepsilon_{\mathrm{P}}) \tag{13}$$

3. Results and discussion

3.1. Conversion of raw materials

The conversion of batch synthesis was defined as:

$$conv = \left(1 - \frac{C_{\mathbf{a},\mathrm{L}}(t_{\mathrm{batch}})}{C_{\mathbf{a},\mathrm{L}}(0)}\right) \times 100 \tag{14}$$

where t_{batch} [min] is the batch reaction time and $C_{\mathbf{a},\mathrm{L}}(0)$ is the initial concentration of **a**. In this work, $t_{\mathrm{batch}} = 60$ min. The conversion of flow synthesis was defined as:

$$conv = (1 - \frac{\sum_{t_0}^{t_1} C_{\mathbf{a},\mathrm{L}}(t) \cdot v_\mathrm{L} \cdot \Delta t}{C_{\mathbf{a},\mathrm{L}}(0) \cdot v_\mathrm{L} \cdot (t_1 - t_0)}) \times 100 \quad (15)$$

where $C_{\mathbf{a},\mathrm{L}}(0)$ is the concentration of **a** at the inlet, v_L [cm^3 min^{-1}] is the volumetric flow rate of the liquid phase and Δt is the time interval for the calculation output (set as 1 min in the simulation). The parameters t_0 [min] and t_1 [min] represent the beginning and the end of production using flow synthesis (start-up and shut-down were excluded).

3.2. Sensitivity analysis of reaction rate constant and viscosity

Sensitivity analysis was performed to investigate the impact of the hydrogenation reaction rate constant and viscosity on the conversion in batch and flow syntheses. The parameters were changed under the following ranges.

$$k_1 \in \{5, 10, \cdots, 100\}$$

$$\eta_{\mathrm{solvent}} \in \{0.2, 0.25, \cdots, 2.0\}$$

Here, η_{solvent} [mPa s] is the viscosity of the solvent. Figure 3 shows the conversion of **a**. The reaction rate constant affects the conversion of both batch and flow. It showed less impact in flow synthesis, especially with smaller viscosity values. This is because almost 100% of conversion was obtained even under small reaction rate constant values. On the other hand, the viscosity showed a greater impact on the conversion of flow compared to batch. The viscosity affects the mass transfer between phases (gas-liquid and liquid-solid), and it has less impact in batch synthesis because mass transfer in batch is sufficient even with larger viscosity values.

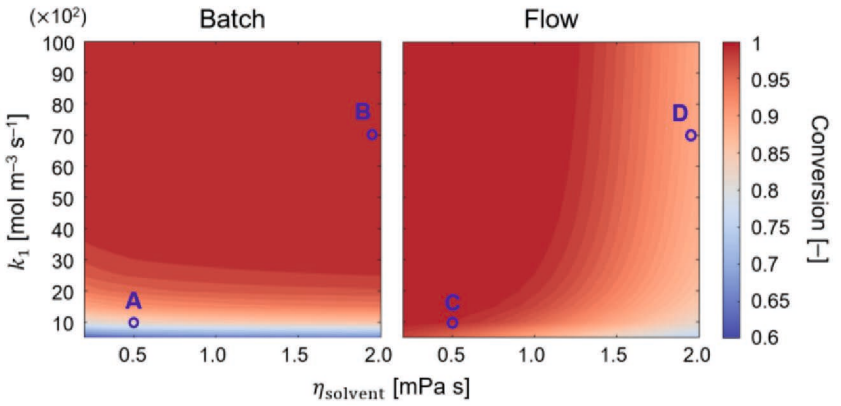


Figure 3. Conversion in batch and flow syntheses

Table 1. Parameter values and conversion at four conditions

Condition	η_{solvent} [mPa s]	k_1 [10^2 mol m^{-3} s^{-1}]	Conversion [–]
A	0.5	10	0.83
B	1.95	70	1.00
C	0.5	10	0.99
D	1.95	70	0.90

The results suggest that the optimal condition may differ between batch and flow synthesis, and could result in a need for extra flow experiments at all relevant conditions when changing from batch to flow synthesis. For example, the optimal operation was obtained at condition B (Figure 3) in batch synthesis, which shows 17% higher conversion than the condition A (Table 1). However, condition C shows 9% higher conversion than condition D in flow synthesis. Further investigation of the impact of other parameters such as solvent density could contribute to reducing the number of or avoiding extra experiments needed for the change from batch to flow.

4. Conclusions and outlook

Sensitivity analysis was performed to compare the impact of reaction rate constant and viscosity in batch and flow syntheses. Reaction rate constant influenced the conversion of both batch and flow syntheses. Viscosity showed much greater impact on the conversion of flow than that of batch. The differences in optimal condition between batch and flow hinder the direct transfer of process experiences between the different production modes. Thus, dedicated experiments are required to optimize production in flow synthesis. Extensive simulation of flow synthesis incorporating the effects of other parameters could help lower the expected experimental load.

References

A.R. Bogdan, S.L. Poe, D.C. Kubis, S.J. Broadwater, D.T. McQuade, 2009. The continuous-flow synthesis of ibuprofen. Angew. Chem., Int. Ed. 48, 8547–8550.

A. Burghardt, G. Bartelmus, M. Jaroszyński, A. Kołodziej, 1995. Hydrodynamics and mass transfer in a three-phase fixed-bed reactor with cocurrent gas-liquid downflow. Chem. Eng. J. Biochem. Eng. J. 58, 83–99.

J. Hartwig, S. Ceylan, L. Kupracz, L. Coutable, A. Kirschning, 2013. Heating under high-frequency inductive conditions: Application to the continuous synthesis of the neurolepticum olanzapine (Zyprexa). Angew. Chem., Int. Ed. 52, 9813–9817.

J. Kim, H. Yonekura, T. Watanabe, S. Yoshikawa, H. Nakanishi, S. Badr, H. Sugiyama, 2022, Model-based comparison of batch and flow syntheses of an active pharmaceutical ingredient using heterogeneous hydrogenation, Comput. Chem. Eng., 156, 107541.

T. Kilpiö, P. Mäki-Arvela, M. Rönnholm, V. Sifontes, J. Wärnå, T. Salmi, 2012. Modeling of a three-phase continuously operating isothermal packed-bed reactor: Kinetics, mass-transfer, and dispersion effects in the hydrogenation of citral. Ind. Eng. Chem. Res. 51, 8858–8866.

F. Lévesque, P.H. Seeberger, 2012. Continuous-flow synthesis of the anti-malaria drug artemisinin. Angew. Chem., Int. Ed. 51, 1706–1709.

M.J. Pedersen, T. Skovby, M.J. Mealy, K. Dam-Johansen, S. Kiil, 2018. Redesign of a Grignard-based active pharmaceutical ingredient (API) batch synthesis to a flow process for the preparation of melitracen HCl. Org. Proc. Res. Dev. 22, 228–235.

Y. Sano, N. Yamaguchi, T. Adachi, 1974. Mass transfer coefficients for suspended particles in agitated vessels and bubble columns. J. Chem. Eng. Jpn., 7 (4) 255–261.

Satake MultiMix Corporation, https://www.satake.co.jp/product/catalog/download/data/17_impeller.pdf (accessed November 5, 2021)

T. Tsubogo, H. Oyamada, S. Kobayashi, 2015. Multistep continuous-flow synthesis of (R)- and (S)-rolipram using heterogeneous catalysts. Nature 520, 329–332.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50147-0

Assessment on the heat integration potential for different pressure thermally coupled distillation structures

J. Rafael Alcántara-Avila[a]*, Rodrigo Tinoco Saenz[a,b]

[a]*Department of Chemical Engineering, Kyoto University, Katsura Campus Nishikyo-ku, Kyoto 615-8510, Japan*
[b]*Sales del Istmo, S.A de C.V. Grupo CYDSA. Complejo industrial Pajaritos, Coatzacoalcos, 96400, México*
jrafael@cheme.kyoto-u.ac.jp

Abstract

This work presents a comparison of several distillation structures that can attain energy and cost savings. The Different Pressure Thermally Coupled Distillation (DPTCD) was studied to increase its energy-savings potential by exploiting heat integration among stages. Therefore, a Mixed-Integer Linear Programming (MILP) problem was proposed and solved to find heat-integrated distillation column (HIDiC) structures that can improve energy and cost savings of DPTCD. The separation of a closed-boiling point mixture was taken as a case study. The HIDiC slightly outperformed DPTCD in energy and cost terms.

Keywords: Process Synthesis, MILP, DPTCD, HIDiC, Heat Integration

1. Introduction

Distillation is the most widely used technique to separate liquid mixtures. However, it uses large amounts of energy because it boils liquid streams. Typically, more than half of the process heat distributed to a plant is dedicated to supplying heat in distillation columns' reboilers (Caballero and Grossmann, 2013). To improve energy efficiency in distillation processes, several alternatives, including heat integration and thermal coupling, have been proposed. Heat integration between hot streams and cold streams has been widely used in the Chemical Industry.

Different Pressure Thermally Coupled distillation (DPTCD) has been proposed as an energy-saving distillation alternative for the separation of close-boiling point mixtures (Liu *et al.*, 2017) and in reactive distillation (Li *et al.*, 2015). DPTCD claims that heat integration between the top vapor of a high-pressure rectifying column and the bottom liquid of a low-pressure stripping column reduces the energy requirements in comparison with its conventional counterpart. The Vapor Recompression Column (VRC) can also realize condenser-reboiler heat integration by compressing the top vapor stream leaving the column. Finally, the Heat Integrated Distillation Column (HIDiC) can also realize condenser-reboiler heat integration. However, it also exploits the idea of realizing heat integration among stages in the rectifying and stripping columns. This work compares the economic and energy performance of DPTCD, VRD, and HIDiC. The economic criterion is assessed by minimizing the total annual cost (TAC), while the energy requirement criterion is assessed by minimizing energy consumption.

2. Problem Statement

The HIDiC structure is expected to be the most promising alternative to reduce the energy consumption in the chemical plants because outstanding energy savings up to 60% in comparison with its conventional distillation counterparts have been obtained theoretically and experimentally in pilot plants (Matsuda, Iwakabe and Nakaiwa, 2012). However, the typical HIDiC structure (i.e., concentric arrangement of inner and outer tubes) has presented operating issues such as difficult maintenance, HIDiC structures with few heat integrations have been proposed (Wakabayashi and Hasebe, 2013, 2015).

The simulation of DPTCD and VRC alternatives is straightforward because the only heat integration is between the condenser and reboiler. However, the simulation and optimization of the HIDiC structure is more challenging because the following questions must be answered: (1) how many heat integrations are needed? and (2) what is the best amount of heat transferred at the chosen locations?

This work adopts a synthesis methodology for the generation of near-optimal HIDIC structures where the stages in a high-pressure rectifying column are regarded as heat sources, and the stages in a low-pressure stripping column are regarded as heat sinks. The synthesis problem is represented by a superstructure in which all possible heat integrations are explicitly included. Finally, the superstructure is formulated as a mixed-integer linear programming (MILP) problem that successively combines process simulation and optimization (Alcantara-Avila, 2019; Herrera Velázquez *et al.*, 2021).

In this work, the synthesis problem for the heat integration among stages in a HIDiC structure is formulated based on the following assumptions:

1. The top pressure in each section is given in advance, and the pressure drop is considered (i.e., 75 kPa per stage).
2. The feed composition and product specifications are given in advance.
3. The feed and products are saturated liquids.
4. Only one stage in the high-pressure column (i.e., rectifying section) can supply heat to only one stage in the low-pressure column (i.e., stripping section).
5. Only one stage in the low-pressure column (i.e., stripping section) can receive heat from only one stage in the high-pressure column (i.e., rectifying section).
6. The minimum temperature difference value that enables heat integration is given in advance.
7. The compressor type and efficiency are given in advance.
8. The process is at the steady-state.
9. Vapor-liquid equilibrium is reached in each distillation stage.
10. Utility costs per unit amount of heating and cooling are given in advance.

The synthesis problem is represented by the superstructure in Figure 1. *REC* and *STR* are the sets of stages in the rectifying and stripping sections. *LC* and *HC* are the light and heavy component(s), respectively. The arrows connecting stages of different sections represent the heat transfer between stages.

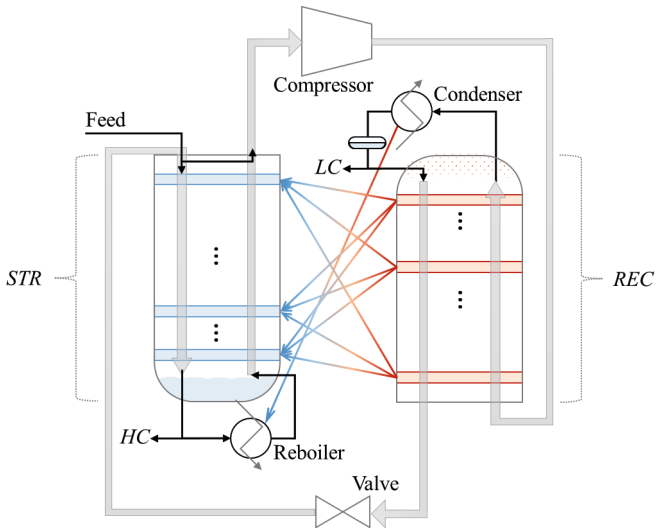


Figure 1. Superstructure representation

3. Mathematical formulation

The superstructure in Figure 1 can be reformulated as the optimization problem represented by Eq. (1). The detailed optimization model can be found in Herrera Velázquez *et al.*, (2021).

$$UC = \sum_{\substack{i\in REC \\ j\in CU}} Q_{i,j}^{\mathrm{hx}} C_j^{\mathrm{cool}} + \sum_{\substack{i\in HU \\ j\in STR}} C_i^{\mathrm{heat}} Q_{i,j}^{\mathrm{hx}} + WC^{\mathrm{elec}} \tag{1}$$

s.t.

- Heat balance at each stage
- Heat balance in the condenser, reboiler, and compressor
- Heat integration feasibility
- Work, condenser and reboiler duty estimation

where UC is the utility cost, CU and HU are the set of cooling and heating utilities, respectively. $Q_{i,j}^{\mathrm{hx}}$ is the heat exchange between a heat source i and a heat sink j. C_j^{cool} and C_i^{heat} is the cooling and heating cost of utilities per unit amount of heat. W is the compressor work duty and C^{elec} is the electricity cost.

The optimization variables are the pressure ratio between the rectifying and stripping sections (PR), the heat removed at a stage in the rectifying section (Q_i), the heat removed at a stage in the stripping section (Q_j), the location of a heat integration ($Y_{i,j}^{\mathrm{hx}}$), and the number of heat integrations (N^{hx}), respectively. $Y_{i,j}^{\mathrm{hx}}$ is an integer variable that becomes one if heat integration is realized between stages i and j while it becomes zero otherwise. Therefore, the optimization problem can be solved as an MILP problem.

The stage-by-stage material balance, thermodynamic relationships, and summation restrictions are considered in the process simulation Aspen Plus V11. The iterative execution of simulations and optimizations result in the optimal HIDiC structure when the convergence criterion in Eq. (2) is met.

$$\sqrt{\sum_{i\in REC}\left(T_{i,s}-T_{i,s-1}\right)^2+\sum_{j\in STR}\left(T_{j,s}-T_{j,s-1}\right)^2}\leq(NR+NS)\varphi \tag{2}$$

where T is the temperature of stage i and j at an iteration s. NR and NS are the number of stages in the rectifying and stripping sections of the HIDiC structure, and φ is the tolerance for the termination criterion, which is a small value (e. g., 0.01).

4. Case Study: n-butanol/isobutanol separation

The separation of a close-boiling point mixture of n-butanol and isobutanol studied by Liu *et al.*, (2017) is taken as a case study to compare economic and energy criteria for the DPTCD, VRC, and HIDiC structures. The key ideas in this comparison are to know the true energy-saving potential of DPTCD and to know if HIDiC can outperform DPTCD.

Table 1. shows the parameters used in the optimization problem.

Parameter	Value
Cooling cost (C_j^{cool}) [$/GJ]	0.354
Heating cost (C_i^{heat}) [$/GJ]	7.720
Electricity cost (C^{elec}) [$/GJ]	16.81
Payback time [yr]	5
Overall heat transfer coefficient [BTU/h F ft^2]	
Condenser:	190
Heat exchanger / Reboiler:	175
Big-M parameter (M) [-]	250
Minimum Temperature difference ($\Delta T_{\min}$) [F]	18

The total annual cost (TAC) was calculated using the Marshall & Swift index for 2020 (i.e., 2171.6). The TAC was calculated according to Eq. (3)

$$\min TAC = OC * \text{AOH} + \frac{\left[C_{\text{shell}}+C_{\text{tray}}+C_{\text{comp}}+\sum_{i\in HS}\sum_{j\in CS}C_{hex_{i,j}}\right]}{\text{PBP}} \tag{3}$$

where $i \in HS$ is the set of heating streams ($HS = REC \cup HU$), $j \in CS$ is the set of cooling streams ($CS = STR \cup CU$), C_{shell}, C_{tray}, C_{comp}, and $C_{hex_{i,j}}$ is the cost of the column shell, trays, compressor, and heat exchangers. AOH and PBP are the annual operation hours (i.e., 8000 h/y) and the payback time (i.e., 5 years), respectively.

5. Results and Discussions

Table 2 shows the results of the conventional distillation column (CD), VRC, DPTCD, and HIDiD structures after executing the optimization problem in Section 3. It is worthy of keeping in mind that even the solution of the MILP needs to determine the following questions: 1) how many heat integrations does the optimal result have, and 2) what is the maximum amount of heat integrated at stages. Table 2 shows the simulation and optimization results, while Table 3 summarizes the energy and cost savings.

Table 2. TAC results for the different distillation structures

	CD*	**VRC***	**DPTCD***	**HIDiC**
Heating [$/y]	542,169	---	---	23,072
Cooling [$/y]	25,554	4,187	3,300	3,856
Electricity [$/y]	---	165,934	137,710	104,147
Operating Cost [$/y]	567,723	170,120	141,009	131,075
Condenser [$]	211,959	79,699	55,954	64,231
Reboiler [$]	456,595	---	---	42,009
Heat Exchangers [$]	---	761,662	522,257	656,844
Column [$]	1,327,953	1,327,953	1,360,239	1,519,093
Tray [$]	156,973	156,973	133,271	151,423
Compressor [$]	---	1,903,129	1,633,325	1,298,955
Equipment Cost [$/y]	2,153,479	4,229,415	3,705,045	3,732,554
TAC [$/y]	998,419	1,016,003	882,018	877,586

*Adapted from the simulation results in Liu *et al.*, (2017).

Table 3. Energy and cost savings comparison for the different distillation structures

	CD	**VRC**	**DPTCD**	**HIDiC**
Energy Consumption [kW]	2438.5	935.7	776.5	691.1
Energy savings [%]	0.0	61.6	68.2	71.7
TAC savings [%]	0.0	-1.8	11.7	12.1

The energy consumption was calculated according to Eq. (4) as the summation the energy from heating utilities and the electricity used in the compressor

$$\text{Energy consumption} = \sum_{\substack{i\in HU \\ j\in STR}} Q_{i,j}^{\text{hx}} + 2.73W \tag{3}$$

From the results in Tables 2 and 3, it can be seen that in terms of energy consumption, the HIDiC uses less energy because it operates at a compression ratio smaller than that of DPTCD (2.0 and 2.6, respectively) despite HIDiC needs steam at the bottom reboiler in the stripping section. Although HIDiC needs energy, electricity, and compressor costs less than DPTCD, it increases the column, trays, and heat exchangers cost.

Therefore, overall, DPTCD and HIDiC have similar TAC, but there is a trade-off between energy consumption and equipment cost. This same behavior has been presented in previous works when compressors are included in the synthesis and optimization of distillation sequences (Cabrera-Ruiz *et al.*, 2012; J. Alcántara-Avila, Kano and Hasebe, 2012; J. R. Alcántara-Avila, Kano and Hasebe, 2012). The HIDiC structure had one heat integration between stage 9 in the rectifying section and the reboiler in the stripping section.

6. Conclusions

This work presented a comparison between several distillation structures that can realize heat integration between a condenser and reboiler as well as among stages in different column sections. When one heat integration was done, the HIDiC structure attained energy and cost savings higher than DPTCD. Although the HIDiC and DPTCD can have very similar performances, there is a trade-off between these alternatives.

References

Alcántara-Avila, J., Kano, M. and Hasebe, S. (2012) 'Environmental and economic optimization of distillation structures to produce anhydrous ethanol', *Computer Aided Chemical Engineering*.

Alcantara-Avila, J. R. (2019) 'Optimization methodologies for intensified distillation processes with flexible heat integration networks', in Gomez Castro, F. I. and Segovia-Hernández, J. G. (eds) *Process Intensification: Design Methodologies*. 1st edn. Berlin: De Gruyter, pp. 181–209.

Alcántara-Avila, J. R., Kano, M. and Hasebe, S. (2012) 'Multiobjective optimization for synthesizing compressor-aided distillation sequences with heat integration', *Industrial and Engineering Chemistry Research*, 51(17), pp. 5911–5921.

Caballero, J. A. and Grossmann, I. E. (2013) 'Synthesis of Complex Thermally Coupled Distillation Systems Including Divided Wall Columns', *AIChE Journal*, 59(4), pp. 1139–1159.

Cabrera-Ruiz, J. *et al.* (2012) 'Optimal dynamic controllability in compressor-aided distillation schemes using stochastic algorithms', *Computer Aided Chemical Engineering*.

Herrera Velázquez, J. J. *et al.* (2021) 'Hybrid two-step optimization of internally heat-integrated distillation columns', *Journal of the Taiwan Institute of Chemical Engineers*. In press.

Li, L. *et al.* (2015) 'Design and Control of Different Pressure Thermally Coupled Reactive Distillation for Methyl Acetate Hydrolysis', *Industrial and Engineering Chemistry Research*, 54(49), pp. 12342–12353.

Liu, L. *et al.* (2017) 'Simulation and optimization of different pressure thermally coupled distillation for separating a close-boiling mixture of n-butanol and isobutanol', *Applied Petrochemical Research*. Springer Berlin Heidelberg, 7(2–4), pp. 143–150.

Matsuda, K., Iwakabe, K. and Nakaiwa, M. (2012) 'Recent advances in internally Heat-integrated distillation columns (HIDiC) for sustainable development', *Journal of Chemical Engineering of Japan*, 45(6), pp. 363–372.

Wakabayashi, T. and Hasebe, S. (2013) 'Reversible distillation curve and operating locus as criteria for appropriate of HIDiC', *Kagaku Kogaku Ronbunshu*, 39(6), pp. 493–502.

Wakabayashi, T. and Hasebe, S. (2015) 'Higher Energy Saving with New Heat Integration Arrangement in Heat-Integrated Distillation Column', *AIChE Journal*, 61(10), pp. 3479–2488.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50148-2

Python platform for Tennessee Eastman Process

Chengyu HAN, Shiping HUANG, and Wei SUN*

College of Chemical Engineering, Beijing University of Chemical Technology, Beijing 100029, CHINA.
sunwei@mail.buct.edu.cn

Abstract

Tennessee Eastman process (TEP) is a widely used test benchmark in the field of process monitoring and process control. A complete chemical plant is simulated, which can provide a large amount of process data as DCS (Distributed Control System), and people can introduce various faults artificially as case studies in their research. original TEP program was written in FORTRAN. Ricker used f2c (Fortran to C) program to convert Fortran code to C code, and wrote a mex function to connect TEP to MATLAB. Users can run the TEP process graphically and view the curves of variables. The graphical interface of MATLAB/Simulink is easy to use, but it takes a lot of glue code to set up TEP's parameters and run simulations in other programming languages. In addition, the fault adding logic of TEP is put in the code written in C language, which prevents the fault from being added dynamically at the expected time.

In order to make better use of the TEP process and make it convenient for people to use it, a method of calling TEP through python instead of MATLAB is proposed. Through the interaction between python's C-API and TEP's C program, all the functions of TEP are exported. The stop time of running open-loop simulation is consistent with the result of MATLAB, and the relative error between the modified program running results and the data generated by the original TEP program with open-loop is less than 10^{-6}. Ricker developed a Simplified TEP (S-TEP), and kept only one tank reactor with two phases. In the same way, the method of connecting S-TEP with python is proposed. Then the C library with python interface is bundled into a python package. This allows users to install all programs needed to run S-TEP, and only one python package needs to install through the python package manager without configuring the C language compilation environment, which greatly facilitates the use of S-TEP. A GUI interface was added to the modified S-TEP program to make it easier for people to use.

Keywords: Process Monitoring; Fault detection; Tennessee Eastman Process; Process models; Process simulators.

1. Introduction

Modern chemical industry is large in scale and high in complexity, which runs a multivariable control system. With the upgrading of chemical processes and the use of new reaction equipment, the chemical process has become more and more complex and highly nonlinear. Large-scale complex systems will produce a large amount of process data. Although more information from the process can be collected, the amount of data provided by existing distributed control system (DCS) is quite tremendous, which makes it hard for operators or engineers to evaluate the effectiveness of the process only by observing process data. Especially when a fault occurs, it is more difficult to find the real cause of the fault from overwhelming information provided by DCS and solve it quickly.

In order to ensure that the process operations meet the performance specifications, faults in the process need to be detected, diagnosed and eliminated. These tasks are related to process monitoring, more specifically, it is data-driven process monitoring. Benefited from proper process monitoring, downtime is minimized, safety of plant operations is improved, and production costs is reduced. With the increase of the number of instruments in the plant, the amount of data available for monitoring and troubleshooting is also increasing.

Data-driven methods require a large amount of process data, including data under normal working conditions and data in case of failure. For the stable operation of the chemical process, most of the time is in normal state. Faults only occur in unusual situations, and the operator will try to shorten the duration of faults as much as possible. Factories will not provide its process data for data security or confidentiality. Therefore, it is difficult to extract the fault state data from the actual chemical process.

Downs and Vogel (1993) presented the Tennessee Eastman Challenge in their paper. This process is also known as Tennessee Eastman Process (TEP). TEC is based on a real industrial process and consists of a set of FORTRAN subroutines. Running these subroutines can obtain instrumental information in a process like DCS. Furthermore, by introducing faults manually, we can capture process data in case of faults. TEP has been used to research and evaluate many aspects, such as process monitoring algorithms, control system design, optimization algorithms. Many authors have used it to demonstrate their procedures for the design of control systems (Ricker, 1995; Ricker and Lee, 1995; Ricker, 1996; Larsson and Skogestad, 2001).

Bathelt and Ricker (2015) have studied the underlying computer code of the well-known TEP model. A bug was found in the original TEP. The causes of bug are discussed in this paper and a solution in terms of a modified code is presented. Furthermore, some additional changes are discussed, widening the usability of the simulation model. This improvement to the TEP has improved TEP, but the current TEP is still difficult to extend programmatically. Martin-Villalba and Guodong Shao (2018) presented the development of two different Modelica libraries, TE-process and TE-Simplified. Using Modelica is an important attempt to improve the programmability of TEP, which proves that TEP does not completely rely on MATLAB. However, the modelling language used by Modelica is not very popular, and there are not many packages related to statistical algorithms, so we tried to use other languages to replace MATLAB, and after some comparison we finally chose python to encapsulate the TEP code.

2. Different versions of TEP

Ignoring the control part and considering only the core process simulation part, there are many versions of the TEP. The most important versions are: the original TEP, the simplified version of the TEP and the latest version of the TEP.

2.1. Original TEP

The TEP was created by the Eastman Chemical Company to provide a realistic industrial process for evaluating process control and monitoring methods. The testing process is based on a simulation of an actual industrial process, in which components, dynamics and operating conditions are modified due to patent reasons. This process is well suited for a wide variety of studies including both plant-wide control and multivariable control problems. FORTRAN code representing the process is available from Downs and Vogel (1993), but they have chosen not to publish the model equations. Instead, they provide a

flowsheet, a steady-state material balance, and a qualitative description of the key process characteristics. The process is nonlinear and open-loop unstable. Without control, it reaches shutdown limits within an hour, even for very small disturbances.

All codes are written in FORTRAN 77 in a fixed format. According to the notes in the source code, MEX interface only supports the MATLAB version 3.x and 4.x, and cannot be used with the new version of MATLAB at present. Only FORTRAN source files and MEX interface codes are provided, and no control codes of MATLAB are provided. All symbols in the FORTRAN code are capitalized, and there is no space or indentation between statements. This version of the code is the original TEC code, and is often used as the basis for improvements in other versions, but it can neither be compiled nor read because of its long history and incomplete code. Therefore, papers will use the dataset that comes with the original TEP directly instead of running the program to generate a new dataset. If not specified in papers, the TEP dataset usually refers to the dataset accompanying this version.

2.2. Simplified TEP

Ricker (1995) proposed the Simplified Tennessee Eastman Challenge (S-TEP) which is a simplification of the TEP model, with only one process unit and eight state variables. The process consists of a single vessel that represents a combination of the reactor and separation system in the original TE process. The process unit has two input flows and two output flow. The S-TEP model is also a well-known benchmark process. It is a multi-input multi-output, nonlinear system, open-loop unstable, and contains fast and slow dynamics.

FORTRAN77 with fixed format is used in the code of process simulation. The control part of the code is written by MATLAB, but only supports MATLAB version 3.x or 4.x. Therefore, the S-TEP code needs to be updated to use.

2.3. The Newest TEP

With the update of MATLAB and the improvement of the complexity of TEP control system, a bug was found in the original process simulation code: when using different integration methods or different integration steps, the results of process simulation are inconsistent. After detailed debugging, the problem lies in the generation of random numbers. Different integrators need different convergence times, which leads to different calls of random number generators, which makes the generated random numbers different, and finally leads to inconsistent process simulation results. Bathelt and Ricker (2015) implemented a new random number generation method to fix this problem. Furthermore, some additional changes are discussed, widening the usability of the simulation model.

The core process simulation part is written in C language, and the control part is written in MATLAB/Simulink. The tested code can run on MATLAB 2014~2020. C language can be recompiled by MSVC 2017 compiler. The work in this article is also based on this version.

2.4. Possible way to improve TEP

At present, TEP mainly uses FORTRAN, C and MATLAB. The core part of process simulation is written in FORTRAN and C languages, which runs well and is rarely modified. Most of the modifications are in the MATLAB. MATLAB is used to call the process simulation, integrate the process variables and implement control system. Using TEP to generate data requires using graphical user interface, adjusting parameters, setting the running time length, then clicking the Run button, waiting for the simulation to be

completed, and finally manually exporting the required data sets. If you need data for multiple faults, you need to manually repeat the above steps several times. Using MATLAB script some of the above steps can be completed automatically to a certain extent, but the copyright and toolbox package of MATLAB somehow limited TEP revision. Python or R is usually used in statistical study of data processing. Among the above languages, Python has the most mature data science ecology, and many data-driven algorithms can be used directly. Therefore, we chose Python as a substitute for MATLAB, so that data generation and post-processing can be carried out in one language. The advantage of doing this is to improve the usability of TEP, and to make TEP more widely used.

3. Connect TEP to Python

3.1. System Architecture

The architecture of the 2015 modified version of TEP is shown on the left side of Figure 1. The core module is written in C and the interface is exported for MATLAB through a wrapper of S-functions. The control logic is written in MATLAB, and the input of parameters and the output of results are performed in MATALB.

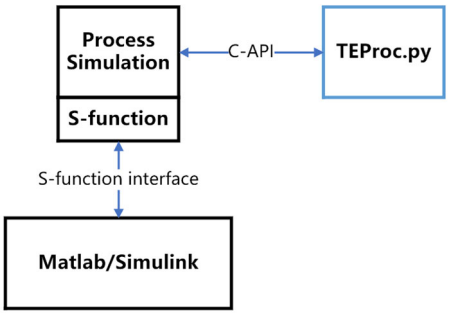


Figure 1 Original and modified TEP architecture diagram

Such an architecture involves encapsulating the details of the process simulation, and the control logic can be modified in MATALB, which is very convenient for researchers in the control field. However, for the field of process monitoring, the control methods of the system are generally not modified, and researchers are more concerned with the introduction of faults, the superposition of noise, and the input of parameters and output of results. The generation and export of the dataset is performed manually in MATLAB's GUI, and the processing of the data and application of the algorithm is performed in python. These can be partially automated in MATLAB, and a more desirable situation would be to use one language to perform these operations as well.

To solve the problem of interaction between TEP and other languages, we propose to use python to call the core modules directly, without MATLAB. The overall architecture is shown on the right side of the figure, where the python wrapper script calls the modules of TEP by using C-API directly, and the control logic is moved to the python script. We use the ctypes module in python standard library to wrap constants, variables, and structures from C code into python, and then define the prototypes of the functions in python. Once the correct C compiled shared library is loaded, the corresponding C functions can be called in python.

3.2. Precision Comparison

To verify that the accuracy of the TEP process simulation does not degrade after switching from MATLAB to python. We used the open-loop TEP model without control system as a test, compared the stopping time of the original TEP and the python wrapped TEP, and calculated the deviation of each process variable. Additional noise was added to the output of the TEP, so we calculated the relative error of the 50 state variables in the TEP. As shown in Figure 2, the relative errors of component A in vapor phase of reactor are within acceptable limits (less than 10^{-6}). This indicates that the same accuracy can be achieved by using python to call the TEP core module directly as by using MATLAB to call the TEP.

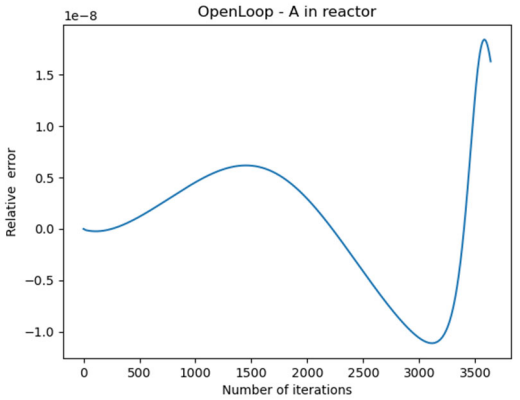


Figure 2 Errors in TEP.py vs. TEP-2015

3.3. GUI for S-TEP

The TEP process is complex and requires many variables to be wrapped. After verifying the feasibility of calling TEP in python, we used the S-TEP process as an example, wrapped the code using the same method, and created a graphical user interface (GUI) for ease of use. The GUI for model parameter setting is shown in Figure 3.

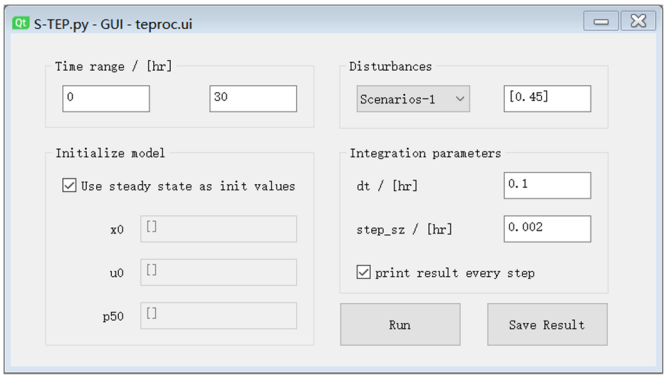


Figure 3 GUI for S-TEP.py

4. Conclusions

The existing TEP still can be further improved, such as the need to compile before use, the use of commercial software MATLAB/Simulink, the lack of standardization of process, and the weak repeatability of process simulation, etc. This paper proposes a method to connect C code in python through C-API, and successfully use python to call the core code of TEP directly. And it achieves the same accuracy as the original TEP in the open-loop test without control system. A GUI interface of the modified S-TEP program has been added to make it easier for people to use. Then a package of S-TEP process has been made by using modern scripting language Python and new object-oriented programming technology. A standardized S-TEP can promote the use of S-TEP, and make all kinds of research based on S-TEP comparable.

Next step, the original control logic of S-TEP will be transplanted into python, which will prove that python can substitute for Simulink. When the output accuracy of the python S-TEP program with control logic reaches an acceptable range, and the bundled S-TEP python module will be uploaded to the python package hosting website for everyone to use.

References

A. Bathelt, N. L. Ricker, M. Jelali, 2015, Revision of the Tennessee Eastman Process Model, IFAC-PapersOnLine, 9th IFAC Symposium on Advanced Control of Chemical Processes ADCHEM 2015, 48, 309-314.

J. J. Downs, E. F. Vogel, 1993, A plant-wide industrial process control problem, Computers & Chemical Engineering, Industrial challenge problems in process control, 17, 245-255.

A. Kroll, H. Schulte, 2014, Benchmark problems for nonlinear system identification and control using Soft Computing methods: Need and overview, Applied Soft Computing, 25, 496-513.

T. Larsson, K. Hestetun, E. Hovland, S. Skogestad, 2001, Self-Optimizing Control of a Large-Scale Plant: The Tennessee Eastman Process, Ind. Eng. Chem. Res., 40, 4889-4901.

C. Martin-Villalba, A. Urquia, G. Shao, 2018, Implementations of the Tennessee Eastman Process in Modelica, IFAC-PapersOnLine, 9th Vienna International Conference on Mathematical Modelling, 51, 619-624.

T. J. McAvoy, N. Ye, 1994, Base control for the Tennessee Eastman problem, Computers & Chemical Engineering, An International Journal of Computer Applications in Chemical Engineering, 18, 383-413.

N. L. Ricker, 1993, Model predictive control of a continuous, nonlinear, two-phase reactor, Journal of Process Control, 3, 109-123.

N. L. Ricker, 1995, Optimal steady-state operation of the Tennessee Eastman challenge process, Computers & Chemical Engineering, 19, 949-959.

N. L. Ricker, J. H. Lee, 1995, Nonlinear model predictive control of the Tennessee Eastman challenge process, Computers & Chemical Engineering, 19, 961-981.

N. L. Ricker, 1996, Decentralized control of the Tennessee Eastman Challenge Process, Journal of Process Control, 6, 205-221.

N. L. Ricker, 1995, Tennessee eastman challenge archive, URL: http://depts.washington.edu/control/LARRY/TE/download.html.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50149-4

Computational Modeling of Lube-Oil Flows in Pipelines to Study the Efficacy of Flushing Operations

Swapana S. Jerpoth, Robert P. Hesketh, C. Stewart Slater, Mariano J. Savelski, Kirti M. Yenkie*

Department of Chemical Engineering, Rowan University, Glassboro, NJ-08028, USA.
yenkie@rowan.edu

Abstract

Lube-oil industries use a complex network of pipelines for transporting thousands of high-value finished products successively in batches throughout the production plant. Each lube-oil is unique in regard to its properties, and its integrity is extremely crucial. Therefore, during a changeover operation, the lines are flushed using a high-value finished product of the current batch that is desired to be processed. The existing flushing operation typically rely on a trial-and-error procedure, resulting in the downgrading of the finished product. Moreover, it leads to enormous economic losses to the industries. In response to this problem, this work presents an approach for modeling and optimizing the flushing operation by employing first-principles and optimal control strategies. We model the flushing operation by integrating the Kendall and Monroe viscosity blending equations with time-dependent component balance equations for lube-oil pipelines. The models developed are validated against the data collected from well-designed flush-study experiments, and a good agreement is observed. We generate theoretical optimal flowrate profiles and provide insights for designing and controlling the flushing operation.

Keywords: lube-oil pipelines, flushing, optimization, optimal control

1. Introduction

In the lube-oil process industries, pipelines are considered the safest and most cost-effective mode of transportation. Therefore, complex pipeline networks are widely used for the transfer of lube-oils during various phases of manufacturing. A single network of pipelines is used for processing over thousands of different product blends. Product quality is extremely crucial to the lube-oil industries because even the slightest amount of contamination can reduce their market value and desired functionality. Hence, to ensure the integrity of every product and avoid contamination due to product mixing, pipelines are flushed between every changeover operation. The traditional flushing techniques involve the use of a finished product of the new batch for cleaning the residual oil from the previous batch. A finished product is used for flushing because using a different cleaner such as water or other oil-based fluid adds further contamination possibilities. Hence the existing flushing technique leads to the downgrading of the high-value finished product, is labor-intensive, primarily depends on a trial-and-error basis, and adds to long operational downtime. Moreover, it leads to economic losses exceeding over $1M/year and renders the product commercially unviable in the worst-case scenario. However, by integrating chemical engineering principles with process optimization techniques, the operation can be conducted more efficiently (Cafaro et al. 2015). The optimized operation could save the industry over a million dollars per annum as well as minimize material and energy consumption footprint.

Product composition must be within specifications required by the purchaser, as the systems in which these products are used (automotive engines, turbines, gears etc.) are extremely sensitive to contamination. During transportation of lube-oil in a batch changeover, a formerly processed oil mixes with the oil that is desired to be processed next. It results in the creation of a mixed oil at the interface of the two batches. The mixed oil does not match in specification to either of the two batches, hence it is classified as downgraded oil/commingled oil, and is unsuitable for the intended end-use consumer. The knowledge of physical property of the mixture is used as a quality indicator to predict whether the desired purity level of the current product batch is attained. One of the most important physical property of lube oils is viscosity. In commercial lube oil industries, samples are collected at the end of a flush period and tested for their viscosity to confirm the quality specifications prior to final packaging. This is a labour-intensive test, and results in operational downtime, which can be minimized via predictive modeling and optimization for effective flushing operation.

2. Optimal Control Problems and Applications

The flushing operation involves controlling a dynamic system, i.e., the system that evolves with time. Optimal control theory is a branch of mathematics that finds optimal ways to control dynamic systems (Sethi 2019). Here, the system refers to the pipeline flushing operation. The way of controlling the state of our system is through the oil flow rate in the flushing operation. Hence, our control variable is the flush oil flow rate. We achieve the theoretical optimal flow rate profile and report insights for designing and controlling the flushing operation. Optimal control deals with the properties of control functions, such that these functions, when inserted in differential equations, give a solution that minimizes or maximizes a performance index. In engineering applications, the control function is an intended outcome denoted in a mathematical form. The differential equations describe the dynamic response of the mechanism to be controlled and depend on the control strategy employed. The evaluation of the time-dependent operating profiles, in terms of the control variable, is used for optimizing the process performance (Yenkie and Diwekar 2012). In this work, our developed optimal control problem for predicting the optimal flushing policies is solved by using the Pontryagin's maximum principle.

3. Viscosity Blending Equations

To model the flushing operation, it is important to calculate the viscosity of the mixture of the two oils. Viscosity blending equations or mixing rules are widely used in the lubricant industries to calculate the viscosity of the mixture of lube oils (Roegiers and Zhmud 2011). One of the best-known viscosity blending equations was proposed by Kendall and Monroe (1917) (represented by equation (1)). It relies on the power law and calculates blend viscosity as the cubic-root average of the component viscosities. This equation also gives us an understanding of the mass fraction of the individual components of the blend. Through this equation, we would predict how the blend viscosity attains the desired specifications of the new oil with time.

$$\mu_{AB}^{1/3} = x_A \mu_A^{1/3} + x_B \mu_B^{1/3} \quad (1)$$

where: μ_{AB}- viscosity of blend A and B; μ_A and μ_B- viscosities of lubricants A and B;

x_A and x_B - mass fractions of lubricants A and B

Based on these fundamentals, we model the flushing operation by combining the Kendall-Monroe equation with the component balance equations for the lube oil transportation pipelines. The developed models were compared against the experimental data, and a good agreement was observed. The comparison graphs for two different products are shown in Figure 1. A comparison for thirty such changeover operations was analysed. This confirms that our developed models hold value in representing the lube oil pipeline system.

4. Modeling the Flushing Operation

Based on the fundamental first-principles, the component balance equations for the lube oil transportation pipelines are represented by equations (2) and (3). The assumptions made while developing the component balance equations are as follows: The densities of the lube oil A & B is approximately the same. Furthermore, prior to flushing the pipeline with lube oil B, it is completely filled with lube oil A.

$$\frac{dx_A}{dt} = -\frac{x_A Q}{A_C L} \tag{2}$$

$$\frac{dx_B}{dt} = \frac{x_A Q}{A_C L} \tag{3}$$

where: Q- volumetric flowrate of lubricant B; A_C - cross-sectional area; L- pipe length

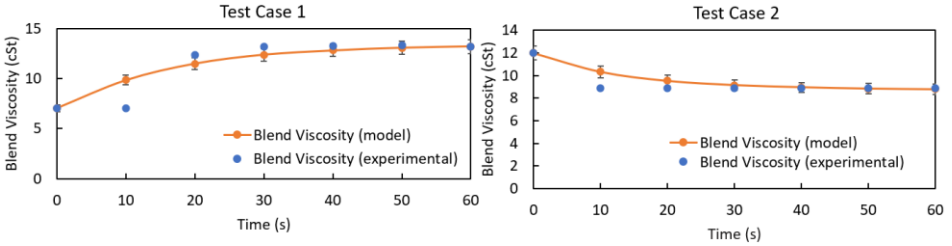


Figure 1. Validation of First Principle Mathematical Models against Experimental Data

5. Solution Methodology

Our objective is to have the pipeline completely free of the old lube oil A and just filled with the pure lube oil B. Mathematically, our objective can be formulated as minimization of the difference between the viscosity of the blend (time-dependent) and the viscosity of lube oil B (constant) at the final flush time. In other terms, the objective is to maximize the negative of the difference in viscosities of the blend and the lube oil B, as shown in equation (4).

$$\text{Max J} = -[\mu_{AB}\,(t_{final}) - \mu_B]^2 \tag{4}$$

The state of our system is controlled through the flow rate of lube oil B. Hence, the variable Q represents the control variable of the system. The process performance is determined by attaining the desired viscosity of lube oil B. Given the values of the state variables x_i $[where\ x_i = (x_A, x_B, \mu_{AB})]$ and the control variable Q at time t, the differential equations (2), (3), and (5) specify the instantaneous rate of change in the state variables. Equation (5) is obtained by differentiating equation (1) w.r.t time and substituting equation (2) and (3) in it.

$$\frac{d\mu_{AB}}{dt} = 3\mu_B^{2/3} x_A{}^3 \frac{Q}{A_C L} [1 - \mu_A^{1/3}] + 6x_A{}^2 x_B \frac{Q}{A_C L} [\mu_A^{1/3}\mu_B^{2/3} - \mu_A^{2/3}\mu_B^{1/3}] + 3\mu_B^{2/3} x_A x_B \frac{Q}{A_C L} [\mu_B^{1/3} x_B - 2\mu_A^{1/3}] \quad (5)$$

The application of the Pontryagin's maximum principle involves the introduction of additional variables known as adjoint variables and reformulation of the objective as a Hamiltonian function. Three adjoint variables 'z_i', corresponding to each of the state variables result in three more adjoint differential equations as shown in equation (7), and the Hamiltonian must satisfy equation (8). The initial values of the state, and control variable, and model parameters are: x_A= 1, x_B= 0, μ_{AB} = 12cSt, Q = 0.007 (m^3/s), μ_A= 12 cSt, μ_B= 8.7cSt, A_C = 0.019m^2, L = 6.1m

$$\frac{dx_i}{dt} = f(x_i, Q, t) \quad (6)$$

$$\frac{dz_i}{dt} = -\sum_{j=1}^{3} z_j \frac{\partial f_j}{\partial x_i} \quad (7)$$

$$H = \sum_{i=1}^{3} z_j f(x_i, Q, t) \quad (8)$$

The system results in a two-point boundary value problem since we have initial conditions for the state variables, $x_i(t_0)$= [1 0 μ_A], and final conditions for the adjoint variables, $z_i(t_f)$= [0 0 -1]. Furthermore, the total flush time is 60 seconds. For evaluating the Hamiltonian derivative, we use an analytical method proposed by Benavides and Diwekar (2013), which introduces an additional variable θ_i corresponding to each state variable and Φ_i corresponding to each adjoint variable.

$$\theta_i = \frac{dx_i}{dQ} \,\&\, \Phi_i = \frac{dz_i}{dQ} \quad (9)$$

$$\frac{d(dx_i/dt)}{dQ} = \frac{d(dx_i/dQ)}{dt} = \frac{d\theta_i}{dt} \quad (10)$$

$$\frac{d(dz_i/dt)}{dQ} = \frac{d(dz_i/dQ)}{dt} = \frac{d\Phi_i}{dt} \quad (11)$$

$$\frac{dH}{dQ} = \sum_{i=1}^{3}\left(\frac{dH}{dx_i}\right)\left(\frac{dx_i}{dQ}\right) + \sum_{i=1}^{3}\left(\frac{dH}{dz_i}\right)\left(\frac{dz_i}{dQ}\right) \quad (12)$$

The solution algorithm shown in Figure 2 starts with the initial guess of flowrate Q(t). Next, state equations are solved for the interval of t_0 to t_f using forward integration then, the adjoint equations are solved using backward integration. Next, the optimal control variable Q(t) is obtained by finding the extremum of the Hamiltonian at each time step, using the optimality condition of [|dH/dQ |] < tolerance. If the optimality condition is not satisfied, the flowrate Q(t) is updated using the gradient, such that the updated flowrate profile improves the objective function.

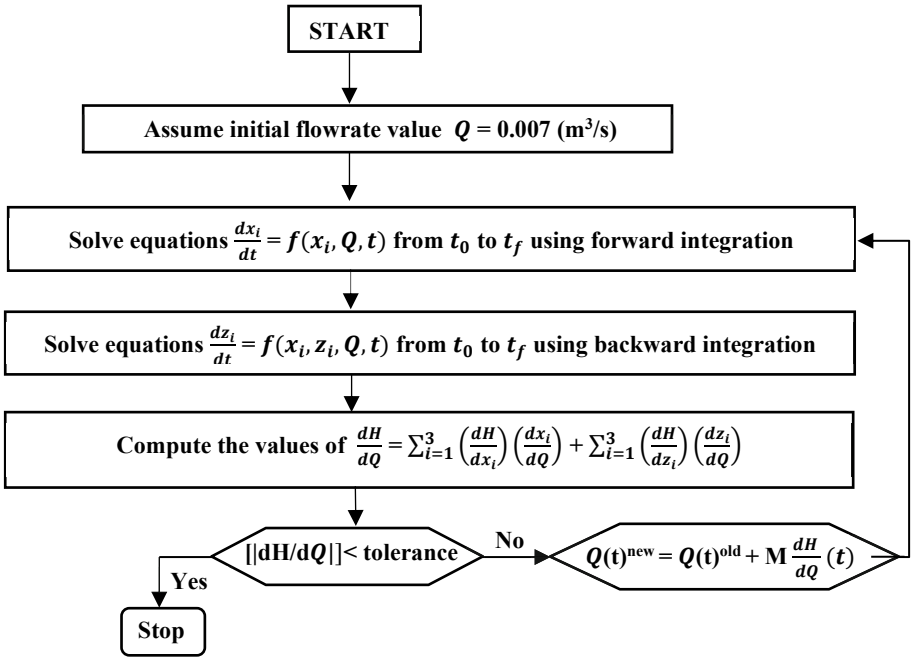


Figure 3. Flowchart of Solution Technique Using Maximum Principle Approach

6. Results and Discussions

The profiles for the derivative of Hamiltonian with respect to the control variable 'Q' is shown in Figure 4(a). It can be seen that the value of dH/dQ decreases with every iteration. The final iteration value lies within the given tolerance limit hence, we conclude the flowrate to be optimal and the corresponding flowrate profile 'Q (flow)' is shown in Figure 4(b). The variation of the state variables with respect to time is shown in Figure 5. The desired viscosity value for the lube oil B was 8.9 cSt and was attained at approximately 40 s.

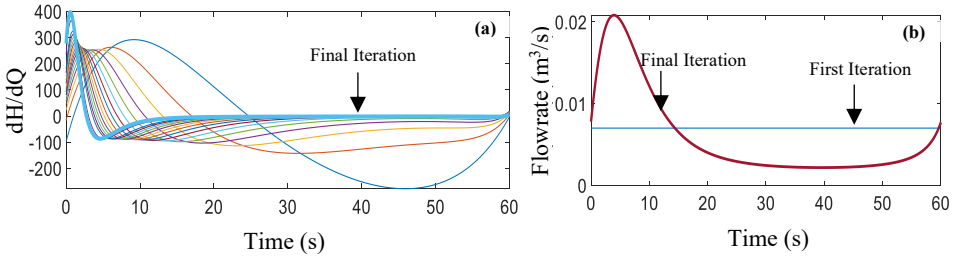


Figure 4. (a) Profiles of Hamiltonian Gradients for All Iterations (b) Profile of Control Variable (Flowrate) for Final Iteration

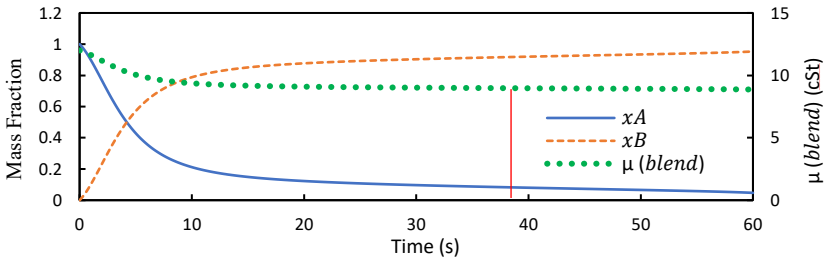


Figure 5. Variation of State Variables $[x_A, x_B, \mu_{AB}]$ with Time

7. Conclusion

In this work, the flushing operation of the lube-oil transportation pipelines was modelled by integrating the Kendall-Monroe mixing rule with component balance equations. Optimal control theory was employed in the form of the Pontryagin's maximum principle for solving the optimal flushing flowrate prediction problem. The results indicated that the optimal flushing time was 40 seconds with a modified flowrate profile. Thus, our solution approach will provide a platform for strategically optimizing the flushing operation in the lube-oil transportation pipelines and will eliminate long operational downtimes, product downgrade, and economic losses.

References

Benavides, Pahola T., and Urmila Diwekar. 2013. "Studying Various Optimal Control Problems in Biodiesel Production in a Batch Reactor under Uncertainty." *Fuel* 103 (January): 585–92. https://doi.org/10.1016/j.fuel.2012.06.089.

Cafaro, Vanina G., Diego C. Cafaro, Carlos A. Méndez, and Jaime Cerdá. 2015. "Optimization Model for the Detailed Scheduling of Multi-Source Pipelines." *Computers & Industrial Engineering* 88 (October): 395–409. https://doi.org/10.1016/j.cie.2015.07.022.

Kendall, James, and Kenneth Potter Monroe. 1917. "The Viscosity of Liquids. II. The Viscosity-Composition Curve for Ideal Liquid Mixtures." *Journal of the American Chemical Society* 39 (9): 1787–1802. https://doi.org/10.1021/ja02254a001.

Roegiers, Michel, and Boris Zhmud. 2011. "Property Blending Relationships for Binary Mixtures of Mineral Oil and Elektrionised Vegetable Oil: Viscosity, Solvent Power, and Seal Compatibility Index." *Lubrication Science* 23 (6): 263–78. https://doi.org/10.1002/ls.154.

Sethi, Suresh P. 2019. *Optimal Control Theory: Applications to Management Science and Economics*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-98237-3.

Yenkie, Kirti M., and Urmila Diwekar. 2012. "Stochastic Optimal Control of Seeded Batch Crystallizer Applying the Ito Process." *Industrial & Engineering Chemistry Research*, June, 120604103933002. https://doi.org/10.1021/ie300491v.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50150-0

Comparison of ammonia synthesis plants of different scale with a dynamic model

Laron Burrows, George Bollas*

Department of Chemical & Biomolecular Engineering, University of Connecticut, 191 Auditorium Road, Unit 3222, Storrs, CT 06269, USA.
geroge.bollas@uconn.edu

Abstract

Distributed ammonia production has received a lot of attention as a means of converting stranded fossil resources to a valuable product or for energy storage. However, little attention is given to the potentially problematic pairing of a process sensitive to disturbances with a feedstock variable in rate and composition. The purpose of this work is to explore the stability of a small-scale ammonia production system and thus the viability of distributed ammonia production with processes and chemistries similar to those of large-scale plants. A state-of-the-art large-scale ammonia production system is modelled in Aspen Plus® and converted to a flow driven model in Aspen Dynamics® to serve as a reference model. The reference model is then scaled down to 10% of its original capacity using constant length to diameter ratios for the reactors, and both models are subjected to disturbances in reactor temperature, reactor feed flow rate, and natural gas feed. In the small-scale model, a 5% pulse in the natural gas flow rate is shown to cause 100 °C swings in reactor temperature and 550 kmol/h swings in reactor flow rate, while the reference model remains stable. We conclude with an outlook of the need of decoupled ammonia synthesis pathways that allow for independent control of the extent of reactions and therefore their heat generation or consumption.

Keywords: Distributed Chemical Manufacturing, Ammonia Production, Haber-Bosch

1. Introduction

In recent years, there has been an increasing interest in the concept of using distributed ammonia production systems to convert stranded resources such as wind, biomass, and natural gas into ammonia-based fertilizers for local markets (Arora et al., 2016). Using stranded resources as a feedstock has the advantages of proximity to local markets, low acquisition costs and, in the case of wind, being available in a virtually infinite supply (Allman and Daoutidis, 2016). However, commingled with these advantages are disadvantages of the variable quality, intermittency, and volatility associated with distributed resources. While some chemical manufacturing processes are resilient enough to withstand volatile feedstocks, there is sufficient evidence to suggest that ammonia production processes may not be robust and stable to handle variability and disturbances in the process boundaries (Morud and Skogestad, 1998).

Distributed ammonia production proposals generally attempt to replicate the Haber-Bosch process at small scale, fed by a stranded resource, despite evidence that small changes in the Haber-Bosch process conditions can lead to system failures (Morud and Skogestad, 1998; Gullberg, 2018). Although considerable attention is being given to assessing the economic feasibility (Arora et al., 2016), optimal location (Allman and

Daoutidis, 2016), and coproduction options for distributed ammonia production systems (Andersson and Lundgren, 2014), little is being done to assess the difficulty in pairing of a temperamental process with a sporadic feedstock at a small scale.

Simulations are a powerful tool for assessing the feasibility of a proposed process and are the primary method used to evaluate distributed ammonia production proposals. Ammonia synthesis reactors exhibit multiple steady states and may move between steady states if perturbed (Morud and Skogestad, 1998), thus necessitating the need for dynamic as well as steady state simulations to accurately assess the feasibility of new process concepts. Nevertheless, the literature is a medley of distributed ammonia production proposals done at steady state without regard for process dynamics, stability and robustness to uncertainty. Arora et al. (2016) and Andersson and Lundgren (2014) investigated the techno-economic potential of a distributed ammonia production process fed by biomass at steady state using reactors modelled as Gibbs reactors, without studying the potential for process instabilities. Conversely, Gullberg (2018) studied the instability of a small-scale ammonia reactor in a steady state and dynamic simulation and showed that small changes in reactor feed temperature, pressure, or flow can induce oscillatory behavior and instability. Gullberg's study concluded with a control structure to reduce the oscillatory behavior brought on by process changes; however, the reactor was studied in isolation and the effects of feedstock changes upstream of the reactor were not evaluated. To date, scant attention has been paid to the potential of process instability jeopardizing distributed ammonia production proposals; thus, motivating this study into the effects of feedstock variation on the process stability of a small-scale ammonia production system.

In this study, a state-of-the-art ammonia production system fed by stranded natural gas was modelled in Aspen Plus® with the capacity of 1300 tNH_3/d and validated against industrial data to serve as a reference model. The reference model was then scaled down to 130 tNH_3/d using standard scaling principles for the reactors and heat exchangers. Both models were then converted to dynamic models in Aspen Dynamics® and subjected to disturbances in reactor feed temperature, reactor feed flow and natural gas flow. The study concludes with a stability comparison at each scale.

2. Methodology

2.1. Model Description

Figure 1 shows the block diagram of the model used in this work segmented into the following three sections: synthesis gas generation, gas purification, and ammonia synthesis. Unit operations in the synthesis gas generation and gas purification sections of the model were adapted from the ammonia plant model (Aspen Technology, 2008) and amine absorber model (Aspen Technology, 2013) included in the Aspen Plus® software examples respectively. The ammonia synthesis model was adapted from Morud and Skogestad (1998) using a kinetic model for a modern catalyst, and a 3-bed adiabatic plug flow reactor system with intermediate quench. The pressure drop across each bed is calculated by the Ergun equation and heat transfer to the catalyst is allowed with a catalyst heat capacity of 1100 J/kg-K and bulk density of 2200 kg/m^3.

The process modelled begins with the combustion of a split stream of natural gas to produce steam and preheat the feed to the primary reformer to 700 °C. The primary

reformer is fed with a steam-to-carbon ratio of 3/1 at 2.5 MPa to steam-reform methane and produce hydrogen. The products of the primary reformer are sent to the secondary reformer where air is introduced to combust natural gas in an autothermal reformer design. The products of the secondary reformer are then cooled to 450 °C using the process feed water before entering the water gas shift reactors (high temperature shift, HTS and low temperature shift, LTS). Next, the process gas is fed to the gas purification section, where CO_2 is removed using an amine absorber adapted from the model developed by Aspen Technology (2013). Downstream the absorber, the process gas is passed through the methanator to remove unreacted CO and CO_2 from the process stream and convert it to CH_4 before being compressed to 21 MPa and fed to the ammonia reactor loop at a H_2/N_2 ratio of 3/1 at 250 °C. Finally, the products of the ammonia reactor are cooled to -10 °C to recover ammonia as a liquid product and the unreacted H_2 and N_2 is recycled to the reactor to increase NH_3 yield.

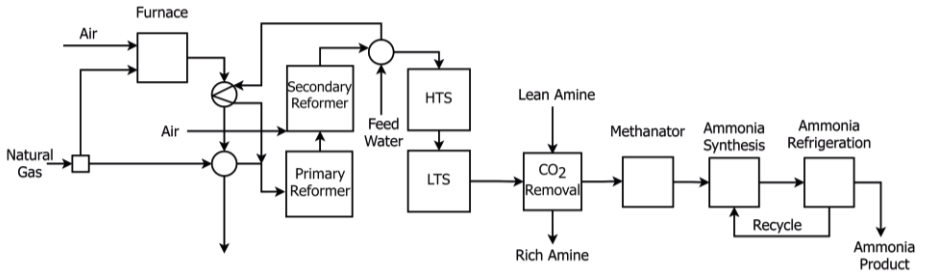


Figure 1. Ammonia production flow sheet.

2.2. Model Validation and Downscaling

The reference model was validated in three steps: first, individual components were validated against literature data; then, the ammonia loop was validated in a dynamic environment; and finally, the full model was validated against data from similar-sized plants. After the reference model was validated, it was scaled down from 1300 tNH_3/d to 130 tNH_3/d using fixed length to diameter ratios for the major reactors. Heat exchangers used in the reference model were scaled to 10 % of their heat exchange area in the small-scale model, and columns were scaled using fixed height to diameter ratios.

2.3. Dynamic Simulations and Disturbance Testing

The steady state reference model and small-scale model were converted to flow driven dynamic models in Aspen Dynamics®. The amine absorbers of each model were replaced with separators of equivalent CO_2 removal efficiency, due to the lack of support for rate-based absorption in Aspen Dynamics®. Additional modifications were made to heat exchangers and valves used in the model according to Aspen Plus® recommendations to improve model robustness.

The following studies were performed on the open-loop of the models in Aspen Dynamics®: (a) Haber-Bosch (HB) reactor stability analysis; (b) and full plant stability analysis. In the HB reactor stability analysis, HB reactors were studied in isolation to explore the effect of size on their stability. Each HB reactor was fed with a feed of H_2, N_2, and NH_3 at a mole fraction of 0.73, 0.23, and 0.04, respectively, at 250 °C and 20.0 MPa. The following disturbances were introduced; (a) a reactor feed temperature pulse of +20 °C and -20 °C for 1h; (b) a reactor feed flow pulse of +23.1 % and -23.1 % for 1 h; and (c) a reactor feed flow noise modelled as a sine wave with amplitude 23.1 % and

a period of 30 minutes for 8 h. In the full plant stability analysis, the HB reactor stability was examined in response to a +5 % bump in natural gas flow to the ammonia process.

3. Results and Discussion

3.1. Model Validation

A summary of the model tuning and validation is presented in Table 1, in terms of reactor outlet composition of the primary reformer (PR), the high temperature shift reactor (HTS), the methanator (MTH), and the ammonia reactor (AM) against their respective reference values, marked as Ref. Specifically, the PR outlet was compared to the data of an industrial primary reformer (Tran et al., 2017), the HTS was compared to data by Fabian et al. (2020), the MTH was compared to data by Er-Rbib and Bouallou (2014), and the ammonia reactor outlet was compared with the industrial plant data provided in Appl (2011). Overall, each process component was tuned to be in excellent agreement with literature data.

Table 1: Component validation results.

	CH_4	H_2	CO	CO_2	H_2O		CH_4	H_2	N_2	NH_3	Ar
PR	3.7	45	5.7	7.6	37.6	MTH	13	25	-	-	49.9
Ref	3.8	47	5.7	8.8	34.3	Ref	12	24.8	-	-	50.5
HTS	-	34	7.1	26.9	31.7	AM	5	57	18	19	1
Ref	-	35	6.5	27.4	31.5	Ref	6	56	18	15	2

3.2. Disturbance Testing

The Haber-Bosch (HB) reactors of the reference model and small-scale model were subjected to disturbances in feed temperature and flow rate, and an excerpt of disturbance and response plots is shown in Figure 2. First, the HB reactors were subjected to a -20 °C pulse in reactor feed temperature, which resulted in limit cycle behavior in both models in agreement with the literature (Morud and Skogestad, 1998; Gullberg, 2018) shown in Figure 2(a) and Figure 2(b) for the reference model and small-scale model respectively. After the pulse, the reference model returned to steady state while the small-scale model sustained oscillations of 10 °C for over 30 simulated hours. Next, the HB reactors were subjected to a +20 °C pulse in reactor feed temperature and a -23.1 % reduction in reactor feed flow rate, where no process instabilities were observed in agreement with Gullberg (2018). Thereafter, the HB reactors were subjected to a +23.1 % pulse in reactor feed flow rate, which resulted in reaction extinction in both models, as evident by the rapid drop in reactor outlet temperature shown in Figure 2(c) and Figure 2(d) for the reference model and small-scale model, respectively. The immediate increase in flow rate to the system began to induce the limit cycle behavior shown in Figure 2(c) and Figure 2(d), however the sustained increase in flow rate resulted in a decrease in reactor feed temperature beyond limit cycle to reactor extinction. Lastly, the HB reactors were subjected to a sinusoidal disturbance in feed flow rate, which resulted in sustained oscillations of 7 °C in the reference model and reaction extinction in the small-scale model.

In testing the full plant model at different scales, a +5 % bump in natural gas flow rate resulted in a HB reactor outlet temperature decrease of 10 °C in the reference model, shown in Figure 2(e), and large oscillations in the HB reactor of the small-scale model,

shown in Figure 2(f). In response to the +5 % bump in natural gas flow rate, the HB reactor outlet temperature of the small-scale model oscillated within 20-100 °C for over 20 simulated hours, with a corresponding flow rate oscillation between 100-550 kmol/h. The precipitous changes experienced by the small-scale model could plausibly result in system failure similar to those reported by Morud and Skogestad (1998).

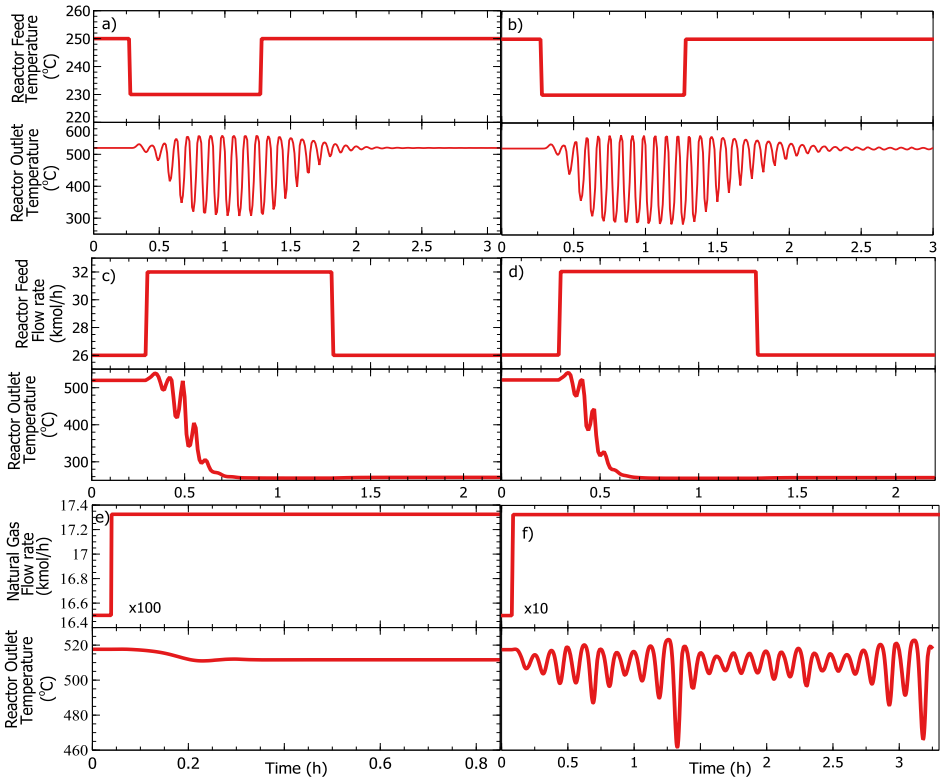


Figure 2. Disturbance and response plots for: a -20 °C pulse in HB reactor feed in the reference model (a), and small-scale model (b); a +23.1 % pulse in HB reactor flow in the reference model (c), and small-scale model (d); and a +5 % bump in natural gas flow to the reference model (e), and small-scale model showing the response in HB reactor outlet.

In previous studies, it has been shown that decreasing the HB reactor feed temperature can lead to limit cycle behavior (Morud and Skogestad, 1998), in this study we show that an increase in natural gas flow rate to the production plant can also lead to such behavior, seen in Figure 2(f). Evidently, the behavior shown in Figure 2(f) is also due to a reduction in HB reactor feed temperature. As the stream of higher flow rate enters the preheater of the HB reactor, it exchanges heat with the smaller reactor outlet flow, and is not sufficiently heated to the desired operating temperature. Thereafter, the decreased inlet temperature increases the reaction rate of the forward exothermic reaction, which in turn increases the reactor bed temperature, until the reverse endothermic reaction dominates, and the reactor bed temperature is decreased. As the reactor beds cycle between the forward exothermic reaction and the reverse endothermic reaction of the Haber-Bosch process, the effects are transferred through the reactor by the intermediate quenches and preheater, creating a feedback loop of snowballing effects (Morud and Skogestad, 1998).

Ammonia synthesis reactors operate with a small margin of stability, which this study suggests decreases with system size. To improve ammonia production stability at small scale the following could be explored: (a) a reactor design with no intermediate quenches or heat integration, thus eliminating the ability for snowball effects to permeate through the system at the expense of ammonia yield and potentially economic feasibility; (b) advanced control structures such as those proposed by Gullberg (2018) or Araújo and Skogestad (2008); and (c) a chemical looping reactor (Burrows et al., 2021), to decouple ammonia synthesis and eliminate the potential of the exothermic forward reaction competing with the endothermic reverse reaction.

4. Conclusion

In this study, ammonia production stability was explored at different scales to assess the impact of stranded resource variability on small-scale ammonia production. A state-of the-art ammonia production process was modelled, scaled down and subjected to disturbances in reactor feed temperature, flow rate and natural gas feed to the process. It was found that the reference model is more robust than the small-scale model, when subjected to the same disturbances. Interestingly, after only a 5% increase in natural gas flow rate, the small-scale model exhibited large oscillations in the ammonia reactor temperature and flow, whereas the reference model remained stable. Future distributed ammonia production work should focus on improving the stability of the ammonia synthesis reactor by: reducing the ability of large oscillations to permeate through the reactor, employing advanced controls, or decoupling the ammonia synthesis reaction using intermediates that assist with N_2 or H_2 fixations, such as those in chemical looping.

References

W. A, Allman, P. Daoutidis, 2016, Ammonia supply chains: A new framework for renewable generation with a case study for Minnesota. Comput. Chem. Eng. 38, 1395-1400.
J. Andersson, J. Lundgren, 2014, Techno-economic analysis of ammonia production via integrated biomass gasification. Appl. Ener. 130, 484-490.
M. Appl, 2011, Ammonia, 2. Prod. Pro. Ull. Enc. Ind. Chem. 2, 2019-211, Dan-Sch, Germany.
A. Araújo, S. Skogestad, 2008, Control structure design for the ammonia synthesis process, Comput. Chem. Eng. 32: 2920-2932.
P. Arora, A. Hoadley, S.M Mahajani, A. Ganesh, 2016, Small-Scale Ammonia Production from Biomass: A Techno-Enviro-Economic Perspective, Ind. Eng. Chem. Res. 55, 6422-6434.
L. Burrows, P. Gao, G.M. Bollas, 2021, Thermodynamic feasibility analysis of distributed chemical looping ammonia synthesis, Chem. Eng. J, 426, 131421.
H. Er-Rbib, C. Bouallou, 2014, Methanation Catalytic Reactor, Comp. R. Chemie. 17 (7): 701–6.
R. Gullberg, 2018, Controllability Analysis of Ammonia Synthesis Loops, Ph.D. Thesis, Norwegian University of Science and Technology, Trondheim, Norway.
J. Morud, S. Skogestad, 1998, Analysis of instability in an industrial ammonia reactor, AIChE J., 44, 888-895.
F. Rosner, A. Rao, S. Samuelsen, 2020, Water gas shift reactor modelling and new dimensionless number for thermal management/design of isothermal reactors, Appl. Ther. Eng. 173: 115033.
Aspen Technology, 2008, Aspen Plus® Ammonia Plant Model, Bedford MA, USA.
Aspen Technology, 2013, Aspen Plus® Rate Based Absorption Model, Bedford MA, USA.
A. Tran, A. Andres, H. Durand, M. Crose, P. D. Christofides, 2017, CFD Modeling of a Industrial-Scale Steam Methane Reforming Furnace, Chem. Eng. Science 171: 576–98.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50151-2

Simulation Analysis of Gas Feed Method for Development of Ru-Based Catalyst for Ammonia Production

Hideyuki Matsumoto[a*], Masashi Kikugawa[b], Anthony Basuni Hamzah[a], Marie Ishikawa[c], Yoshihiro Goto[b], Shinichi Ookawara[a], Yuichi Manaka[d], Masayasu Nishi[e], Tetsuya Nanba[d]

[a] *Department of Chemical Science Engineering, Tokyo Institute of Technology, Tokyo 152-8550, JAPAN*
[b] *Toyota Central R&D Laboratories, Aichi 480-1192, JAPAN*
[c] *Toyota Motor Corporation, Aichi 471-8571, JAPAN*
[d] *Renewable Energy Research Center, AIST, Fukushima 963-0215, JAPAN*
[e] *Energy Process Research Institute, AIST, Ibaraki 305-8569, JAPAN*
**Corresponding Author's E-mail: matsumoto.h.ae@m.titech.ac.jp*

Abstract

In the present study, we investigate a new strategy for design of modular, energy-efficient green ammonia production system, which is equipped with a reactor system utilizing novel Ru-based catalyst and novel heat management. In order to maximize the performance of the reactor, which employs two stages consisting of Fe-based (first stage) and the new Ru-based (second stage) catalysts, we propose a split feed method for main feed gas, of which a portion of the feed gas serves as quench gas at inlet of the second stage. Through sensitivity analysis, we demonstrate that split feed method for nitrogen gas could achieve higher process efficiency than other feed methods.

Keywords: Process Synthesis; Green ammonia; Process simulation; Optimization.

1. Introduction

To realize the net-zero emission of carbon dioxide by 2050, a supply chain of green ammonia (NH_3) needs to be established. A key component of such supply chain would be modularized, decentralized, small- and medium-scale plant for production of NH_3 (10-100 ton-NH_3/day) from renewable-derived hydrogen. It is necessary to overcome some key problems, namely high specific energy consumption, harsh operating condition, and slow response towards rapidly-fluctuating operating parameters. These challenges have to be addressed with development of new catalysts, as well as novel process design tailored to realize their potentials.

Ruthenium (Ru) catalysts have been widely studied and utilized in industrial scale due to its high activity under milder operating temperature (~400 ºC) and pressure (<12 MPa). In addition, Ru-based catalysts are considered to be superior to the start-up performance in comparison with conventional iron-based catalysts. On the other hand, hydrogen poisoning has been identified as major obstacle for application of Ru-based catalysts, and therefore various types of catalyst supports have been put forward in the literature. Our research group has developed new Ru/lanthanoid oxide catalysts that performed high NH_3 yield and less hydrogen poisoning under high pressure conditions.

From process system design perspective, unique optimum operating condition window of NH_3 synthesis over such type of catalysts, which includes lower H_2/N_2 ratio (1.0 – 1.5) and temperature (380 – 400 °C) than that of conventional catalysts, warrants specific attention and eventually new strategy for the intensification of the process. This is achievable by adjusting the gas feed condition, however, relative complexity of the process itself (e.g. unreacted gas recycle and multi-bed configuration of reactors) presents challenge towards its implementation.

In the present study, we introduce and demonstrate a new strategy for design of modular, energy-efficient green NH_3 production system, which utilizes integration of reactor system equipped with novel Ru-based catalyst and internal quench gas feed, as well as novel heat management. To the authors' knowledge, there is a scarcity of publications and studies in regards of gas feed methods specifically designed for the above-mentioned Ru/lanthanoid oxide-type catalysts. Therefore, a purpose of the present contribution is to quantitatively clarify the effect of the gas feed method on the process efficiency for the overall process system using the catalyst by process simulation.

2. Derivation of reaction rate equation for developed Ru-based catalyst

Various reaction rate models based on the Langmuir-Hinshelwood mechanism have been proposed in the previous literature on NH_3 synthesis reactions using Ru-based catalysts (Aika, 2017; Y. Kobayashi et al., 2017). Most reaction rate equation models have been derived assuming that the reaction rate determining step (RDS) is the dissociative adsorption of nitrogen. In addition, various assumptions have been made regarding the adsorption of reaction intermediates on the surface of the catalyst, and it was considered difficult to deductively determine the RDS for Ru/lanthanoid oxide catalysts that we have developed. Therefore, in derivation of the reaction rate equation model for our developed catalyst, RDS is also assumed to be dissociative adsorption of nitrogen. Furthermore, all possible reaction intermediates, namely H*, N*, and nitrogen hydrates groups NH_x*, have to considered for the surface coverage balance.

Table 1 Elementary reactions and reaction rate equations

Reaction		Rate expression	Eq.
A-1	$N_2(g) + * \leftrightarrow N_2*$	$\theta_{N2*} = k_{A,1}\theta_* P_{N2}$	(1)
A-2	$N_2* + * \leftrightarrow 2N*$	$r = k_{A,2,f}\theta_*\theta_{N2*} - k_{A,2,b}\theta_{N*}{}^2$	(2)
B-1	$N* + H* \leftrightarrow NH* + *$	$\theta_{NH*} = k_{B,1}{}'\frac{\theta_{N*}\theta_{H*}}{\theta_*}$	(3)
B-2	$NH* + H* \leftrightarrow NH_2* + *$	$\theta_{NH2*} = k_{B,2}{}'\frac{\theta_{NH*}\theta_{H*}}{\theta_*}$	(4)
B-3	$NH_2* + H* \leftrightarrow NH_3* + *$	$\theta_{NH3*} = k_{B,3}{}'\frac{\theta_{NH2*}\theta_{H*}}{\theta_*}$	(5)
B-4	$NH_3* \leftrightarrow NH_3(g) + *$	$\theta_* = k_{B,4}\frac{\theta_{NH3*}}{P_{NH3}}$	(6)
C-1	$H_2(g) + 2* \leftrightarrow 2H*$	$\theta_{H*} = \theta_*\sqrt{k_{C,1}{}'P_{H2}}$	(7)

Table 1 shows the assumed elementary reactions and their respective reaction rate equations in deriving the reaction rate equation model based on the Langmuir-Hinshelwood mechanism. Reactions A-*i*, B-*i* and C-*i* represent the dissociative adsorption of nitrogen, the formation of nitrogen hydrides (NH_x), and the dissociative adsorption of hydrogen, respectively. The balance of the coverage of the reaction intermediate on the catalyst surface species was assumed to be Eq. (8).

$$\theta_{\mathrm{H}*} + \theta_{\mathrm{N}*} + \theta_{\mathrm{NH}*} + \theta_{\mathrm{NH2}*} + \theta_{\mathrm{NH2}*} + \theta_{\mathrm{NH3}*} + \theta_{*} = 1 \tag{8}$$

The reaction rate equation based on the Langmuir-Hinshelwood mechanism, which can be derived from Eqs. (1) - (8), was shown in Eq. (9). The second to sixth terms of the denominator of the Eq. (9) represent the reaction suppression by adsorption of H*, N*, NH*, NH_2*, and NH_3*, respectively. The Arrhenius-type estimation equations for the forward reaction rate constant k_f and the adsorption constant K_j are Eq. (10) and Eq. (11), respectively.

$$r = \frac{k_f P_{N2} - \left(\frac{k_f}{K_{eq}}\right)\frac{{P_{NH3}}^2}{{P_{H2}}^3}}{\left(1 + K_{H*}{P_{H2}}^{0.5} + K_{N*}\frac{P_{NH3}}{{P_{H2}}^{1.5}} + K_{NH*}\frac{P_{NH3}}{{P_{H2}}^{1}} + K_{NH2*}\frac{P_{NH3}}{{P_{H2}}^{0.5}} + K_{NH3*}P_{NH3}\right)^2} \tag{9}$$

$$k_f = A_f \, exp\left(\frac{-E_{a,f}}{RT}\right) \tag{10}$$

$$K_j = A_j \, exp\left(\frac{Q_{j,ads}}{RT}\right) \tag{11}$$

By using experimental data for the catalytic activity test, the parameters in Eqs. (9) to (11) based on the isothermal plug flow reactor (PFR) model were estimated. Figure 1 shows the results of parameter estimation. The plots in Figure 1 represent the activity test data, and the solid lines (red and blue) represent the simulation data based on estimated values for the parameters. In the region for operating condition far from the equilibrium concentration (dash line in the figure), an acceptable agreement was seen between the simulation data and the activity test data. Hence, by considering not only H* and N* but also the reaction intermediate NH_x* for the balance of coverage of the reaction intermediate on the surface of the catalyst / support, a reaction rate equation with higher predictive performance could be derived.

The estimated values of parameters in the reaction rate equation model based on the Langmuir-Hinshelwood mechanism were analyzed afterwards. The value of activation energy $E_{a,f}$ in the reaction rate constant of the forward reaction (Eq. (10)) was estimated to be 21.0 kJ/mol. It was found that the $E_{a,f}$ value was slightly lower than the activation energy (~ 24-50 kJ/mol) of the Ru-based catalyst using an alkali oxide as a carrier reported in the previous literatures (Dahl et al., 2000; Siporin and Davis, 2004). It was considered that the support of lanthanoid oxide exhibited higher degree of electron

donation to ruthenium, which resulted in the cleavage of nitrogen triple bonds at lower energy levels.

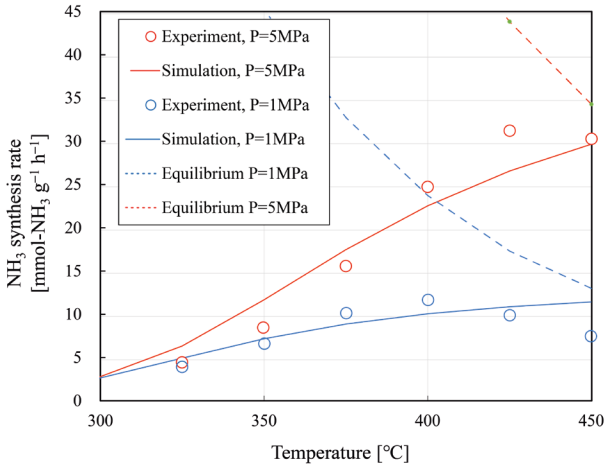


Figure 1 Comparison of experimental data and simulation data of NH_3 synthesis rate for Ru/lanthanoid oxide catalyst (Operating pressure P = 1 MPa, 5 MPa; H_2/N_2 = 3.0).

The orders of value of the adsorption equilibrium constants of the reaction intermediates calculated from the estimated values of the parameters was as follows.

$K_{NH3*} < K_{NH*} < K_{H*} < K_{NH2*} < K_{N*}$

It was considered that the adsorbed hydrogen H* was certainly an important intermediate for the balance of coverage of absorbed species, since it depended on the partial pressure of hydrogen, which was higher than partial pressure of ammonia under operating conditions.

3. Simulation analysis for application of methods for split feed of gas

Upon obtaining the kinetic rate equation of the synthesis reaction, process simulation of a small-scale NH_3 production system (250 kg-NH_3/h) operating under low pressure conditions ($\leq$ 8 MPa) was carried out to investigate favorable operating conditions, consisting of inlet composition and flow ratio of quench gas (N_2 or H_2). Schematic diagram of the system is seen in Figure 2, which shows the dual-stage reactor configuration. In the present paper, a conventional Fe-based catalyst and the Ru/lanthanoid oxide catalyst developed in the present study were filled in the first stage (hereinafter, "Bed 1") and the second stage (hereinafter, "Bed 2") of the reactor, respectively. In addition, heat integration of the entire system was investigated. An in-situ heat removal system was adopted for both of two stages in the reactor, and a system of heat exchange of the reactor inlet gas with reactor outlet gas and recycled gas was also applied in this simulation.

The inlet temperature of the Bed 1 of the reactor was set at 673 K (400 °C). A quench flow (573 K) of N_2 or H_2 between Bed 1 and 2 was added not only to shift the Bed 2 inlet composition to favorable ratio, but also to decrease the Bed 1 outflow to favorable temperature of 673 K. Variable parameters in this process simulation include the

composition (H_2/N_2 ratio) and ratio of flow rate of quench gas to the main feed. The system performances were evaluated by using two indices: process efficiency ($\eta_{process}$) and specific energy consumption (SEC).

$$SEC = \frac{System\ energy\ consumption}{NH_3\ production} \tag{12}$$

$$\eta_{process} = \frac{Q_{NH_3\ combustion}}{Q_{H_2\ combustion} + energy\ consumption} \times 100 \tag{13}$$

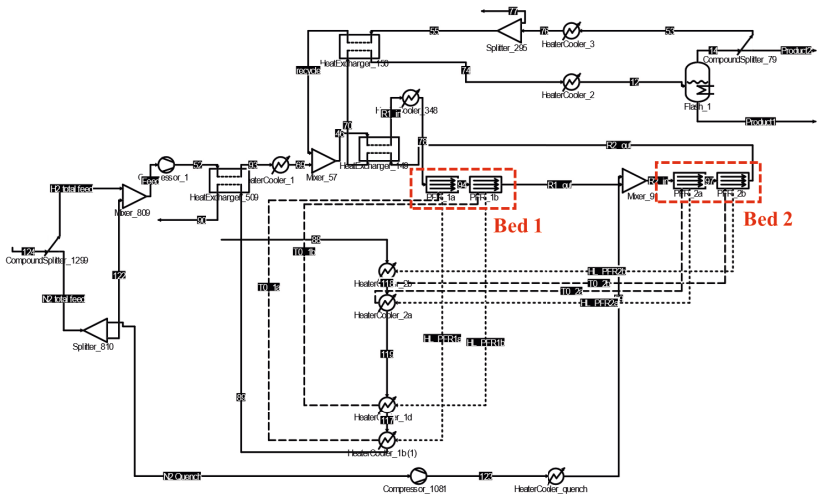


Figure 2 Process flow diagram of the NH_3 production system

Figures 3 and 4 show results of sensitivity analysis of composition of the main feed gas and the flow ratio of quench gas to the main feed with respect to the process efficiency and the SEC. It was seen that process operation with high H_2/N_2 ratio of ~2.75 at the inlet of Bed 2 adjusted by H_2 (or N_2) quench with medium rate (50% of total feed of H_2 or N_2) was preferable. Interestingly, when the H_2/N_2 ratio of the main feed was constant, change in values for the above evaluation indices with respect to the operating conditions for quenching (gas type and feed ratio) was extremely small. Through the process simulations, the optimum performance of the evaluated process was ~61% and ~ 7500 kJ / kg-NH_3 for process efficiency and SEC, respectively.

In addition, the ratio of each energy consumption of the reactor, separator, etc. to that of the entire system was calculated for each case using split quench gas of hydrogen and nitrogen. The proportion of energy consumption associated with the separation of NH_3 and the recycling of separated gas was also seen to be significant, and thus it is considered that reduction of the energy consumption for NH_3 separation / recovery unit will continue to be an important issue, which will be subject in further study.

4. Conclusions

A novel small-scale green ammonia synthesis process system design was proposed herein, which employs dual-stage reactor with Fe- and newly-developed Ru/lanthanoid

oxide catalysts and split feed method. For the development of the process system model, the reaction rate equation for ammonia synthesis over the novel Ru catalyst was derived based on the activity test data collected under realistic operating pressure conditions. To optimize the system performance, we proposed a split feed method for main feed gas to utilize as quench gas at inlet of Bed 2. Through sensitivity analysis by process simulation, it was demonstrated that the split feed method for nitrogen gas could achieve higher process efficiency than other feed methods.

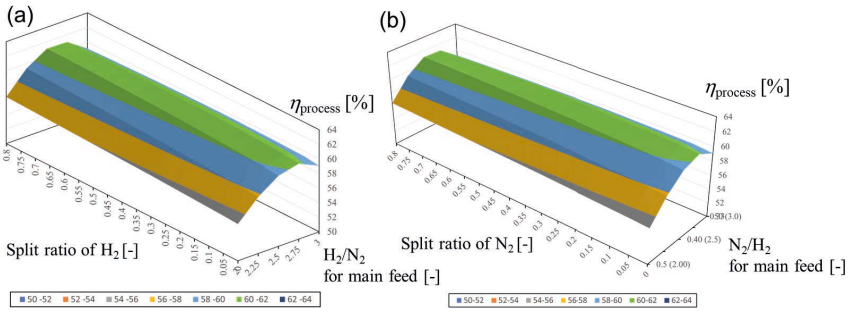


Figure 3 Sensitivity analysis of operating conditions to the process efficiency; (a) H_2 and (b) N_2 quench

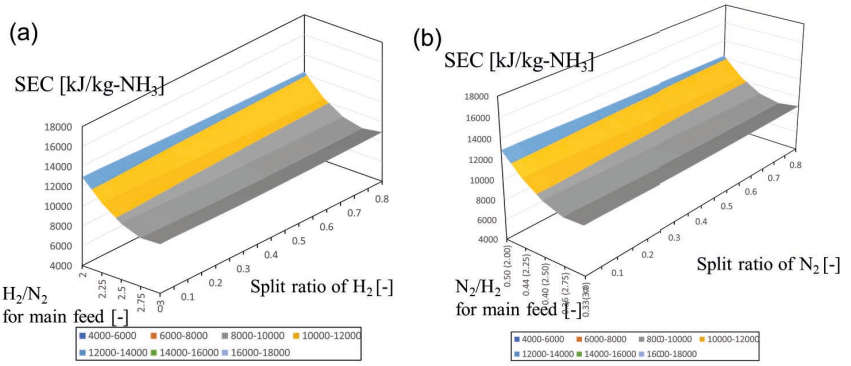


Figure 4 Sensitivity analysis of operating conditions to the SEC; (a) H_2 and (b) N_2 quench

References

K. Aika, 2017, Role of alkali promoter in ammonia synthesis over rutheniumcatalysts—Effect on reaction mechanism, Catalysis Today, **286**, 14–20.

S. Dahl et al., 2000, Surface science based microkinetic analysis of ammonia synthesis over ruthenium catalysts, Journal of Catalysis, **192**(2), 391-399.

Y. Kobayashi et al., 2017, Kinetic evidence: the rate-determining step forammonia synthesis over electride-supported Rucatalysts is no longer the nitrogen dissociation step, Catal. Sci. Technol., **7**, 47–50.

S.E. Siporin and R.J. Davis, 2004, Use of kinetic models to explore the role of base promoters on Ru/MgO ammonia synthesis catalysts, Journal of Catalysis, **225**(2), 359-368.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50152-4

Estimation of the effect of liquid viscosity on gas-liquid mass transfer in a bioreactor using CFD-PBM coupled model

Young Seok Bak[a] and Jong Min Lee[a*]

[a]*School of Chemical and Biological Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, KOREA*
jongmin@snu.ac.kr

Abstract

Stirred tanks are widely used as aerated fermenters for cell cultivation. Mass transfer inside the bioreactor is important for reactor control and design, as it directly affects productivity. Even if mixing and aeration are considered sufficient, insufficient local oxygen mass transfer may occur due to the highly inhomogeneous nature of reactor hydrodynamics. Furthermore, it is well known that the increase in viscosity greatly reduces the amount of oxygen mass transfer, which is inevitable as cell cultivation progresses. In order to analyze the effect of viscosity on mass transfer of bioreactor , this work proposes a modeling and computational framework that combines CFD (Computational Fluid Dynamics) and PBM (Population Balance Model). A 6-blade Rushton tank was selected as the target system, and the model was validated with the measurement results at the nominal water-viscosity, 1cp. Then, the change in mass transfer rate was analyzed. As the viscosity increases, the mass transfer rate gradually decreases, and the level of decrease was consistent with the empirical correlations.

Keywords: Bioreactor, Computational fluid dynamics, Population balance model, Gas-liquid mass transfer, Liquid viscosity

1. Introduction

Control and design of bioreactors aim to maximize productivity by creating an optimized environment for cell cultivation. However, due to the lack of understanding of reactor hydrodynamics, maintaining an optimal environment for bioreactors with various scales remains a challenging task.

Oxygen is one of the most important materials required by cells. However, the inhomogeneous nature of the hydrodynamics of the stirred tank bioreactor creates the regions with locally deficient oxygen supply. Traditionally, the mass transfer rate between gas and liquid has been estimated from the interfacial area and mass transfer coefficient. Since these two factors are influenced not only by the hydrodynamics of the continuous phase, the liquid, but also by the dynamics of the disperse phase, the bubble, a comprehensive simulation of both phases is required. Using the population balance model (PBM), it is possible to describe the bubble dynamics such as coalescence and breakage.

In this study, an integrated model framework consisting of both CFD and PBM is suggested to reliably predict the oxygen mass transfer rate inside the bioreactor. As a 6-blade Rushton turbine tank system was targeted, the model was validated through

measurement results using the viscosity of water. Then, the change in oxygen mass transfer rate was further analyzed by increasing the viscosity tenfold.

2. Methodology

The Eulerian-Eulerian approach was employed for multiphase simulation. Only the drag force was considered as the interactive force between the gas and liquid phases. In order to reflect the drag change due to the non-spherical shape of bubbles, the Ishii-Zuber model was employed. In addition, to reflect the dense dispersion of bubbles and the bubble cavity formed after the impeller, Behzadi correlation-based drag coefficient modification was included. The final drag model is given by

$$C_{D,modified}=\begin{cases} C_{D,IZ}\left(e^{3.64\alpha_g}+\alpha_g^{0.864}\right) & 0<\alpha_g\leq 0.6 \\ C_{D,IZ}(-62.5815\alpha_g+51.0651) & 0.6<\alpha_g<0.8 \\ C_{D,IZ} & 0.8\leq\alpha_g<1 \end{cases} \tag{1}$$

$$C_{D,IZ}=\max\left\{\min\left(\frac{24}{Re}(1+0.15Re^{0.687}),\frac{72}{Re}\right),\frac{8}{3}\frac{Eo}{Eo+4}\right\} \tag{2}$$

, where α_g is the gas volume fraction, and E is the Eotvos number. The SST $k-\omega$ model, one of the Reynolds averaging Navier-Stokes approaches, was employed for the turbulence model.

Population balance equation (PBE) was used to simulate the bubble coalescence and breakage. In this study, a quadratic method of moments (QMOM) based approximation was employed to solve the equation. The number of ODEs in the PBM was determined by the number of tracked moments and the number of bins. For the gas-liquid systems, such as bubble columns or stirred tanks, three nodes and 6-moments equations are widely selected as an option to guarantee both accuracy and stability. The equations of PBE and QMOM are

$$\frac{\partial n(L;\mathbf{x},t)}{\partial t}+\nabla\cdot[\vec{u}n(L;\mathbf{x},t)]= B_{ag}(L;\mathbf{x},t)-D_{ag}(L;\mathbf{x},t)+B_{br}(L;\mathbf{x},t)-D_{br}(L;\mathbf{x},t) \tag{3}$$

$$\frac{\partial m_k}{\partial t}+\nabla\cdot[\vec{u}m_k]=\overline{S_{m_k}} \tag{4}$$

$$m_k=\int_0^\infty n(L)L^k dL\approx\sum_{i=1}^{N=3}w_i L_i^k \tag{5}$$

$$\overline{S_{m_k}}=\frac{1}{2}\sum_{i=1}^{N=3}w_i\sum_{j=1}^{N=3}w_j h_{ij}\left[\left(L_i^3+L_j^3\right)^{\frac{k}{3}}-L_i^k-L_j^k\right]+\sum_{i=1}^{N=3}w_i g_i\left(\overline{b_i^k}-L_i^k\right) \tag{6}$$

$n(L;\mathbf{x},t)$ is a function of property and m represents the moments of property function. g_{ij} and h_{ij} are called kernels and have different forms depending on physical phenomena. The above equation assumes QMOM with three nodes.

Since QMOM employs approximation using moments and nodes, it is necessary to transform the source term, which reflects the physical phenomenon, into the moment space. Because the problem must be closed in the moment space, the source term has a rather complicated form. In this study, the kernels for bubble coalescence and breakage phenomena are given by

$$\mathrm{h}(L_1, L_2) = C_1(L_1+L_2)^2\epsilon^{\frac{1}{3}}\sqrt{L_1^{2/3}+L_2^{2/3}}\exp\left(-C_2\frac{\mu_l\rho_l\epsilon}{\sigma^2}\left(\frac{L_1L_2}{L_1+L_2}\right)^4\right) \quad (7)$$

$$g(L) = C_3\epsilon^{\frac{1}{3}}\mathrm{erfc}\left(\sqrt{C_4\frac{\sigma}{\rho_L\epsilon^{\frac{2}{3}}L^{\frac{5}{3}}}+C_5\frac{\mu_l}{\sqrt{\rho_L\rho_G}\epsilon^{\frac{1}{3}}L^{\frac{4}{3}}}}\right) \quad (8)$$

$$\beta(\mathrm{L},\lambda) = 180\left(\frac{L^2}{\lambda^3}\right)\left(\frac{L^3}{\lambda^3}\right)^2\left(1-\frac{L^3}{\lambda^3}\right)^2 \quad (9)$$

, where h is the coalescence kernel, g is the breakage kernel, and β is the daughter bubble size distribution. The kernel of Prince and Blanch with the collision efficiency of Coulalglou and Tavlarides was employed for the coalescence kernel, and the kernel of Laakkonen was employed for breakage and daughter bubble size distribution. The built-in QMOM framework in Fluent® was used, and other optional kernels and drag models were implemented via the user-defined macros in Fluent®.

The volumetric mass transfer rate was calculated as the product of the mass transfer coefficient K_L and the interfacial area *a*. In this study, the interfacial area was directly obtained from the CFD simulation results, and the mass transfer coefficient was estimated by the slip velocity model:

$$a = a_{Sauter} * R \quad (10)$$

$$\mathrm{a_{Sauter}} = \sum_{\mathrm{i}}\pi d_{b,i}^2 n_i \approx \frac{6a_g}{d_{32}} \quad (11)$$

$$\mathrm{R} = 1 + 0.163\mathrm{E_o}^{0.757} \quad (12)$$

$$\mathrm{k_L} = \frac{2}{\sqrt{\pi}}\sqrt{\frac{D_L v_b}{d_b}} \quad (13)$$

where a_{Sauter} is the interfacial area calculated from the Sauter mean diameter, and R is the aspect ratio of bubbles with ellipsoidal shape. D_L is the diffusivity of oxygen in water, and v_b is the slip velocity.

3. Numerical strategy

A 6-blade Rushton turbine tank with a ring sparger was selected as the target system. The geometry of Laakkonen was referenced, and the thickness of blades and baffles was set to 4mm, which is similar to that of the actual Rushton tank. The multiple reference frame (MRF) method was employed to model the impeller movement. A Green-Gauss node-based formulation was used for calculating gradients, QUICK scheme for the momentum equations, PRESTO! scheme for the pressure balance equation, and the 1st order upwind scheme was used for all other equations. The pressure outlet was used for the outlet boundary condition. A grid sensitivity study was conducted using 4 fully structured hexagonal meshes of 419k, 666k, 729k and 930k elements, and the 666k elements mesh was finally selected. The shape of grid is shown in Figure 2.

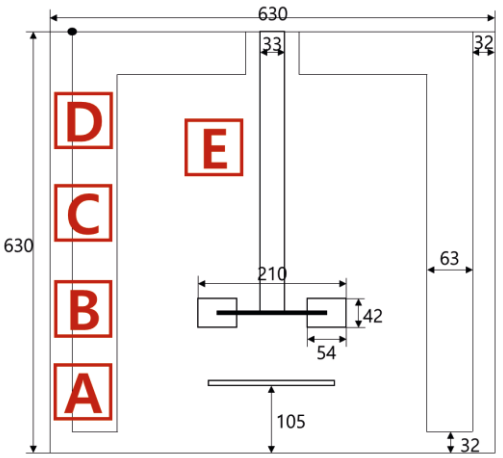


Figure 1 Schematic figure of a 6-blade Rushton tank with a ring sparger. The alphabets indicate where Laakkonen measured the diameter of bubbles. Model validation was performed through comparison with the measurement results.

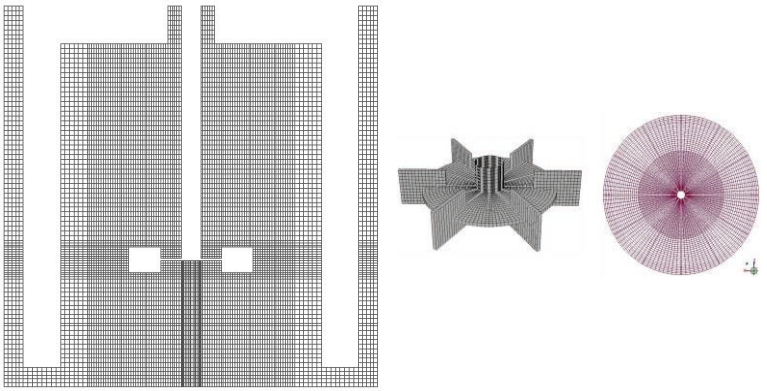


Figure 2 Structure of grid with 666k hexagonal elements. From left, front view, rotor view, and top view

4. Results and discussion

4.1. Model validation

The overall gas hold-up was 4.35%, which is in fairly good agreement with the measured value of 6%. The power number was 2.2, estimated from the turbulent dissipation rate. Although the typical power number of a Rushton tank is about 5, it is widely known that the estimation of power number from the turbulent dissipation rate leads to underestimated values. Moreover, previous studies also obtained power number ranging 1 to 2. Since the power number is used as an indicator for judging whether an appropriate level of turbulence was simulated in this work, the above power number is acceptable for this study. Figure 3 shows the contours of bubble sauter mean within the reactor.

The reactor-volume-average $K_L a$ was 0.0465, which showed good agreement with the measurement. The local bubble size results achieved an error of less than 15% with experimental measurements at all the measured points, which is an improvement over previous studies. Figure 4 shows the results of a local bubble size comparison.

4.2. Effect of liquid viscosity on oxygen mass transfer

From nominal water-viscosity, the change in oxygen mass transfer was analyzed by increasing the viscosity of liquid up to ten times. At 10 cp, it was estimated that only 28% of the oxygen transfer in nominal water was delivered. The rate of oxygen mass transfer in the reactor was found to be inversely proportional to the viscosity power of 0.653, and the R square value was 0.933. These results are consistent with the empirical correlations widely used. Figure 5 below shows the change of oxygen mass transfer with increasing viscosity.

5. Conclusions

This work showed that the CFD-PBM model can reflect the heterogeneous hydrodynamics of a 200L bioreactor including baffles, rotors, and sparger. The results of local bubble size distribution and overall gas volume fraction are consistent with the measurement results. In addition, the volume-averaged oxygen mass transfer rate was predicted to decrease by 28%, and a relationship between viscosity and oxygen mass transfer was estimated to be inversely proportional to the 0.653 power.

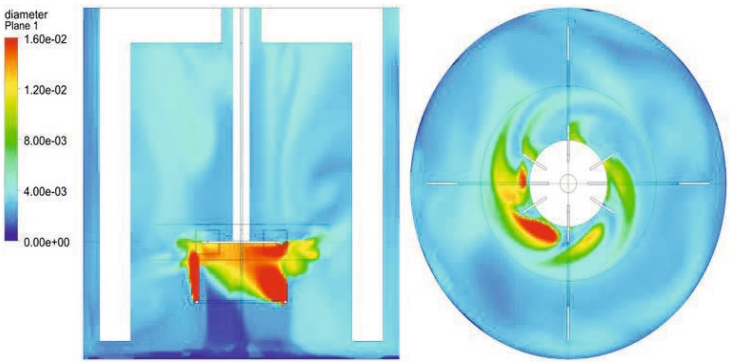


Figure 3 Contour of bubble diameter (1cp Case)

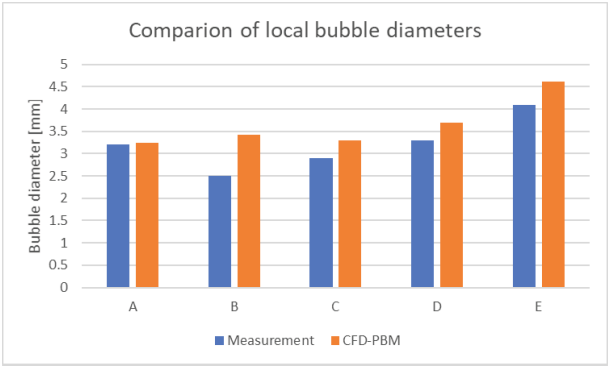


Figure 4 Comparison of local bubble diameters of 5 points

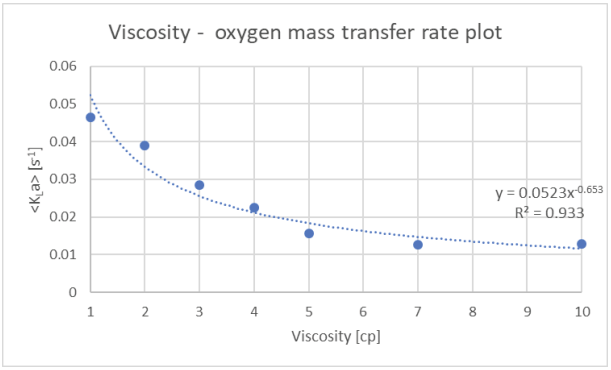


Figure 5 Viscosity and oxygen mass transfer rate plot

Acknowledgements

This paper was supported by Korea Institute for Advancement of Technology(KIAT) grant funded by the Korea Government(MOTIE) (P0008475, The Competency Development Program for Industry Specialist)

References

Y.S. Bak, and J. M. Lee, “Integration of Population Balance Model with Computational Fluid Dynamics for Estimation of Oxygen Mass Transfer Rate in Bioreactor”, Mater’s Thesis, https://www.dcollection.net/handler/snu/000000167087

A. Behzadi, R. Issa, and H. Rusche, "Modelling of dispersed bubble and droplet flow at high phase fractions," *Chemical Engineering Science,* vol. 59, no. 4, pp. 759-770, 2004.

J. Gimbun, C. D. Rielly, and Z. K. Nagy, "Modelling of mass transfer in gas–liquid stirred tanks agitated by Rushton turbine and CD-6 impeller: A scale-up study," *Chemical Engineering Research and Design,* vol. 87, no. 4, pp. 437-451, 2009.

M. Laakkonen, P. Moilanen, V. Alopaeus, and J. Aittamaa, "Modelling local bubble size distributions in agitated vessels," *Chemical Engineering Science,* vol. 62, no. 3, pp. 721-740, 2007.

M. J. Prince and H. W. Blanch, "Bubble coalescence and break-up in air-sparged bubble columns," *AIChE Journal,* vol. 36, no. 10, pp. 1485-1499, 1990.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50153-6

Knowledge-matching based computational framework for genome-scale metabolic model refinement

Kiumars Badr, Q. Peter He and Jin Wang*

Department of Chemical Engineering, Auburn University, Auburn, AL, 36849, USA
**Corresponding Author's E-mail: wang@auburn.edu*

Abstract

Genome-scale metabolic models (GEMs) are mathematically structured knowledge base reconstructed from annotated genome of different organisms. With the advancement of next-generation sequencing technology, many organisms have had their genomes sequenced. However, obtaining a high-quality GEM is highly time-consuming, even with the introduction of several genome-scale reconstruction tools that offer automated draft network generation and gap filling. It has been recognized that the iterative process of manual curation and refinement is the limiting step of GEM development, and how to expedite the GEM refinement is still an open question. As cellular metabolism is a complex system with very high degree of freedom and redundancy, the principles and techniques developed in process systems engineering can be adapted to expedite GEM refinement. In this work we present a knowledge-matching based computation framework for GEM refinement, and demonstrate the effectiveness of the proposed solution using the refinement of a GEM for *Clostridium tyrobutyricum* .

Keywords: Genome-scale metabolic model; manual curation; system identification; multivariate analysis; visualization.

1. Introduction

A genome-scale metabolic model (GEM) is a mathematically structured knowledge base that is reconstructed from annotated genome of an organism (King et al., 2016)). A GEM contains a list of biochemical reactions, metabolites and (annotated) genes involved in the cellular metabolism for a specific organism, as well as a set of biophysical constraints (e.g., nutrient uptake and substrate availabilities, etc.). A high-quality GEM can be used to conduct simulations to answer various questions about the capabilities of the organism, serve as a framework to integrate and interpret omics data collected through experiments, and guide the design of mutant for metabolic engineering, etc.

The GEM reconstruction process generally consists of the following steps (Thiele and Palsson, 2010): (1) a draft network is reconstructed based on the annotation of a genome and the prediction of candidate metabolic functions; (2) the draft reconstruction is refined or curated by the user in an iterative manner through an exhaustive review of each reaction, metabolite and gene in the network; (3) the reconstruction is transformed into a mathematical structure, with an objective function and a set of constraints to account for different culture conditions. With the mathematical representation, the resulting GEM can be evaluated by reproducing the experimental data. If the GEM predictions do not match the experimental data, the manual refinement process will be repeated till the quality of the GEM is satisfactory.

It has been recognized that the iterative process of manual curation and refinement is the limiting step of GEM development (Mendoza et al., 2019)). To expedite the development process, several genome-scale reconstruction tools that offer automated draft network generation and gap filling have been reported. However, a systematic assessment of seven GEM reconstruction tools (Mendoza et al., 2019)) concluded that none of the tools performed well in all of the evaluated categories, and there was a relatively large discrepancy between the automatic reconstruction and the high-quality manual curation. Therefore, it is no surprise that the currently validated high-quality GEMs collected in BiGG Models are all manually curated (King et al., 2016)).

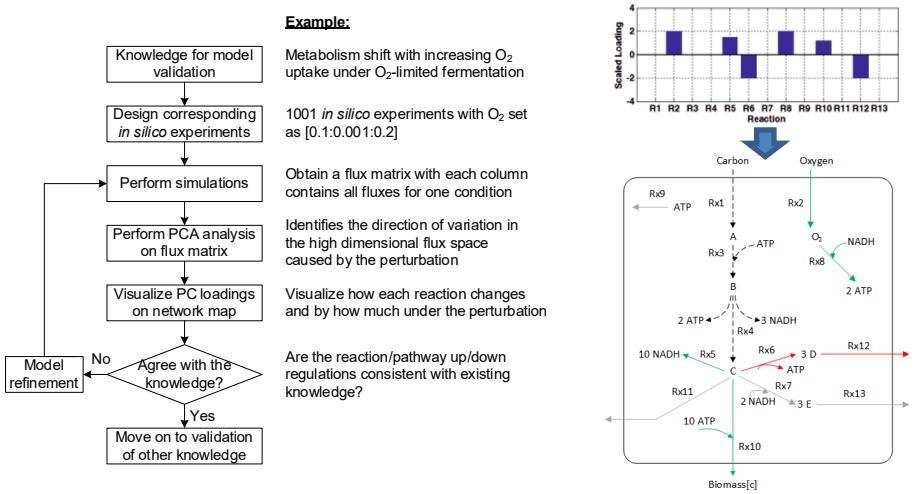
For GEM refinement, one big challenge is to quickly identify the root cause of an erroneous model prediction. Because of the complex interconnectivity in the GEM, many times seemingly unrelated reactions located far away from the "problematic" reactions (i.e., reactions that are not carried out in the expected way) play a key role in correcting the model behavior. Directly comparing the model prediction with experimental data usually yields limited information on the "hidden" relations between the erroneous model prediction and its corresponding root cause reactions. Currently, GEM refinement relies heavily on the modeler's knowledge and capability to sort out clues from various simulation results, and is labor intensive and time consuming.

As cellular metabolism is a complex system with very high degree of freedom and redundancy, we hypothesize that the principles and techniques developed in process systems engineering can be adapted to address some of the challenges associated with GEM refinement. In this work, we first review the foundation of the proposed solution, i.e., a knowledge-matching based computational framework for GEM analysis, then we present the developed GEM refinement approach, and demonstrate the effectiveness of the proposed solution using the development of a GEM for *C. tyrobutyricum*.

2. Knowledge-matching based framework for GEM analysis

In essence, a high-quality GEM is a comprehensive knowledge base of the organism's cellular metabolism. If the key qualitative knowledge captured by a GEM could be extracted and visualized, then GEM refinement could be expedited through knowledge-matching, i.e., by comparing the extracted knowledge with the available ones. The knowledge captured by a GEM is usually embedded in numerical model predictions under various environmental perturbations (e.g., culture media and conditions) and/or genetic perturbations (e.g., mutant), which are difficult to extract and visualize.

To address this challenge, we have developed a system identification (SID) based computational framework for knowledge-matching based GEM analysis (Damiani et al., 2015). In the SID framework, three main steps are involved for GEM analysis, as shown in Figure 1a. First, a set of *in silico* experiments are designed to cover the whole transition path between two metabolic states, where each simulation represents an incremental change along the path. The experiments represent a one-dimensional perturbation to the cellular metabolism and result in a flux matrix containing a series of cellular metabolic states along the transition path; next, system identification tools (such as principal component analysis or PCA) are applied to extract the knowledge contained in the flux matrix (e.g., how the cells respond to the perturbation); finally, the extracted knowledge is visualized against the metabolic network map and compared with the existing knowledge for GEM analysis. Here we use an illustrate example (Figure 1b) to demonstrate how the SID-based GEM analysis works.



(a) (b)

Fig. 1 (a) Illustrated flow chart of the SID-based GEM analysis; (b) an example showing the loadings (top) and their visualization on a simulated network (bottom).

As shown in the bottom part of the Figure 1b, the illustrative example is a toy metabolic network with 13 reactions, which consumes both carbon and oxygen to produce biomass and three potential by-products (C, D and E). To determine if the model correctly captures the metabolic capabilities of the network, we examine if the knowledge captured by the model on how the network respond to an increased O_2 supply under oxygen-limited condition is correct. This response is well understood and well conserved among different microbes, and we expect to see increased flux through electron transport chain due to the availability of additional O_2 (electron acceptor) and increased biomass growth. To do so, we conducted a set of *in silico* experiments, where the carbon uptake was fixed, while the upper limit of O_2 uptake was gradually increased from 0.1 to 0.2 mmol/gDCW/min. These simulations resulted in a 13×1001 flux matrix where each column contains all fluxes within the network under a given O_2 uptake flux. Next, PCA, which we have applied in a novel closed-loop subspace identification algorithm (Wang and Qin, 2006), is applied to extract the knowledge contained in the flux matrix. Due to the linear network structure and one-dimensional perturbation, one principal component (PC) is expected to capture 100% of variation in the flux matrix provided that the whole transition path is located within the same phenotype. Indeed, our result showed that one PC captured 100% of variation. The PC loading captures the knowledge of how the perturbation propagates through the network, i.e., how each reaction flux in the network is affected by the perturbation. The scaled loading is plotted in Figure 1b (top part) and visualized on the network structure in Figure 1b (bottom part). Figure 1b shows that with increasing oxygen supply, the carbon flux in the network shifts from production of by-product D toward biomass production, together with upregulated electron transport chain which converts reducing power (NADH) to produce ATP. This model response agrees with existing understanding as described before, and confirms the model quality. This example further illustrates that although rooted in numerical simulations, the knowledge-matching offered by the SID-based framework is qualitative in nature, which offers enhanced robustness against systematic error among experimental results reported in different literature.

3. SID-based GEM refinement

To identify the root cause of an erroneous model prediction, we adapt the SID-based framework to identify a few candidates which drastically narrows the search space for the potential root causes. The SID-guided GEM refinement process consists of three steps: (1) conducting a set of *in silico* experiments with forced correct model behaviors by applying additional hard constraints; (2) applying PCA to determine how each reaction is affected by the forced correct behavior; (3) identify the candidate root cause reactions – we hypothesize that the reactions affected the most by the forced model behavior are the potential root cause reactions that contribute to the erroneous model behavior. Once the candidate reactions are identified, they will be examined against existing knowledge to see if any of them is the actual root cause. Below we use the development of a GEM for *C. tyrobutyricum* as an example to illustrate how the SID-guided GEM refinement works.

C. tyrobutyricum is a novel and promising industrial chassis strain that plays an important role in carboxylic acids production from lignocellulosic substrates. Since a GEM for *C. tyrobutyricum* is currently not available, we first developed a draft GEM based on a published GEM for *C. beijerinckii*, iCM925 (Milne et al., 2011). Genome analysis revealed that *C. tyrobutyricum* does not use phosphotransbutyrylase nor butyrate kinase for butyric acid production; instead, it uses a CoA transferase to mediate the butyric acid production from butyryl-CoA by reassimilation of acetic acid (Lee et al., 2016). Therefore, the corresponding reaction pathways were deleted from or added to iCM925 to obtain the draft GEM for *C. tyrobutyricum* (iKB917).

Experimental results showed that when cultivated on glucose, *C. tyrobutyricum* excretes butyrate, acetate, hydrogen and carbon dioxide, in addition to cell growth (Lee et al., 2016). However, when the draft GEM was tested by using the experimentally reported glucose uptake rate as the only constraint to predict cell growth and product excretion, the model does not produce butyrate at all, but over-produces acetate, hydrogen and biomass, while under-produces CO_2 (Table 1). Since the draft GEM contains 939 reactions, it is not clear which reaction pathway(s) could be the causes for this erroneous modeling behavior and should be modified. To improve the GEM, we applied the SID-based GEM refinement, as shown in Figure 2.

Table 1. Comparison of measured data with predicted data by the GEM

No.	Condition	Glucose	Acetate	Butyrate	H_2	CO_2	Biomass
1	Experimental measurement	-2.594	0.717	1.895	4.75	4.34	0.052
2	Draft GEM prediction	-2.594	4.223	0	7.96	4.17	0.074
3	Refined GEM prediction	-2.594	0.823	1.834	4.81	4.45	0.053

(1) To force the correct model behavior, i.e., butyrate production, we added a hard constraint to force butyrate production flux gradually increased from 0 to 1.0 mmol/gDCW/min, to produce a flux matrix of 939×1001.

(2) PCA was applied to analyze the flux matrix, and the PC loadings of selected reactions are plotted in Fig. 2 (bottom plot). Clearly, the affected reactions were widespread across multiple pathways. However, the reactions that were affected the most are concentrated in two reaction pathways, i.e., H_2 and H_2O syntheses.

(3) Additional review of literature suggests that due to physiological constraints, the amount of H_2 and CO_2 produced by the cells are proportional to each other (Jo and

Kim, 2016). Since the GEM does not contain any regulatory mechanism, it over produces H_2. To correct this erroneous behavior, we implemented a soft constraint on the ratio of H_2/CO_2 production. Such a soft constraint could provide similar effects as regulatory mechanisms without adding hard constraints, therefore provides more flexibility for the GEM to simulate different phenotypes under different conditions.

(4) The refined GEM (with the added soft constraint) was tested again, which showed significantly improved prediction accuracy (low row of Table 1). The refined GEM not only predicts the excretion of butyrate, but also predicts the excretion of all the byproducts whose fluxes showed excellent agreement with experimental results.

This example highlights the challenge of GEM refinement and effectiveness of the SID-guided refinement. Without the guidance of the SID framework, it can take much longer to figure out what would be the root cause of the erroneous model behavior.

4. Conclusions

GEM has been recognized as a highly effective tool to elucidate the complex cellular metabolisms. It offers a foundation to integrate various omics data, and helps reveal genotype-phenotype relationships, which is fundamental to biology. Despite the recent advancement in automated GEM draft reconstruction and gap filling, manual curation and refinement of GEM remains the limiting step in GEM development. To help address this challenge, we developed a knowledge-matching based computational framework to expedite the GEM refinement. By adapting the principles and techniques in process systems engineering, the proposed SID-guided GEM refinement can quickly identify the "hidden" root cause for the erroneous modeling behavior, therefore significantly expedite the GEM refinement process. The effectiveness of the SID-guided GEM refinement framework is demonstrated through improving a draft GEM of *C. tyrobutyricum*.

References

Damiani, A.L., He, Q.P., Jeffries, T.W., Wang, J., 2015. Comprehensive evaluation of two genome-scale metabolic network models for Scheffersomyces stipitis. Biotechnology and Bioengineering 112, 1250–1262.

Jo, J.H., Kim, W., 2016. Carbon material distribution and flux analysis under varying glucose concentrations in hydrogen-producing Clostridium tyrobutyricum JM1. Journal of Biotechnology 228, 103–111.

King, Z.A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J.A., Ebrahim, A., Palsson, B.O., Lewis, N.E., 2016. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Research 44, D515.

Lee, J., Jang, Y.-S., Han, M.-J., Kim, J.Y., Lee, S.Y., 2016. Deciphering Clostridium tyrobutyricum Metabolism Based on the Whole-Genome Sequence and Proteome Analyses. mBio 7, e00743--16.

Mendoza, S.N., Olivier, B.G., Molenaar, D., Teusink, B., 2019. A systematic assessment of current genome-scale metabolic reconstruction tools. Genome Biology 2019 20:1 20, 1–20.

Milne, C.B., Eddy, J.A., Raju, R., Ardekani, S., Kim, P.-J., Senger, R.S., Jin, Y.-S., Blaschek, H.P., Price, N.D., 2011. Metabolic network reconstruction and genome-scale model of butanol-producing strain Clostridium beijerinckii NCIMB 8052. BMC Syst Biol 5, 130.

Thiele, I., Palsson, B.Ø., 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nature Protocols 2010 5:1 5, 93–121.

Wang, J., Qin, S.J., 2006. Closed-loop subspace identification using the parity space. Automatica 42, 315–320.

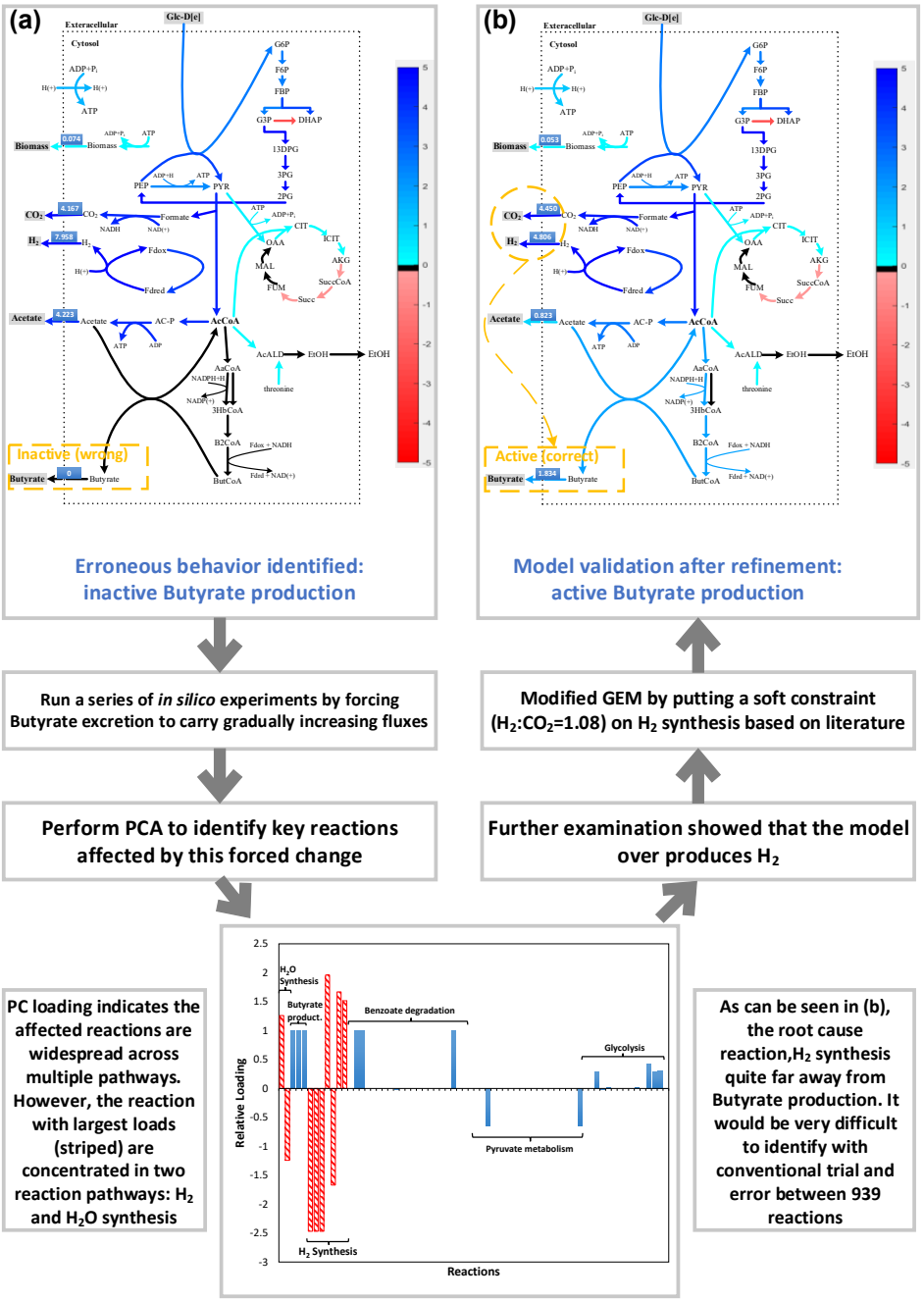


Figure 2. Demonstration of the SID based framework for GEM refinement

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50154-8

Multi-Regional Surrogate Model Selection (MRSMS) approach for the analysis and optimal fitting of univariate responses

Srikar V Srinivas[a], and I.A.Karimi[a*]

[a]*Department of Chemical & Biomolecular Engineering, National University of Singapore, 4 Engineering Drive, 117585 Singapore*
cheiak@nus.edu.sg

Abstract

This study focuses on the development of a Multi-Regional Surrogate Model Selection (MRSMS) approach for the optimal fitting and analysis of univariate responses. Using a library of simple curve fitting and regression models, the proposed approach can be used to provide surrogate model recommendations at each section of the response based on the user specified selection of the residual error metric and its corresponding data fit range. The efficacy of the proposed approach is validated using the Henry Hub Natural Gas price dataset and its fitting performance is compared with Piecewise Linear (PL), Neural Network (NN) and Support Vector Regression (SVR) models. It was found that proposed MRSMS approach outperformed the fitting performance of the considered models.

Keywords: Surrogate Model; Optimal Data Fitting; Univariate Responses, Piecewise Linear Model

1. INTRODUCTION

The idea of employing simple and less complex surrogates (or) surrogate models in estimating a functional relationship of a complex chemical phenomenon has gained traction in the research community over the last few decades. Surrogate models, also known as meta-models (or) response surfaces, are used to accurately mimic the relationship between the inputs and outputs of a system (McBride and Sundmacher (2019)).

The selection of a surrogate model for a particular system (or) phenomenon is an extremely challenging task. Although there is no consensus or clear-cut mantra to select a particular model, the existing popular custom in surrogate-based modelling of complex phenomena is to fit one (or) more surrogate models and select the best fitting model (based on an error metric) over an entire data response.

Several studies have aimed at finding the best surrogate model for a particular dataset. While some research works were heavily reliant on the trial-and-error approach of trying one model after another on a dataset and then comparing its performance with its fellow member models (Williams and Cremaschi (2020)); other research works have focused on the development of ensemble model approaches, where the individual member models are combined to form a weighted ensemble through some optimization framework (Goel et.al (2007)). In recent times, the concept of meta-learning has gained traction, where the best surrogate model can be selected through an automated platform utilizing an

exhaustive library of models, error metrics, rules, and datasets. Cozad et al. (2014) developed the ALAMO model approach to determine the optimal mix of nonlinear basis functions by iteratively solving a MILP based optimization problem. Cui et al. (2016) developed a framework for surrogate selection using a meta-learning approach. Their approach involved the characterization of the datasets into several quantitative features, which were then used as a selection guide in identifying the best modeling technique. S.Garud et al. (2018) developed the LEAPS2 model for finding the best surrogate model based on the given input-output data relationship and drew comparisons with Cui's meta-learning model. M.Ahmad and Karimi (2021) further revised the existing LEAPS2 framework through the inclusion of more surrogate models, attribute sets and weight metrics in their analysis database and drew positive comparisons of their revised framework's accuracy against the earlier LEAPS2 framework.

Based on the above literature, we have found that most of the developed approaches only provide surrogate recommendations valid over the entire sample size of a data response. The number of research works providing surrogate model recommendations at different sections of a data response have been very limited and this study focuses on the development of a multi-regional surrogate model selection (MRSMS) approach for the optimal fitting and analysis of univariate responses. The proposed approach has been applied on the Henry Hub Natural Gas price dataset and its performance has been compared with the widely used Piecewise Linear (PL), Neural Network (NN) and Support Vector Regression (SVR) models.

2. METHODOLOGY

The proposed approach has been implemented with MATLAB v2020b used as the programming language and steps of the proposed MRSMS approach are as follows:

I. Step 1: On acquiring the required data and user-specific inputs, the first segment of acquired data (W_{size}) is fitted using all the member models of the library and their in-sample fitting performances are recorded in a matrix. The list of member models and user-specific inputs can be found in Tables 1 and 2 respectively.

II. Step 2: Subsequent data points and their responses are iteratively added to the existing segment based on the step size (S_{size}) value and all the models respective in-sample performances are recorded and appended to the existing performance matrix. The number of models having a fitting performance within the range of Fit_{min} & Fit_{max} are noted at each iteration. Overfitting models (or) models having a fit greater than Fit_{max} are assigned a 'NaN' or null value at every iteration.

III. Step 3: The performances of all models are recorded at every iteration and the termination of a data segment (or) region of data is executed when none of the models are within the given user-specified range of fit for a given number of user-specified consecutive validation steps (V_{steps}).

IV. Step 4: Once the termination of a data segment is executed, the best characteristic model for this segment is determined to be the model with the best average performance (within the user-specified levels of fit) over this region. In the hypothetical case that if two models were found to have the same fitting performance over a data segment, then the best model for the segment is determined to the model with the lowest SQS score developed by M. Ahmad

and Karimi (2021). Unlike the commonly used Akaike Information Criteria (AIC), which assumes that the model parameters are determined using the 'maximum likelihood estimate' theory and can sometimes favour its goodness of term (gof) over the second penalty term, the SQS metric developed to addresses these limitations, provides a proportional weightage between a model's fitting accuracy and its complexity. The formula of the SQS score is displayed in Eq.(1).

$$SQS_{model} = (1 - R^2_{model}) \cdot \left(\frac{\ln(1 + DOF_{model})}{\ln(1 + K_{segment})}\right)^{0.5} \quad (1)$$

Where, R^2_{model} represents the fit of the model, DOF_{model} represents the number of independent model parameters and $K_{segment}$ represents the sample size of the current data segment under analysis.

Step 5: The last point (or) the point where all the model performances started falling outside the regions of user-specified fit with the current segment of data indicates the termination of that segment and this final point is recorded as a breakpoint. The process is repeated from Step1 from the latest breakpoint until all the points in the response are fitted optimally.

All the member models in this approach have been implemented using the Curve Fitting and Mathematics Toolbox of MATLAB. The model parameters of these models are estimated using the least squares approximation. We would like to point out that the results of fit, segment sizes and the location of these breakpoints obtained from this approach are subject to change depending on the characteristic nature of the dataset, the type of models present in the library and the user specified inputs.

Table 1: MRSMS Member Models

Model	Model Form
2nd Degree Polynomial model	$y = ax^2 + bx + c$
Exponential model	$y = a \cdot exp(bx)$
Gaussian Model	$y = a \cdot exp(-(x - b) \cdot c^{-1})^2)$
Power Law Model	$y = ax^b$
Sinusoidal Function	$y = a \cdot \sin(bx + c)$
Hyperbolic Tangent Function	$y = a \cdot \tanh(bx)$
Root Function	$y = \sqrt[a]{x}$
Rational Function	$y = (ax + b) \cdot (cx + d)^{-1}$
Sigmoid Function	$y = \left(1 + \exp\left(-c1 \cdot (x - c2)\right)\right)^{-1}$
Logarithmic Function	$y = a \cdot \log(bx + c)$

Table 2: MRSMS User-Specific Inputs

Parameter	Meaning	Default Value
Fit_{min}	Minimum threshold of desired fit	-
$Fit_{\max}$	Maximum threshold of desired fit	-
S_{size}	Step size	1
W_{size}	Minimum size of a segment	2·(DOF of largest model in the library)
V_{steps}	Validation steps (or) number of continuous iterations to confirm change of model	6

3. CASE STUDY

The Henry Hub Natural Gas Price dataset containing the monthly prices of natural gas (in Dollars/Millions BTU) from January 1997 onwards, is obtained from the U.S Energy Information Administration (EIA) website (EIA (2021)). The default settings of $S_{size} = 1$, $W_{size} = 8$, & V_{steps}=6 is used here. The aim is to have an R^2 fit between 0.70 and 0.98 for the first 200 points of this dataset.

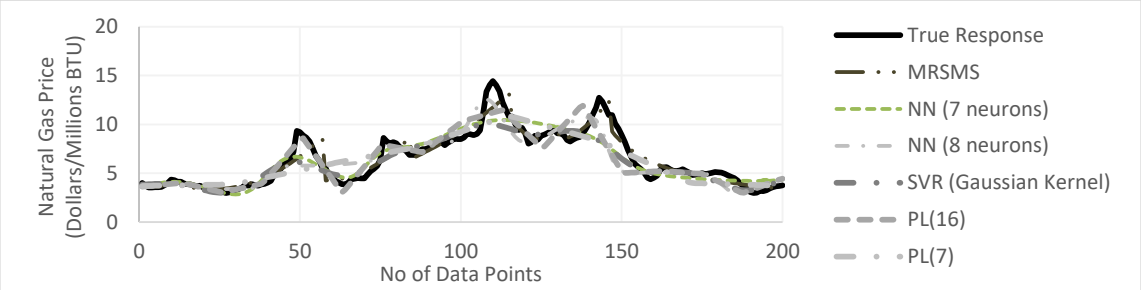


Figure 1: Comparison of the MRSMS fit with various models.

Table 3: MRSMS summary of fitting performance on Henry Hub data

Segment No	Starting Point	Ending Point	Model Form	R^2_{fit}
1	1	11	Polynomial	0.905
2	12	55	Polynomial	0.899
3	56	81	Polynomial	0.875
4	82	111	Gaussian	0.932
5	112	125	Polynomial	0.840
6	126	140	Polynomial	0.892
7	141	200	Rational	0.816

Based on Table 3, it was found that there are 7 optimal segments (or 7 optimal models) required to fit the dataset within the given user-specifications. Out of the 7 models required for fitting the curve, 2nd degree polynomial models (with different coefficients) were found to be the best models in most of the segments except for segments 4 and 7 where the Gaussian and Rational functions were found to the best models respectively.

Due to the concerns about the paper length, the results of fit using the R^2 metric are only presented. However, the MRSMS fitting performance on this dataset has been verified using other commonly used error metrics such as the Root Mean-Squared Error (RMSE), Mean-Squared Error (MSE), and Mean Absolute Error (Metric) as well.

Table 4 provides a comparison of the MRSMS fitting performance against the considered regression models. To establish a valid basis for comparison, the number of hidden neurons for NN models and the no of segments for PL models (D'Errico, J (2009)) were initially set to be equal to the number of segments obtained from the MRSMS approach. These settings were then modified to identify the respective model structures required to outperform the MRSMS fitting performance.

Table 4: Comparison of MRSMS fitting performance with other models

Approach	No of Segments	$Average\ R^2_{fit}$	Parameters	No of Parameters	RMSE
MRSMS	7	0.879	Model-Coeffs	22	0.855
PL	7	0.649	Model-Coeffs	14	1.383
PL	16	0.916	Model-Coeffs	32	0.753
NN (7)	1	0.834	Weights & biases	22	1.105
NN (8)	1	0.897	Weights & biases	25	0.869
SVR (Gaussian)	1	0.842	Support Vectors	106	1.08

From Table 4, it can be inferred that for almost the same number of parameters, the developed MRSMS approach provides a better fitting performance than the PL models (7 segments), NN model (7 neurons) and even the SVR model (with Gaussian Kernel & 106 support vectors). Although the NN (8 neurons) and PL (16 linear segments) models were found to out-perform the fitting ability of the MRSMS approach, the number of parameters involved in their data fitting are much higher than that of the MRSMS approach. Modelling a dataset with a lesser number of model parameters that achieves a similar fitting performance obtained from more complex models demonstrates a significant advantage of using the developed approach. We believe that the proposed MRSMS approach can guide researchers in understanding the different types of model behaviour present at different sections of their univariate response.

5. Conclusion

This study focuses on the development of the MRSMS approach its application for optimally fitting of the Henry Hub Natural Gas price dataset. Relying on a library of models and user-specified inputs, the developed MRSMS approach is compared with several linear and nonlinear models in terms of fitting performance, and we found that the MRSMS approach outperformed these models in optimally fitting the given univariate

response for the given user specified inputs. Our future works will focus on conducting sensitivity analysis of the user-specified inputs on the overall fitting performance, validation on other univariate chemical engineering responses (generated with the help of the Aspen Hysys Software), comparison with other models such as regression trees and finally, the application of the methodology towards the accurate modelling and analysis of Multi-Input Single Output (MISO) responses.

6. Acknowledgements

This research work is supported by the NUS Research Scholarship (co-funded by SDPPVO) at the National University of Singapore (NUS). We would like to acknowledge MathWorks and AspenTech Inc. for providing the academic licenses (for software usage) to NUS. We also would like to express our acknowledgment and gratitude to all the reviewers for their detailed comments and reviews of this manuscript.

7. References

Ahmad, M., & Karimi, I. A. (2021). Revised learning based evolutionary assistive paradigm for surrogate selection (LEAPS2v2). *Computers & Chemical Engineering*, *152*, 107385. https://doi.org/10.1016/j.compchemeng.2021.107385

Cozad, A., Sahinidis, N. V., & Miller, D. C. (2014). Learning surrogate models for simulation-based optimization. *AIChE Journal*, *60*(6), 2211–2227. https://doi.org/10.1002/aic.14418

Cui, C., Hu, M., Weir, J. D., & Wu, T. (2016). A recommendation system for meta-modeling: A meta-learning based approach. *Expert Systems with Applications*, *46*, 33–44. https://doi.org/10.1016/j.eswa.2015.10.021

D'Errico, J. (2009). SLM-shape language modeling. *SLM-Shape Language Modeling.. http://www. mathworks. com/matlabcentral/fileexchange/24443-slm-shape-language-modeling: Mathworks*.

Garud, S. S., Karimi, I. A., & Kraft, M. (2018). LEAPS2: Learning based Evolutionary Assistive Paradigm for Surrogate Selection. *Computers & Chemical Engineering*, *119*, 352–370. https://doi.org/10.1016/j.compchemeng.2018.09.008

Goel, T., Haftka, R. T., Shyy, W., & Queipo, N. V. (2007). Ensemble of surrogates. *Structural and Multidisciplinary Optimization*, *33*(3), 199–216. https://doi.org/10.1007/s00158-006-0051-9

McBride, K., & Sundmacher, K. (2019). Overview of Surrogate Modeling in Chemical Process Engineering. *Chemie Ingenieur Technik*, *91*(3), 228–239. https://doi.org/10.1002/cite.201800091

US Energy Information Administration. (2021). Short-Term Energy Outlook, July 2021.

Williams, B., & Cremaschi, S. (2021). Selection of surrogate modeling techniques for surface approximation and surrogate-based optimization. *Chemical Engineering Research and Design*, *170*, 76–89. https://doi.org/10.1016/j.cherd.2021.03.028

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50155-X

A Digital Reality Pilot Plant for Research and Learning

Deborah E CARBERRY[a*], Mark N JONES[b], Khosrow BAGHERPOUR[a], Christian BEENFELDT[c], Martin P ANDERSSON[a], Seyed Soheil MANSOURI[a*],

[a] Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søltofts Plads, Building 228A, 2800 Kgs. Lyngby, DENMARK
[b]Molecular Quantum Solutions ApS, Maskinvej 5, 2860 Søborg, DENMARK
[c]Knowledge Hub Zealand, Holbækvej 141B 4400 Kalundborg, DENMARK
debca@kt.dtu.dk, seso@kt.dtu.dk

Abstract

Technological advancements increase the demand for more diverse skill sets across industry. To meet this growing demand, digitally-based education and training interventions are a common solution. For practical skills development, digital tools have the potential to increase the capacity of provisioned educational experiences by allowing more people to be trained on more skills using less physical infrastructure. Having analyzed the learning objectives for several PILOT PLANT courses delivered at the Chemical and Biochemical Engineering Department at the Technical University of Denmark (DTU-KT), eight broad competencies have been identified across undergraduate and graduate programs. These competencies, for which the learning objectives are intended, is for the effective operation of a number of unit operations. One way to map these competencies to the respective unit operations and their corresponding theoretical frameworks is via two cognitive interfaces. The first of these interfaces is concerned with cognition and learning whilst the second is concerned with User and spatial Interaction. An educational virtual reality (VR) platform is currently being developed and implemented at DTU-KT. It is designed to introduce learners to the physical structures and components that make up a unit operation, whilst demonstrating the functionality and processes for which the unit operation is intended. This paper will start out by presenting the cognitive framework that was designed and utilized for the design of digital reality learning tools at the PILOT PLANT at DTU-KT. Thereafter, it will present a design study for the recently developed 360° VR learning tool that was developed for a Citric Acid Crystallizer.

Keywords: Education; Pedagogy, Unit Operations; Experiments, Virtual Reality.

1. Introduction

In December 2020, not least because of the limitations imposed on situated learning as a result of COVID 19, a product idea was born at the Department of Chemical and Biochemical Engineering at the Technical University of Denmark (DTU-KT). This idea was to leverage eXtended Reality (XR) technologies to build learning experiences for PILOT PLANT courses. It would take another three months for the concept to take shape, by the end of which time, it was understood that the Team would endeavor to deliver walk-throughs for Unit Operation experiments. The Citric Acid Crystallizer was selected for a project to build a prototype and the Instruction Manual for the experiment was to serve as a foundation in forming a product that could be described as a deconstructed linear narrative (Note: A linear narrative is a narrative (story) with a fixed sequence of steps (events). In other words, there is only one sensible path to navigate the content. Whilst a deconstructed linear narrative breaks the narrative down into smaller steps, facilitating flexible entry and exit points.). At the same time, a technology partner was on boarded, and in April 2021, the development of a prototype commenced.

This paper will set out some of the preceding and initial design activities that informed the prototype project. It will commence by introducing the learning objectives and desired outcomes for existing course offerings at DTU.KT that served as a means to understand the purpose of the respective courses. It will then move on to demonstrate a model that was used to understand the connection between the product features and cognition (or learning), with a view to bridging the product to the learning objectives.

Finally, this paper will present a design study to explore how specific design features were formed in response to anticipated cognitive conditions, and subject to the constraints of the product architecture.

2. Existing Course Offering

At DTU.KT, several courses are on offer for operator training at both undergraduate and graduate level. The design team commenced by reviewing the learning objectives for each course and grouped them according to eight core competencies. This categorization strategy was leveraged to make it easier to think about how to design the product to serve various knowledge and skills requirements for students studying at different levels. The categories, education level, and their corresponding definitions are set out in Table 1 and Figure 1 below.

For the prototype our target user was undergraduate students. Of the four intended competencies for undergraduate students, two were deemed to be appropriate for a mixed reality tool, namely 'Familiarity' and 'Basic Operations'. Hence, the design team would seek to develop a product that would respond to the learning objectives that sought to facilitate students in demonstrating familiarity with the various pieces of equipment and instrumentation, and, to demonstrate a capacity to operate same.

Intuitively, it was felt that spatial reasoning was one of a number of theories that could assist the design team in selecting a number of teaching theories that could further guide the work. Spatial reasoning is a diffused area of research (Davis & Francis 2020). It involves, but is not limited to, a range of abstract reasoning skills that can identify, translate, transform and synthesize information about the geometry, scale, direction and motion of physical objects. More fundamentally, it exploits the physical context of our own body in relation to our environment. In the section that follows, a model will be employed to articulate how relevant learning and teaching theories can be understood in the context of mixed reality education products.

Table 1: Definitions for desired competency categories

Learning Objectives	**Translating theoretical understanding to a practical situation**
Safety	Knowledge of safe and responsible work practices
Planning	Independently able to plan experiments
Familiarity	Familiar with the equipment and instrumentation
Operations	Capable of operating equipment and instrumentation
Calculations	Able to carry out calculations on given or collected data
Interpreting	Able to make sense of the raw data and calculated data
Contrasting	Can anticipate the correct values for collected and calculated data. Where discrepancies occur, can identify the source(s) of error
Communication	Can communicate goals, processes and findings

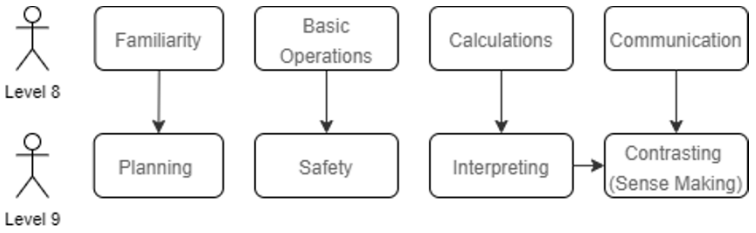


Figure 1: Summary of learning outcomes, based on level of Education

3. A Cognitive Framework for the design of digital reality learning tools

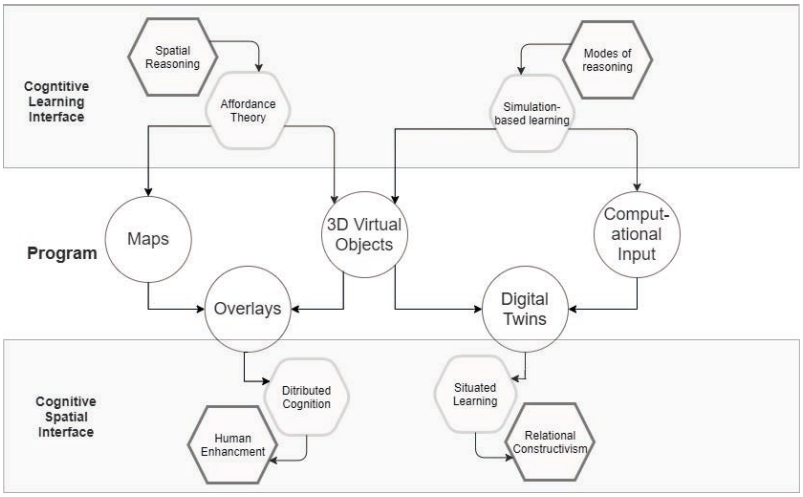


Figure 2: A Cognitive-Program Model

Note: Teaching theories can be identified as hexagons with lighter borders and softer corners than their learning theory counterparts. Affordance Theory (Gibson, 1979) Simulation-based learning (Ton De Jong & Sarti, 1994) (Nazir et al, 2012) Distributed Cognition (Hutchins, 1991) Situated Learning (Lave & Wenger, 1991)

Educational theories can be broken down into theories of learning and theories of teaching (Davis & Francis 2020), where learning theories seek to understand how learning happens and teaching theories are methods to facilitate learning. As was the case for competencies, the design team sought to develop a simple model for rapid conceptualization of pedagogical design and cognitive processes in light of their existing and proposed digital systems. Bearing in mind that this is merely a model, the purpose of which is to facilitate understanding, and, keeping in mind that it is by no means exhaustive, a Cognitive-Program Model was devised for the project and is represented in Figure 2 above.

The model consists of three layers; a cognitive mental interface layer, a program layer (or product layer), and, a cognitive spatial interface layer. The cognitive mental interface is devoted to human learning whilst the cognitive spatial interface is devoted to the relationship one has with their environment. Sitting between the two cognitive interfaces is the digital program, mediating the relationship between how one thinks and how one interacts with their surroundings.

In our Cognitive-Program Model, you can draw a vertical line through the center of the diagram to produce two discrete product groups. To the left, is a system for a mixed reality solution that leverages affordances to create a distributed cognitive system. This mixed reality system exploits human perception, with a view to reducing cognitive load, for the purpose of enabling a user to focus their attention on the task at hand. Distributed cognition, as it relates to cognitive load, can be understood by two actions. The first is the potential to transfer content from our working memories and locate it elsewhere in our environment, for example, as a list on a piece of paper. The second action is the potential to access information that is annotated in our environment such as is the case with signs and symbols that populate motorways. Combining these two actions is a powerful methodology to reduce cognitive overload. Moreover, a learning environment with these capabilities is ideal for learners who are largely unfamiliar or inexperienced with a subject or skill. On the right side of the diagram, is a system for a digital twin which leverages simulation in a situated context. In contrast, the digital twin is suitable for those with a good fundamental understanding of their subject and whom are capable of engaging in a variety of reason-based cognitive activities. This paper will not concern itself with the design of digital twins.

In the section that follows, a design study will be presented that demonstrates how some of the learning and teaching theories from our Cognitive-Program Model were applied to the design of a VR prototype at DTU in 2021.

4. Design Study: The PID

An affordance can be described as 'the quality or property of an object that defines its possible uses or makes clear how it can or should be used' (Merriam-Webster.com). A Piping and Instrumentation Diagram (PID) is a language of related symbols. The symbols in a PID do not necessarily have strong affordances. On the other hand, the components and instrumentation for a unit operation are designed to tell a user how they can be used. In nature it is more common to be exposed to a thing itself before one encounters or creates a symbol for the same thing. However, often in formal education, as is the case for learning PILOT PLANT operations at DTU, it happens in reverse.

At the 1st project meeting, the value of the PID was emphasized. As a language in Chemical Engineering, and one that most students are fluent in, it was a feature that offered significant pedagogical value in the overall design. Hence, it was decided to exploit it as a device that learners could use to comprehend the virtual renditions. An initial design was proposed, however, the choice of hardware and software for the prototype did not facilitate the design of bespoke features. On the contrary, the technical architecture had been selected for the speed of implementation, and not for flexibility. As such the design team needed to be creative.

In this design study, both the proposed and the implemented design for integrating the PID into the walkthrough for a pilot scaled experiment will be outlined. In both cases, the solution seeks to achieve the same goal, which is to use the PID as a way to navigate and make sense of pilot scaled experiments.

4.1 Original Design

Excerpts from the original design specification:

'The PID will serve as a navigation device in the experiment. For each step in the experimental process, the PID will change dynamically from step to step, illustrating which components are involved in each corresponding step.'

'A user can locate individual components of the unit operation by selecting the corresponding unit in the PID, and vice versa'.

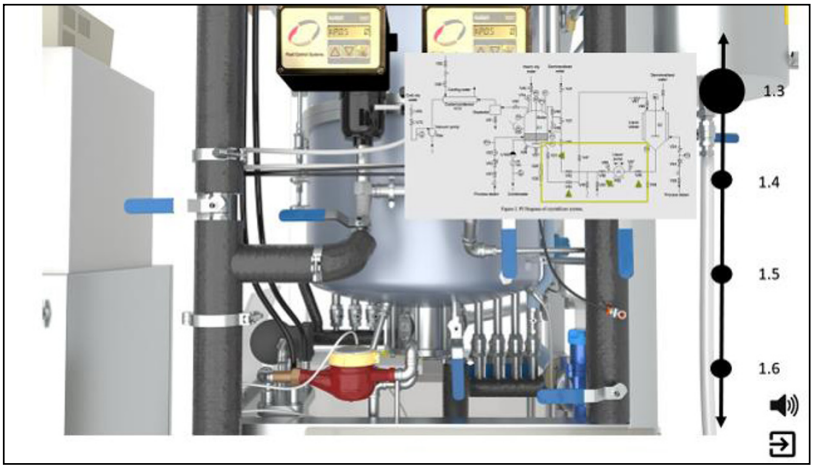


Figure 3: Mockup of how the PID could change dynamically from step to step

Note: The vertical axis located on the right side of the screen acts as a navigation system for tracking the user's current location in the experiment. In Figure 3, steps 1.3 – 1.6 are in view. Arrows on both ends of the axis allow the user to move back and forward through the experiment/narrative. The PID can be visible or invisible (see Figure 4). When the PID is visible, only the objects in the PID that are relevant to that step will be visible (not shown here).

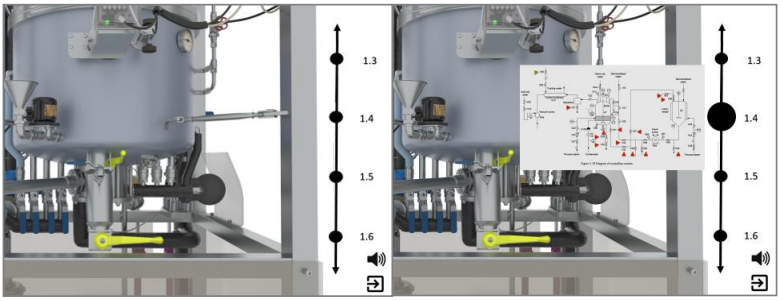


Figure 4: Mock up for how to activate the PID as a navigation tool
Note: A user can activate the PID overlay by performing an action on the corresponding step on the navigation axis. Otherwise, the PID will be invisible.

The first of the two use cases from the requirement specification is illustrated as mockups in Figure 3 and Figure 4 above. In the mockups, the PID has been scaled up to facilitate understanding, however, in a VR environment, the relative size of the PID to the unit operation would be significantly smaller.
This proposed solution responded to the requirements for the product by developing an interactive play between the PID and virtual unit operation. The feature permitted the user to project knowledge about something they already understood, i.e. the PID, onto a new knowledge domain, i.e. physical unit operations. This process is referred to as scaffolding, a theory that states that learners build new knowledge by using their existing knowledge (Wood, Ross, Bruner, 1976). And whilst attractive as a concept, time constraints dictated that an alternative solution was required, at least initially. As such, the proposed features are currently part of the product backlog for future development.

4.2 Implemented Design

The solution that was implemented for the prototype project took advantage of a 360 VR development environment called present4D. The main staple for the content delivery portion of the product was the development of a series of animated videos for each step in the experiment. It was decided that in order to capture the value of the PID for each step in the experiment, cut scenes between a PID (highlighting relevant components for that step) and animated segments (that demonstrated the required action and location of the corresponding components) could serve to bridge the two. This is illustrated in Figure 5 below.
The navigation system for the prototype was designed using an identical architecture to the proposed design, however the UX in the prototype is somewhat different. Rather than using a linear axis for example, the user selects from a menu of modules, and then, from a menu of steps.

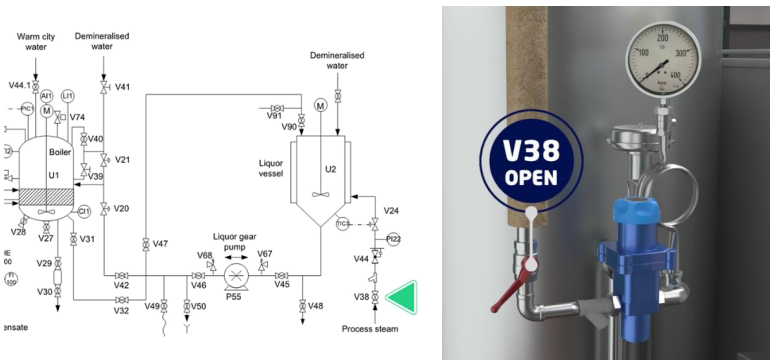


Figure 5: Left: PID with arrow pointing to V38 (valve 38). Right: Screen shot from animation with tag attached to V38 (valve 38).

5. Conclusion

VR is a visual medium and ideal for addressing pedagogical challenges that involve visual learning such as object recognition, process demonstrations and simulations of body active experiences. There are several potential value propositions for developing digital reality tools to teach unit operations as part of chemical and biochemical engineering education, not least for remote learning. Based on our experience, it is advantageous to consider the goals and the assumptions of the same courses offered in their current format, to guarantee that the competencies sought (or part thereof) are consistent.

Technical architectures will shape the possibilities for product design, however, it is often possible to achieve the same requirements using different technologies, albeit via different design features. This does not necessarily mean that the designs for each medium will be of equal value therefore the choice of technical architecture should be informed by their suitability to address the product goals. In the case of the prototype for DTU.KT, a full scale user testing program is yet to be implemented, however, the current implemented solution has recently been made available to both Teachers and Students at the department as a working prototype. Early feedback suggests that the product is a useful tool for experiment preparation for both sets of users.

Acknowledgements

The authors would like to acknowledge the Department of Chemical and Biochemical Engineering at the Technical University of Denmark, Erasmus+: Higher Education – International Capacity Building project TESS (Application number 609925-EPP-1-2019-1-NO-EPPKA2-CBHE-JP), and Novo Nordisk Foundation through Fermentation Based Manufacturing Initiative (Grant number NNF17SA0031362), for providing financial support.

References

Davis, B., & Francis, K., 2020, “Spatial Reasoning” in Discourses on Learning in Education. https://learningdiscourses.com.

Kraus, B., 2017, Plädoyer für den Relationalen Konstruktivismus und eine Relationale Soziale Arbeit - In: Forum sozial, 1, S. 29-35 - URN: urn:nbn:de:0111-pedocs-153817 - DOI: 10.25656/01:15381

Gibson, J.J., 1979,. The Ecological Approach to Visual Perception

Hutchins, E., 1991, The social organization of distributed cognition

Ton De Jong, Sarti, L., 1994, Design and Production of Multimedia and Simulation-Based Learning Material. Dordrecht, Netherlands ; Boston, Kluwer.

Nazir, S., Totaro, R., Brambilla, S., Colombo, S., Manca, D., 2012, Virtual Reality and Augmented-Virtual Reality as Tools to Train Industrial Operators, Editor(s): Lockhart Bogle I.D, Fairweather M., Computer Aided Chemical Engineering, Elsevier, Volume 30, Pages 1397-1401, https://doi.org/10.1016/B978-0-444-59520-1.50138-X.

Lave, J., & Wenger, E. (1991). Situated learning: Legitimate peripheral participation. Cambridge University Press. https://doi.org/10.1017/CBO9780511815355

“Affordance.” Merriam-Webster.com Dictionary, Merriam-Webster, https://www.merriam-webster.com/dictionary/affordance. Accessed 29 Oct. 2021.

Wood, D., Bruner, J.S. and Ross, G., 1976, THE ROLE OF TUTORING IN PROBLEM SOLVING. Journal of Child Psychology and Psychiatry, 17: 89-100 https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50156-1

Soft sensors development for industrial reactive distillation processes under small training datasets

Andrei Torgashov[a], Svetlana Samotylova [a], Fan Yang[b*]

[a]*Process Control Laboratory, Institute of Automation and Control Process FEB RAS, 5 Radio Str., Vladivostok, Russia*
[b]*Beijing National Research Center for Information Science and Technology and Department of Automation, Tsinghua University, Beijing 100084, China*
yangfan@mail.tsinghua.edu.cn

Abstract

Under real industrial conditions, there are usually missing values in the data. This is due to measurement errors, sensor failures, missing values in real-time databases, irregular measurement intervals, and data that are not covering the total operation range of the plant. Due to the lack of data, the soft sensor (SS) model is of poor quality. The functioning of this model is also unsatisfactory in the new operating points of the plant in the case of a small training sample. We propose the use of a calibrated rigorous (first-principles) process model with acceptable limits of parametric uncertainty to extend the training dataset, which allows us to take into account the physicochemical characteristics of the process. It is shown that the extension of the training sample based on the rigorous model makes it possible to obtain a nonlinear SS of higher accuracy.

Keywords: soft sensor; small training dataset, rigorous model; sample extension; reactive distillation.

1. Introduction

Small samples in the development of SSs lead to overfitting and poor interpretability of the models of an industrial plant when using existing approaches of multiple regression and deep learning (Zhu et al., 2018). To improve the accuracy of SSs under conditions of small samples, we propose approaches based on the generation of virtual data. The approach based on the generation of virtual samples (Virtual Sample Generation - VSG) is used to obtain a large number of new virtual data based on information obtained from small data sets in order to increase the accuracy of SSs. To reduce gaps in the training data set, an information-expanded function based on triangular membership has been developed (Chen et al., 2017). For iterative evaluation of the most appropriate training samples, an approach based on particle swarm optimization with VSG (PSOVSG) was proposed. Zhu et al. (2020) propose a novel locally linear embedding based on a virtual sample generation (LLEVSG) approach. In their proposed LLEVSG method, locally linear embedding is first used to extract features from the original data space. Next, back-propagation neural network (BPNN) and a method of random interpolation are utilized to generate effective virtual samples in the sparse region of the original data.

Within the framework of Industry 4.0, one of the important aspects should be noted, which is the use of a physically based first principles model, intended not only to model an industrial plant, but also to validate collected data for building a reliable model by monitoring the process in various operating points (Vaccari et al., 2021). Thus, the use

of a physically based model for virtual sampling is an advanced approach, as it allows the taking into account of the physicochemical features of the chemical process, and it becomes possible to take into account various modes of operation of the industrial unit (Ko et al., 2021).

Hsiao et al., 2021 proposed a methodology for developing SSs that combines a calibrated, physically based model to extend the training dataset and transfer learning to develop them. In this current work, we propose the use of a calibrated rigorous process model for sample extension in case of parametric uncertainty in the phase equilibrium model, which is often seen in practice. The superiority of the proposed approach based on the sample extension compared to the traditional PSO and PSO with bootstrapping is demonstrated based on the industrial reactive distillation of methyl tert-butyl ether (MTBE) production.

2. Statement of problem and its analysis

We have considered The problem of developing an SS for estimating the output product of a reactive distillation column produced by a high-octane additive of gasoline as MTBE in a small training sample (Fig. 1). The mass fraction of MTBE (grade A) should exceed 98% in the product.

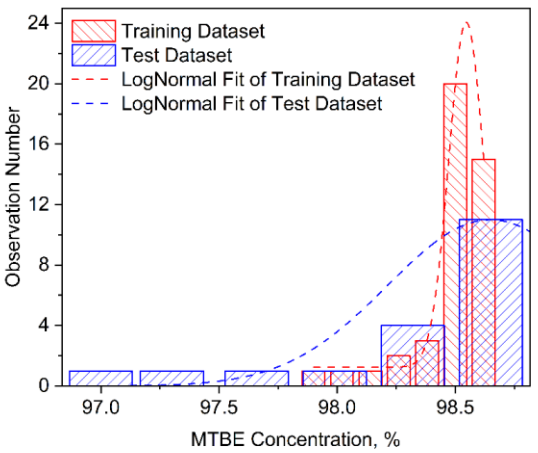


Figure 1: Histograms of the output variable distribution

To build an SS, a training sample with input variables (Table 1) was used $U=[u_1,\ldots,u_n]\in R^{42\times6}$ and output variable observations $Y=[y_1,\ldots,y_n]\in R^{42\times1}$. To test the developed models, a test sample was used with $U=[u_1,\ldots,u_n]\in R^{18\times6}$ and $Y=[y_1,\ldots,y_n]\in R^{18\times1}$.

Basically, the "black box" model is used when building SS. The main drawback of this approach is the lack of use of available a priori knowledge about the process. To solve the problem of a small training sample and improve the accuracy of the SS, we propose the use of a rigorous analytical model for a reactive distillation (RD) process.

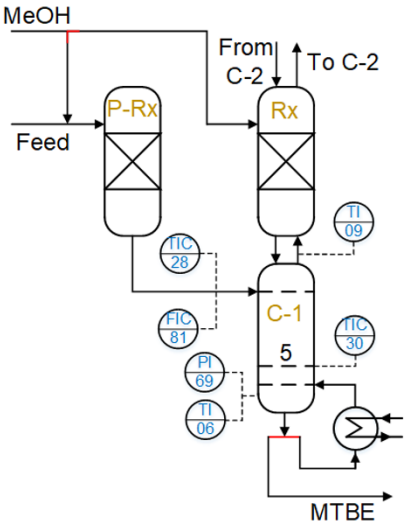


Figure 2. The MTBE synthesis flowsheet diagram

Table 1. Process variables

No	Process variable, *u*	Tag name	Units of meas.
1	Feed of column C-1	FIC81	m^3/hr
2	Bottom pressure of column C-1	PI69	MPa
3	Bottom temperature of column C-1	TI06	°C
4	Top temperature of column C-1	TI09	°C
5	Feed temperature of column C-1	TIC28	°C
6	5th tray temperature of column C-1	TIC30	°C

However, it is also not possible to directly use the rigorous model of RD to build an SS, e.g. using the concept of a "gray box" (Ahmad et al., 2020) here, due to the high dimension of the model. A set of highly nonlinear equations can be obtained by a material balance equation (1), enthalpy balance equation (2), Murphree mass transfer efficiency equation (3), phase equilibrium equation (4), molar fraction summation equation (5) and chemical reaction rate equation (6) (Wang et al., 2020, Mendoza et al., 2013):

$$F\mathbf{x}_{\mathrm{F}i} - V\mathbf{y}_i - L\mathbf{x}_i + \sum_r \nu_{ir} R_r = 0 \qquad (i = 1,\ldots,C) \qquad (1)$$

$$F\mathbf{h}_{\mathrm{F}} - V\mathbf{h}^V - L\mathbf{h}^L = 0 \qquad (2)$$

$$E^{\mathrm{MV}}\left(\mathbf{y}_i^* - \mathbf{y}_i\right) - \left(\mathbf{y}_i - \mathbf{y}_i\right) = 0 \qquad (i = 1,\ldots,C-1) \qquad (3)$$

$$\mathbf{K}_i\mathbf{x}_i - \mathbf{y}_i^* = 0 \qquad (4)$$

$$1 - \sum_{i=1}^{C}\mathbf{y}_i^* = 0, \quad 1 - \sum_{i=1}^{C}\mathbf{y}_i = 0, \quad 1 - \sum_{i=1}^{C}\mathbf{x}_i = 0 \qquad (5)$$

$$r_r = f\left(T, P, \mathbf{x}_i, \mathbf{y}_i\right) \qquad (6)$$

where F, V, and L denote the mass flow rates of the feed; "V" and "L" represent the internal liquid and vapor streams; $\mathbf{x}_{\mathrm{F}i}$ is the mass fraction of the component i of the feed; $\mathbf{x}_i$ and $\mathbf{y}_i$ are the mass fractions of the component i of the liquid and vapor phases; ν_{ir} is

stoichiometric coefficient component *i* of the reaction *r*; R is the extent of the reaction, mol·s^{-1}; $\mathbf{h}_F$, $\mathbf{h}^V$ and $\mathbf{h}^L$ are the feed, vapor and liquid mass enthalpies; $\mathbf{y}_i^*$ is the equilibrium mass fractions of the component *i* of the vapor phase; E^{MV} is Murphree mass transfer efficiency; r_r is the reaction rate of the reaction *r*; T is the temperature, K; *P* is the pressure, MPa.

3. Soft sensor evaluation based on the sample extension

Therefore, we propose the use of a rigorous model of the RD plant to simulate the steady-state operating points under various conditions in order to extend the training sample. The proposed algorithm for developing the SS for estimating the quality of MTBE production in case of a small training dataset is shown in figure 2.

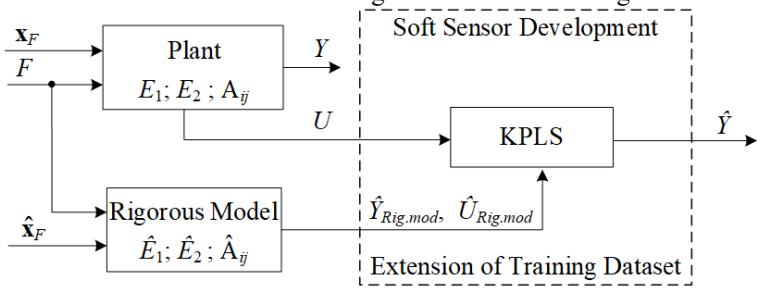


Figure 3. The scheme of algorithm for developing the SS in case of a small sample of industrial data.

Kernel partial least squares (KPLS) (Jin et al., 2014, Liu et al., 2010) are used for developing the SS to estimate the quality of MTBE. Giving the input matrix $U=[u_1,\ldots,u_n]\in R^{n\times m}$ and output matrix $Y=[y_1,\ldots,y_n]\in R^{n\times p}$ used, where *n* is the number of observations, and *m* and *p* represent the number of process variables and quality variables, respectively.

The rigorous model of reactive distillation was calibrated on industrial data and the following Murphree efficiencies were estimated: $\hat{E}_1=0.8$ (trays 1-5); $\hat{E}_2=0.6$ (trays 6-28). It should be noted that the calibration of the rigorous model is also influenced by the parameters of binary interaction. For the MTBE production process, the parameters of the binary interaction of isobutylene (DIB) dimers (2,4,4-trimethyl-1-pentene, and 2,4,4-trimethyl-2-pentene) are practically absent in the literature. However, DIB is often detected in the output product as an impurity component. Binary parameters for MTBE and DIB are $A_{MTBE,DIB}=-17.13$ K, $A_{DIB,MTBE}=45.75$ K. (Sundmacher et al., 1999). For a rigorous model, the activity coefficients of the MTBE-system are calculated from the UNIQUAC-model. In this regard, the deviations of the calibrated rigorous model parameters $\Delta\hat{A}_{ij}$ of the process under study were introduced as $\Delta\hat{A}_{ij}=\left|A_{ij}-\hat{A}_{ij}\right|$. To determine the threshold values of deviations $\Delta_{\hat{A}_{ij}}^{opt.th}$, the % of reduction of R^2 (δR^2) was estimated for the range of parametric uncertainty $\Delta\hat{A}_{ij}\in[0.1,0.4]$ relative to the base case when the model exactly corresponds to the plant, i.e. relatively to $\Delta\hat{A}_{ij}=0$. If δR^2 exceeded 5%, then a threshold deviation value was assigned for the corresponding

binary parameters under which the extension of the training dataset leading to improvement of soft sensor model accuracy (Table 2).

Table 2. R^2 and *MAE* obtained on the test sample before and after the extension of the training sample

Criterion	Extension of the training dataset (number of added observations)			
	0	10	20	30
Particle swarm optimization				
MAE	0,1626	0,1595	0,1548	0,1503
R^2	0,5514	0,6060	0,6897	0,7025
Particle swarm optimization and bootstrap				
MAE	0,1626	0,1579	0,1495	0,1489
R^2	0,5514	0,6243	0,7077	0,7139
Rigorous model				
MAE	0,1626	0,1501	0,1483	0,1408
R^2	0,5514	0,7053	0,7245	0,7316

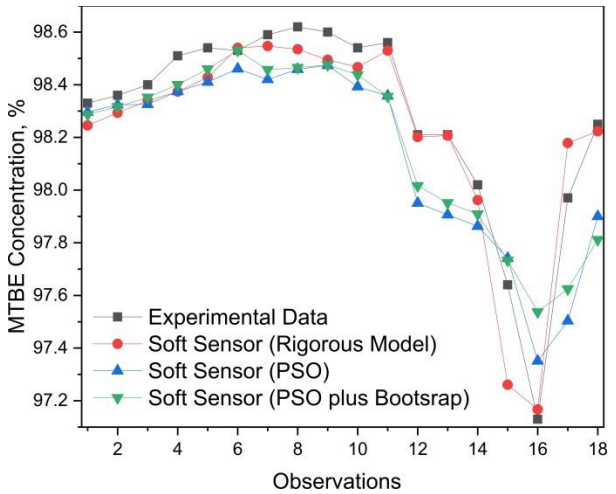


Figure 4: Comparative study of several SSs

The proposed algorithm based on the rigorous model for extending the training sample is compared with well-known approaches for building models to evaluate the quality of end products in conditions of small training samples (Fig.4)—for example, expanding the training sample using particle swarm optimization (Chen et al., 2017), as well as the joint application of particle swarm optimization with bootstrap (Fortuna et al., 2009, Zhang et al., 2021). The value of *MAE* for an SS to estimate the concentration of MTBE using the proposed approach is reduced by $((0.1626\text{-}0.1408)/0.1626)\times 100 \approx 13.41\%$.

4. Conclusions

The use of a calibrated, physically based model of the process with acceptable limits of parametric uncertainty of binary parameters to extend the training sample allows us to take into account the physico-chemical characteristics of the process. The use of nonlinear methods for building soft sensors on an extended training sample allows to increase the accuracy of the developed SS.

5. Acknowledgements

The reported study was partially funded by RFBR and NSFC (project numbers 21-57-53005 and 62111530057).

References

Ahmad, I., Ayub, A., Kano, M., and Cheema, I.I., 2020, Gray-Box Soft Sensors in Process Industry: Current Practice, and Future Prospects in Era of Big Data, Processes, 8, 243

Chen, Z.S., Zhu, B., He, Y.-L., and Yu, L.A., 2017, A PSO Based Virtual Sample Generation Method for Small Sample Sets: Applications to Regression Datasets, Engineering Applications of Artificial Intelligence, 59, 236–243

Fortuna, L., Graziani, S., and Xibilia, M. G., 2009, Comparison of Soft-Sensor Design Methods for Industrial Plants Using Small Data Sets, IEEE Transactions on Instrumentation and Measurement, 58, 2444–2451

Hsiao, Y.D., Kang, J.L., and Wong, D.S.H., 2021, Development of Robust and Physically Interpretable Soft Sensor for Industrial Distillation Column Using Transfer Learning with Small Datasets, Processes, 9(4), 667

Jin H., Chen X., Yang J., Wu L., 2014, Adaptive Soft Sensor Modeling Framework Based on Just-In-Time Learning and Kernel Partial Least Squares Regression for Nonlinear Multiphase Batch Processes, Computers & Chemical Engineering, 71, 77-93

Ko, C., Lee, H., Lim, Y., and Lee, W.B., 2021, Development of Augmented Virtual Reality-Based Operator Training System for Accident Prevention in a Refinery, Korean Journal of Chemical Engineering, 38(8), 1566-1577

Liu, J., Chen, D.S., and Shen, J.F., 2010, Development of Self-Validating Soft Sensors Using Fast Moving Window Partial Least Squares, Industrial & Engineering Chemistry Research, 49(22), 11530-11546

Mendoza, D. F., Palacio, L. M., Graciano, J. E., Riascos, C. A., Vianna Jr, A. S., Le Roux, G. C., 2013, Real-Time Optimization of an Industrial-Scale Vapor Recompression Distillation Process. Model Validation and Analysis, Industrial & Engineering Chemistry Research, 52(16), 5735-5746

Sundmacher, K., Uhde, G., Hoffmann, U., 1999, Multiple Reactions in Catalytic Distillation Processes for the Production of the Fuel Oxygenates MTBE and TAME: Analysis by Rigorous Model and Experimental Validation, Chemical Engineering Science, 54, 2839-2847

Vaccari, M., Bacci di Capaci, R., Brunazzi, E., Tognotti, L., Pierno, P., Vagheggi, R., Pannocchia, G., 2021, Optimally managing chemical plant operations: An example oriented by Industry 4.0 paradigms, Industrial & Engineering Chemistry Research, 60, 7853−7867

Wang, L., Sun, X., Xia, L., Wang, J., Xiang, S., 2020, Inside–Out Method for Simulating a Reactive Distillation Process, Processes, 8(5), 604

Zhang, X.-H., Xu, Y., He, Y.-L., Zhu, Q.-X., 2021, Novel Manifold Learning Based Virtual Sample Generation for Optimizing Soft Sensor with Small Data, ISA Transactions, 109, 229-241

Zhu, J., Ge, Z., Song, Z., and Gao, F., 2018, Review and Big Data Perspectives on Robust Data Mining Approaches for Industrial Process Modeling with Outliers and Missing Data, Annual Reviews in Control, 46, 107-133

Zhu, Q.X., Zhang, X.-H., He, Y.L., 2020, Novel Virtual Sample Generation Based on Locally Linear Embedding for Optimizing the Small Sample Problem: Case of Soft Sensor Applications, Industrial & Engineering Chemistry Research, 59(40), 17977-17986

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50157-3

Comprehensive Quantification of Model Prediction Uncertainty for Simulated Moving Bed Chromatography

Kensuke Suzuki, Tomoyuki Yajima, Yoshiaki Kawajiri*

Department of Materials Process Engineering, Nagoya University, Furo-cho 1, Chikusa, Aichi 464-8603, Japan
kawajiri@nagoya-u.jp

Abstract

To realize robust process design and operation of simulated moving bed (SMB) chromatography, prediction uncertainty in a model must be quantified. In this study, we quantify the prediction uncertainty as predictive distributions stemming from a model parameter uncertainty estimated from experimental data. The resulting predictive distributions provide insights into selecting experimental methods to assure reliability of the model predictions and parameters.

Keywords: Simulated Moving Bed; Uncertainty Quantification; Bayesian Inference; Predictive Distribution; Sequential Monte Carlo.

1. Introduction

Simulated moving bed (SMB) chromatography is a continuous separation process widely used for the separation of petrochemicals, enantiomers, and chiral compounds (Schmidt-Traub et al., 2012). The SMB process consists of multiple chromatographic columns connected in a semi-closed-loop structure, and a mixture is continuously separated by cyclic operation. Between each column, there are inlet valves for supplying feed and desorbent streams and outlet valves for withdrawing extract and raffinate streams. The positions of these four streams are simultaneously switched along the flow direction at a regular time interval, called step time.

A lot of research has been carried out on modeling of SMB, process design, and optimization using mathematical models. In most of these studies, the SMB model consists of a system of partial differential algebraic equations (PDAEs). In this model, there are unknown parameters that are dependent on adsorbents and components to be separated. To estimate parameters, many estimation methods have been studied; in general, these unknown model parameters are estimated from single-column batch experiments and applied to the SMB model. Research to optimize operating conditions—the four flow rates and step time—by model simulation using estimated parameters has also been extensively conducted.

To assess the SMB model in terms of reliability and robustness, uncertainty in the model prediction—product concentration, purity, recovery, and internal concentration profile—should be quantified, but there has been little research on the quantification of uncertainty of the SMB model prediction so far. Uncertainty in model prediction stems from uncertainties in model parameters, implementation of design and operation, and measurement. This prediction uncertainty can be quantified by estimating the predictive distribution. By obtaining the predictive distribution, the reliability of the model prediction can be assessed under various uncertainties, which is crucial for robust

process design and optimization; however, the predictive distributions of purity and recovery, which are essential for evaluating the process performance of SMB, have not been sufficiently investigated, especially for systems with nonlinear adsorption behavior.

In this study, we rigorously estimate uncertainty in SMB model predictions derived from uncertainty in model parameters and analyze an influence of the parameter uncertainty. We consider the separation of the mixture of two components described by the Langmuir isotherm (Bentley et al., 2013). Uncertainty in model parameters is quantified as the posterior distribution via Bayesian inference from the artificial simulation data of single-column batch experiments. We employ sequential Monte Carlo (SMC), which parallelizes numerical sampling for efficient implementation of Bayesian inference (Yamamoto et al., 2021). The predictive distributions are obtained based on the estimated parameter uncertainty. The resulting predictive distributions provide insights into the reliability of parameter estimation results and model predictions, as well as the robustness of processes and operating conditions.

2. Methodology

2.1. SMB Process Model

A system of PDAEs, called the LDF model that is widely used to describe adsorption phenomena inside columns, is adopted in this study. Details of this model are given in Schmidt-Traub et al. (2012). The mass balance in the liquid phase is,

$$\varepsilon_{\mathrm{b}} \frac{\partial C_i^j(x,t)}{\partial t} + (1-\varepsilon_{\mathrm{b}}) \frac{\partial q_i^j(x,t)}{\partial t} + u^j(t) \frac{\partial C_i^j(x,t)}{\partial x} = 0 \tag{1}$$

where $C_i^j(x,t)$ and $q_i^j(x,t)$ are concentrations in the liquid and solid phases, respectively, of component $i \in \{A,B\}$ in the $j \in \{1,2,\dots,N_{\mathrm{col}}\}$ th column at axial position x at time t; ε_{b} is the overall bed porosity; $u^j(t)$ is the superficial liquid velocity; A and B are components to be separated; and N_{col} is the number of columns in the SMB. The mass balance in the solid phase is

$$\frac{\partial q_i^j(x,t)}{\partial t} = k_i \left(q_i^{\mathrm{eq},j}(x,t) - q_i^j(x,t) \right) \tag{2}$$

where k_i is the overall mass-transfer coefficient of component i; and $q_i^{\mathrm{eq},j}(x,t)$ is the equilibrium concentration in the solid phase. This equilibrium concentration is given by the Langmuir isotherm:

$$q_i^{\mathrm{eq},j}(x,t) = \frac{H_i C_i^j(x,t)}{1 + b_A C_A^j(x,t) + b_B C_B^j(x,t)} \tag{3}$$

where H_i is Henry's constant and b_i is the affinity coefficient.

2.2. Bayesian Inference and Predictive Distribution

In this study, we estimate the predictive distribution due to uncertainty in the model parameters. Generally, an observation $\boldsymbol{y} \in \mathbb{R}^M$ is modeled using a deterministic model $f(\boldsymbol{\theta})$ and an observation error as $\boldsymbol{y} = f(\boldsymbol{\theta}) + \boldsymbol{\varepsilon}$, where $\boldsymbol{\theta}$ is the model parameter and $\boldsymbol{\varepsilon}$ is the observation error. The observation error is often assumed to follow a multivariate

normal distribution with mean $\mathbf{0}$ and a covariance matrix $\boldsymbol{\Sigma}$. Under this assumption, the probability density $p(\boldsymbol{y}|\boldsymbol{\theta})$ is given by

$$p(\boldsymbol{y}|\boldsymbol{\theta}) = (2\pi)^{-\frac{M}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}} exp\left\{-\left(\boldsymbol{y}-f(\boldsymbol{\theta})\right)^T \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y}-f(\boldsymbol{\theta})\right)\right\} \tag{4}$$

The parameters $\boldsymbol{\theta}$ in the model $f(\boldsymbol{\theta})$ are estimated from observed data $\boldsymbol{Y}^{\text{data}}$ via Bayesian inference:

$$p\left(\boldsymbol{\theta}|\boldsymbol{Y}^{\text{data}}\right) \propto p\left(\boldsymbol{Y}^{\text{data}}|\boldsymbol{\theta}\right)p(\boldsymbol{\theta}) \tag{5}$$

where $p\left(\boldsymbol{\theta}|\boldsymbol{Y}^{\text{data}}\right)$ is the posterior distribution; $p(\boldsymbol{\theta})$ is the prior distribution; and $p\left(\boldsymbol{Y}^{\text{data}}|\boldsymbol{\theta}\right)$ is the likelihood distribution formulated as $p\left(\boldsymbol{Y}^{\text{data}}|\boldsymbol{\theta}\right) = \prod_{k=1}^{N_{\text{data}}} p\left(\boldsymbol{y}_i^{\text{data}}|\boldsymbol{\theta}\right)$ when the data is given by $\boldsymbol{Y}^{\text{data}} = \left[\boldsymbol{y}_1^{\text{data}}, \boldsymbol{y}_2^{\text{data}}, \dots, \boldsymbol{y}_{N_{\text{data}}}^{\text{data}}\right]^T$.

The predictive distribution $p\left(\boldsymbol{y}|\boldsymbol{Y}^{\text{data}}\right)$ using the posterior distribution of the parameter, $p\left(\boldsymbol{\theta}|\boldsymbol{Y}^{\text{data}}\right)$ obtained by Eq. (5), is given as follows:

$$p\left(\boldsymbol{y}|\boldsymbol{Y}^{\text{data}}\right) = \int p(\boldsymbol{y}|\boldsymbol{\theta})p\left(\boldsymbol{\theta}|\boldsymbol{Y}^{\text{data}}\right) d\boldsymbol{\theta} \tag{6}$$

where $p\left(\boldsymbol{y}|\boldsymbol{Y}^{\text{data}}\right)$ is usually approximated using a Monte Carlo method.

In this study, the model $f(\boldsymbol{\theta})$ is Eqs.(1)-(3), $\boldsymbol{\theta} = [H_A, H_B, k_A, k_B, b_A, b_B]^T$, and the observation is $\boldsymbol{y} = \left[\bar{C}_{\text{Ext},A}, \bar{C}_{\text{Ext},B}, \bar{C}_{\text{Raff},A}, \bar{C}_{\text{Raff},B}\right]^T$ where $\bar{C}_i$ is the average product concentration and subscripts Ext and Raff refer to extract and raffinate, respectively. The data $\boldsymbol{Y}^{data}$ are the time-series concentration data sets obtained from the artificial simulation data of single-column batch experiments as follows: $\boldsymbol{Y}^{\text{data}} = \left[\boldsymbol{C}_1^{\text{data}}, \boldsymbol{C}_2^{\text{data}}, \dots, \boldsymbol{C}_{N_{\text{data}}}^{\text{data}}\right]^T$ where $\boldsymbol{C}_k^{\text{data}}$ refers to the kth experimental data set as $\boldsymbol{C}_k^{\text{data}} = \left\{\left(t_l, C(L,t_l)\right)\right\}_{l=1}^{M}$; L is the length of the column; and t_l is observation time in kth experiment.

3. Experimental

3.1. SMB Model, Operating and Simulation Condition

The model parameters, as well as design and operating conditions in Bentley et al. (2013) for the separation of cyclopentanone (c5) and cyclohexanone (c6) with SMB were used in this study. The final parameter values repetitively refined in their study were employed as true values in this study. Their design conditions—the number of columns $N_{\text{col}} = 4$, column length $L = 0.25$ [m]—were adopted, and their operating conditions—flow rates and step time—were also employed. The covariance matrix $\boldsymbol{\Sigma}$ in Eq.(4) was assumed to be a diagonal matrix, and the diagonal element σ_{ii} was assumed to satisfy the following equation for the element y_i of $\boldsymbol{y}$: $1.96\sigma_{ii} = 0.03y_i$ [g/L].

3.2. Artificial Data of Single-Column Experimental

Simulation of the single-column batch experiment was conducted to obtain the artificial data. Two experimental data were generated using a mixture of c5 and c6 with 34 g/L of each component as a feed and a flow rate $Q = 3.0$ [mL/min]: (A) 40 μL injection

test. (B) 20 mL injection test. For (A) and (B), the number of experimental points was $M = 1000$; the observation time was $t_M = 1200$ [s]; and Gaussian noise was added as the observation error. Bayesian inference was performed with two different data sets—Set α and Set β: Set α contains only (A) ($N_{\mathrm{data}} = 1$) while Set β contains (A) and (B) ($N_{\mathrm{data}} = 2$).

3.3. Implementation of sequential Monte Carlo

The posterior distribution of the model parameters was estimated using SMC with the likelihood tempering (Yamamoto et al., 2021). SMC approximately estimates the posterior distribution in Bayes' theorem (Eq.(5)), which cannot be solved analytically in general, from data using a large number of particles. The number of particles was 1,000 and the number of likelihood tempering depended on data sets, 24 for Set α and 31 for Set β. The prior distribution was set with the method shown in Yamamoto et al. (2021).

4. Results

4.1. Posterior Distribution of Model Parameters

Comparing the posterior distributions estimated using Set α and Set β, the posterior distributions estimated from Set β were found to have smaller uncertainty than that of Set α. Table 1 shows the mode values and 95% credible intervals of the posterior distributions from Set α and Set β, respectively. The 95% credible intervals for all parameters were narrower for Set β than for Set α; especially, the intervals of the equilibrium constants H_i and b_i were reduced from Set α to Set β by 86%, 89%, 99%, and 96%, for H_{c5}, H_{c6}, b_{c5}, and b_{c6}, respectively. This uncertainty reduction of the equilibrium constants is due to the addition of the breakthrough test (B) data, where the interaction between b_{c5} and b_{c6} sufficiently appears, making it easier to determine values of H_i and b_i.

Table 1. True values, modes, and 95% credible intervals of posterior distributions of model parameters estimated from Set α and Set β.

Parameter	True	Posterior distribution (Set α)	Posterior distribution (Set β)
H_{c5} [−]	2.13	$2.13^{+5.23\times10^{-3}}_{-6.55\times10^{-3}}$	$2.13^{+5.53\times10^{-4}}_{-1.07\times10^{-3}}$
H_{c6} [−]	3.82	$3.82^{+4.15\times10^{-3}}_{-4.75\times10^{-3}}$	$3.82^{+4.72\times10^{-4}}_{-5.36\times10^{-4}}$
k_{c5} [$\mathrm{s^{-1}}$]	0.312	$0.312^{+1.14\times10^{-3}}_{-1.44\times10^{-3}}$	$0.312^{+1.23\times10^{-3}}_{-1.13\times10^{-3}}$
k_{c6} [$\mathrm{s^{-1}}$]	1.19	$1.19^{+1.09\times10^{-2}}_{-7.26\times10^{-3}}$	$1.19^{+5.20\times10^{-3}}_{-4.22\times10^{-3}}$
b_{c5} [g/L]	1.7×10^{-2}	$1.57\times10^{-2}{}^{+7.03\times10^{-3}}_{-6.51\times10^{-3}}$	$1.70\times10^{-2}{}^{+6.88\times10^{-5}}_{-4.42\times10^{-5}}$
b_{c6} [g/L]	4.2×10^{-2}	$4.35\times10^{-2}{}^{+2.28\times10^{-3}}_{-1.67\times10^{-3}}$	$4.21\times10^{-2}{}^{+6.90\times10^{-5}}_{-1.00\times10^{-4}}$

4.2. Prediction uncertainty in internal concentration of SMB

The two parameter uncertainties estimated from Set α and Set β were propagated to the SMB model predictions. Figures 1 (a) and (b) show the simulation results at the true parameter and the uncertainty of the internal concentration profile by median and 68 %, 95 %, and 99.9 % prediction intervals (PI). These PIs were estimated by solving the SMB model with 1,000 sets of model parameters sampled from each of the posterior

distributions from Set α and Set β. For each simulation, 20 cycles of continuous operation were performed to ensure that the process reached the cyclic steady state (CSS). The width of the 99.9 % PI for the internal concentration of each component is clearly narrower in (b) than in (a).

The reduction in the uncertainty on the internal concentration profiles of Set β from Set α can be explained by the data addition of the breakthrough test (B) used for parameter estimation. In the region where the two components are mixed, indicated by the arrow in Figure 1 (a), the median of PIs deviates significantly from the true value; in the same interval in Figure 1 (b), there is almost no spread in PIs. The interaction between b_{c5} and b_{c6} may appear significantly in the interval in which the two components mix at high concentrations, where the uncertainty can be reduced remarkably by including the data of the breakthrough test (B).

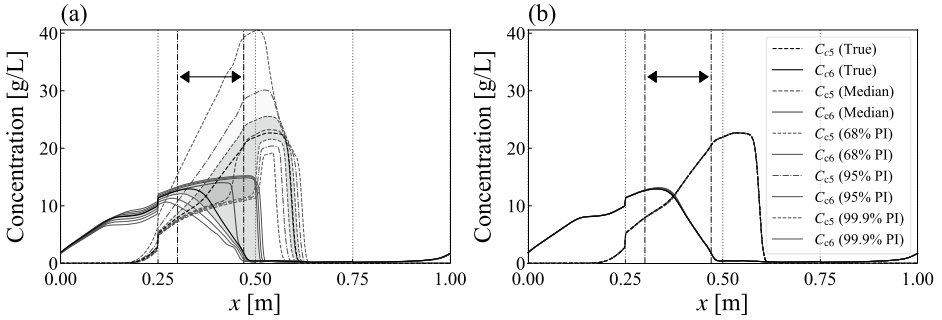


Figure 1. Simulation using the true parameter values and prediction intervals (PI) of the internal concentration profile. (a) PIs estimated from the posterior distribution from Set α. (b) PIs estimated from the posterior distribution from Set β.

4.3. Predictive Distribution of Product Purities and Recoveries

Similarly to the uncertainty of the internal concentration profile (Figure 1), the predictive distribution of product purity and recovery estimated from the posterior distribution of Set β was much less uncertain than that of Set α. Figure 2 (a1) to (a4) and (b1) to (b4) show the predictive distributions of purity $Pur_{\mathrm{Raff,c5}}$ and $Pur_{\mathrm{Ext,c6}}$ and recovery $Rec_{\mathrm{Raff,c5}}$ and $Rec_{\mathrm{Ext,c6}}$ estimated from the posterior distributions of Set α and Set β, respectively. In each graph, the abscissa is the purity or recovery, and the ordinate is the probability density, obtained from the model using the true parameters and two types of predictive distributions: one propagating only the parameter uncertainty and the other including observation error. As in Figure 1, the predictive distribution was estimated using 1,000 sets of parameters sampled from the posterior distribution. The predictive distributions for (b1) through (b4) are much sharper than those for (a1) through (a4).

By comparing the predictive distributions with and without observation error, we conclude that the influence of the parameter uncertainty estimated from Set β is sufficiently smaller than that of the observation error. For Set α, the two predictive distributions shown in Figure 2 (a1) to (a4) overlap each other, indicating that the parameter uncertainty is more dominant than the observation uncertainty. In contrast, for Set β shown in (b1) to (b4) of Figure 2, the magnitudes of the variances for the two predictive distributions differ significantly; the predictive distributions without observation error are very sharp, while those with observation error is widely distributed

as nearly symmetric distributions; i.e., the observation uncertainty given as Gaussian noise is more dominant than the parameter uncertainty. This result confirms that, under the observation error assumed in this study, the influence of the parameter uncertainty estimated from Set β on the SMB model predictions is reduced sufficiently.

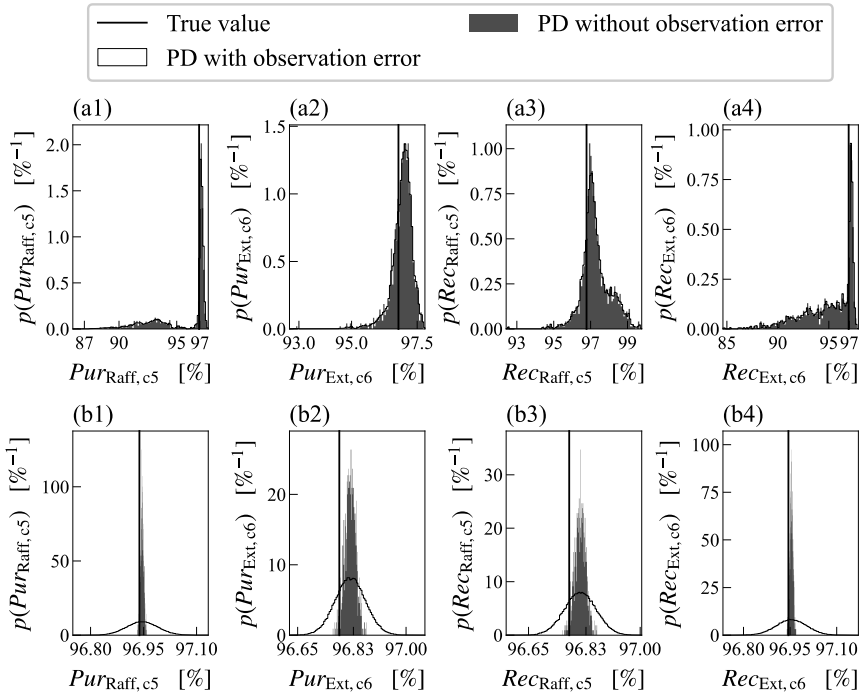


Figure 2. Predictive distributions (PD) of purity and recovery estimated from the posterior distributions of Set α—(a1), (a2), (a3), and (a4)—and of Set β—(b1), (b2), (b3), and (b4).

5. Conclusions

In this study, we quantified the prediction uncertainty in an SMB model and analyzed the influence of the parameter uncertainty on predictions from two different experiments. The resulting distributions allow us to evaluate the reliability of the model predictions and parameters and identify the necessary experiments. Estimating a predictive distribution that considers the uncertainty of the operating conditions such as flow rates and step time will be future work.

References

Bentley, J., Sloan, C., Kawajiri, Y., 2013. Simultaneous modeling and optimization of nonlinear simulated moving bed chromatography by the prediction-correction method. J. Chromatography. A 1280, 51–63.

Schmidt-Traub, H., Schulte, M., Seidel-Morgenstern, A., 2012. Preparative chromatography, 2nd ed. WILEY-VCH Verlag GmbH & Co.

Yamamoto, Y., Yajima, T., Kawajiri, Y., 2021. Uncertainty quantification for chromatography model parameters by Bayesian inference using sequential Monte Carlo method. Chem. Eng. Res. Des. 175, 223–237.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50158-5

A predictive model for multi-criteria selection of optimal thermochemical processing pathways in biorefineries

Mohammad Alherbawi, Ahmed AlNouss, Rajesh Govindan, Gordon McKay, Tareq Al-Ansari*

College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar.
*talansari@hbku.edu.qa

Abstract

The rapid growth of the global economy, combined with the growing demand for energy, environmental degradation from greenhouse gas emissions, and fluctuating fossil fuel prices, have emphasised the importance of renewable sources of energy. Biofuels produced from biomass conversion processes accounts for at least 13% of the gross global energy consumption and 70% of the world renewable energy mix. The biomass resources, including municipal, industrial and forestry waste, were proven to have a great potential for deriving various forms of energy in an affordable and reliable manner. In this regard, several thermochemical technologies have been developed to convert biomass waste into energy. However, due to the extremely heterogeneous characteristics of biomass resources, the feasibility and efficiency of these processes may greatly vary, depending on the biomass category and composition. Intensive experiments, simulations and optimisation models were developed to select the optimal processing pathway for each feedstock, which generally consume significant time and effort. To address this issue, it is desirable to seek novel and accurate mathematical representations that enable rapid performance estimation and multi-criteria selection for the optimal biomass processing pathway based on the physical properties and chemical compositions of different biomass categories, without the need for expensive experimental setup or time-consuming simulations. The objective of this study is to develop a mathematical model which links the biomass' proximate and elemental analyses to three crucial technology performance criteria: including the return on investment, energy efficiency, and carbon intensity. For this purpose, intensive simulations and sensitivity analyses were carried out using Aspen Plus to examine three main processes including gasification, pyrolysis and hydrothermal liquefaction (HTL). The comprehensive simulation data were subsequently used to develop and compare multiple meta-models for their accuracy of representation using regression algorithms. This model is believed to expedite the ongoing research on biomass thermo-processing and play a significant role towards enhancing biomass sustainability.

Keywords: Prediction model, Biomass, Gasification, Pyrolysis, Liquefaction.

1. Introduction

Biofuels, which are produced from biomass, nowadays contribute to nearly 70% of renewable mix and considered as a promising alternative fuel to reduce greenhouse gas emissions (WBA, 2019). In this context, wastes are the most abundant biomass for almost every country to be

processed into valuable energy products. While thermochemical conversion pathways (i.e., gasification, pyrolysis, HTL) are now gaining more interest at the expense of the time-consuming biochemical processes like anaerobic digestion. Nevertheless, there is no single technology can efficiently accommodate all different types of wastes due to their extremely heterogeneous nature. Therefore, the selection of technologies that can handle this diversity is very important for bio-refineries. Selecting an optimal technology to process specific biomass resources have been presented earlier using different optimization techniques (Al-Ansari et al., 2020; Alherbawi et al., 2021a). However, these models are biomass-specific and do not often accommodate a wide range of biomass resources. As such there is a persistent need to provide a general model that can predict the optimal technology to process different biomass resources based on their composition and characteristics without the need to perform any intensive experiments or simulations. The assessment and the selection of the optimal waste to energy technology is a tricky process as it is closely linked with socio-economic and environmental factors, hence an appropriate solution for the above challenge has to be arrived at based on multi-criteria decision-making. For example, Yap and Nixon employed the analytic hierarchy process (AHP) model to identify the most suited waste to energy technology in the UK and India (Yap and Nixon, 2015). It was reported that gasification is the most appropriate technology for treating wastes in UK while anaerobic digestion is the best technique for India. Therefore, this study presents a prediction model for the selection of optimal thermochemical technology based on the proximate and elemental characteristics of biomass. The targeted processes are modelled using Aspen Plus, and evaluated in terms of their technical, environmental, and economic performance for a wide set of biomass compositions. The obtained data are utilized to develop a regression prediction model that correlates the different biomass compositions to the energy recovery, the return on investment and the possible greenhouse gas reduction relative to fossil resources.

2. Methodology

2.1. Processes Modelling

Advanced System for Process Engineering (Aspen Plus V.10) software is utilised for the development of the three thermochemical processes. All processes are modelled based on the assumptions of steady-state and isothermal conditions. For gasification modelling, Peng-Robinson package is selected to estimate fluid properties. In addition, biomass is initially defined as a nonconventional component referring to its proximate and elemental attributes. The nonconventional attributes are then converted into their corresponding components using a Fortran code. Steam is used as a gasifying medium with a steam to biomass ratio of 0.75, while the process is conducted at 850 °C and atmospheric pressure (AlNouss et al., 2021). The “RGibbs” reactor is used to simulate the key reactions which operates based on the minimization of the Gibbs’ free energy. By the end of the process, biochar is collected using a cyclone, while the condensable volatiles are collected upon the cooling of the stream. The gasification process flowsheet is presented in Figure 1.

For pyrolysis process, biomass is initially dried then introduced into a grinder, whereby, the energy requirement for drying and grinding is adapted from (Cheng et al., 2020). The nonconventional stream is then converted into conventional components using a Fortran code based on the products distribution model of (Swagathnath et al., 2019). The process is simulated in an “RYield” reactor at 600 °C and atmospheric pressure with the supply of nitrogen gas to ensure an inert atmosphere (Alherbawi et al., 2021a). Char is then collected using a cyclone, while the volatile stream is cooled down for the collection of syngas and bio-oil. The process flowsheet of pyrolysis is presented in Figure 2.

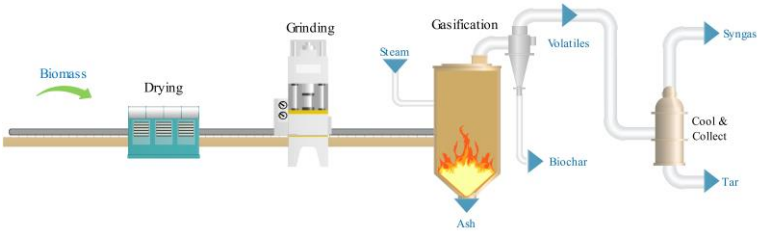


Figure 1: A simplified process flowsheet of gasification.

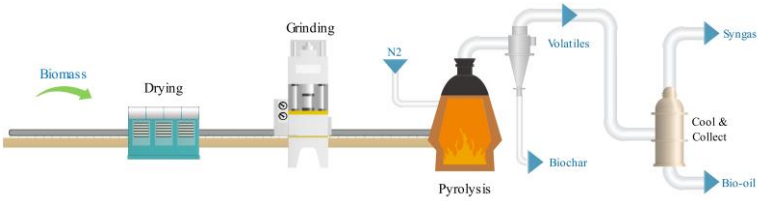


Figure 2: A simplified process flowsheet of pyrolysis.

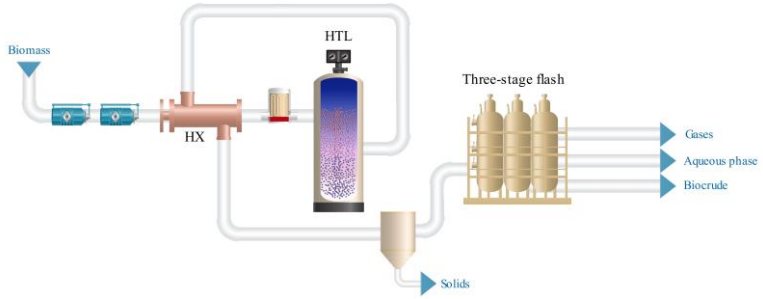


Figure 3: A simplified process flowsheet of hydrothermal liquefaction.

For HTL modelling, the non-random two-liquid (NRTL) thermodynamic package is used, while biomass is defined based on its proximate and elemental attributes. Water as added to create a slurry ensuring the organic solids do not exceed 20%. The slurry is then pumped using two consecutive high-performance pumps into the HTL reactor. The key sub-process is simulated using an "RYield" reactor, whereby, the maximum biocrude and minimum hydrochar yields are restricted based on the model developed by (Zhong and Wei, 2004) using a Fortran code. While the biocrude and hydrochar compositions are adapted from (Pedersen et al., 2017) and (Lentz et al., 2019) respectively. The syngas composition is then calculated in an "RGibbs" reactor via the minimization of the Gibbs' free energy. The process is conducted at 350 °C and 100 bar. In addition, solids are collected using a hydro-cyclone, while a three-stage flash drum is utilised to split the remaining stream into three phases: gas, biocrude and an aqueous phase. The HTL flowsheet is illustrated in Figure 3.

2.2. Technical, environmental and economic studies

Three criteria are selected to develop the prediction model for optimal processing pathway selection, including technical, environmental and economic aspects. At technical level, energy recovery is evaluated as a ratio of the generated energy to the consumed energy for each process. The generated energy is estimated based on the lower heating value of all products, while the consumed energy comprises utilities requirement and the

calorific value of biomass. Whereas for the environmental study, the global warming potential (GWP) at process level (gate-to-gate approach) is evaluated. It comprises the greenhouse gases emitted to atmosphere during the process, as well as the GWP of utilities based on an emission factor of 0.58 $kgCO_2$-e/kWh (Spath and Mann, 2000). The GWP of each process is then evaluated in relation to that of charcoal (Bhattacharya et al., 2002). Nevertheless, an economic study is conducted for all processes using Aspen Process Economic Analyzer (APEA V.10). To achieve a high accuracy of capital costs evaluation, the key equipment prices are based on actual market prices, while scaled up/inflated based on the Chemical Engineering Plant's Cost Index (CEPCI) (Alherbawi et al., 2021b). The year 2019 is considered as the base year of analysis, while all processes plants are assumed to have a feed capacity of 30 t/h and a lifespan of 25 years. The return of investment (ROI) parameter is used for the comparison between the different processes and runs, which provides a ratio between net income and investment cost.

2.3. Regression Model

Regression approach is utilized to develop a prediction model that correlates the biomass characteristics and thermochemical processes performance. Six input parameters are considered in the model including moisture, ash, carbon, hydrogen, oxygen and nitrogen contents of biomass. Whereas three responses are observed including the energy recovery (%), the GHG reduction relative to charcoal (%), and the return on investment (%). Each input parameter is varied, while the remaining parameters are normalised with reference to a baseline biomass composition. The parameters are varied as follows: moisture (10-50 %), ash (6-18 %), carbon (30-70 %), hydrogen (4-12 %), oxygen (20-40) and nitrogen (1-5 %). The fixed carbon is assumed to be 50% of total carbon for all runs, while the volatile matter is found based on the difference (100% - fixed carbon – ash). In total 54 runs are conducted and evaluated. The obtained responses are then processed to develop a regression prediction model using Excel's Data Analysis tool.

2.4. Model Evaluation

The models have then been evaluated through providing three different biomass feedstocks to test the model performance against reported results from Aspen simulation.

Table 1: Biomass types for model evaluation.

Biomass	M*	FC	VM	Ash	C	H	N	O
Food waste	70.0	33.07	61.13	5.80	43.97	4.97	2.14	43.12
MWS	7.56	24.21	57.99	17.8	48.47	5.14	1.16	27.43
Waste Plastics	0.41	0.28	97.28	2.44	83.93	12.83	0	0.80

* M: moisture content on wet basis (%).

3. Results and Discussion

The developed model is evaluated by analysing the values of regression coefficients and analysis of variance (ANOVA). Besides, the extent of fit of the model equation is expressed by the determination coefficient R^2, while the model significance is evaluated through F-significance as presented in Table 2. The predicted R^2 values of (70.57 – 99.83 %) reflect a good correlation between the actual and predicted values of the responses. While the model significance is high since the lack of fit values (F-significance) are extremely low (0.01-$8.5x10^{-13}$). A maximum "F" value of 0.01 means that only 1% of the input data may possibly not fit into the model. The model suggested that the moisture, carbon and oxygen contents are the key influencing parameters for all responses.

Table 2: Regression models and their data fitting evaluation.

	Regression model*	F-Significance	R^2
Gasification	$Energy\ recovery\ (\%) = 1.005 * C + 1.868 * O - 0.292 * M$	$2.9x10^{-07}$	97.07 %
	$Emissions\ reduction\ (\%) = 4.036 - 4.456 * C - 5.217 * O - 0.639 * M$	$2.5x10^{-05}$	92.23 %
	$Return\ on\ investment\ (\%) = 0.638 - 0.131 * C - 0.891 * O - 0.292 * M$	$8.5x10^{-13}$	99.83 %
Pyrolysis	$Energy\ recovery\ (\%) = 1.303 - 0.316 * C - 0.476 * O - 1.060 * M$	$8.2x10^{-11}$	99.52 %
	$Emissions\ reduction\ (\%) = 1.112 - 1.033 * O - 0.437 * M$	$1.0x10^{-02}$	70.57 %
	$Return\ on\ investment\ (\%) = 0.777 - 0.307 * C - 0.974 * O - 0.453 * M$	$1.5x10^{-08}$	98.49 %
HTL	$Energy\ recovery\ (\%) = -4.148 + 5.627 + 7.040 * O - 0.961 * M$	$2.8x10^{-07}$	97.10 %
	$Emissions\ reduction\ (\%) = 2.619 - 2.006 * C - 2.887 * O + 0.248 * M$	$1.6x10^{-04}$	88.28 %
	$Return\ on\ investment\ (\%) = -0.742 + 1.088 * C + 1.263 * O - 0.3 * M$	$1.5x10^{-04}$	88.40 %

* M: moisture content on wet basis (%), C & O: carbon and oxygen mass contents on dry basis (%).

The three responses are combined in one formula to indicate an overall efficiency for each process, considering an equal relative weight for each response. The final prediction model is presented in Eq.(1-3). The overall efficiency is a function of biomass characteristics which facilitates selecting the optimal processing pathway for each biomass, by satisfying the technical, environmental and economic aspects of the process. The highest overall efficiency amongst the three processes indicates the optimal technology for the tested biomass.

$$\boldsymbol{Gasification\ overall\ efficiency}\ (\%) = 1.668 - 1.057 * C - 1.413 * O - 0.408 * M \quad (1)$$

$$\boldsymbol{Pyrolysis\ overall\ efficiency}\ (\%) = 1.064 - 0.2076 * C - 0.828 * O - 0.65 * M \quad (2)$$

$$\boldsymbol{HTL\ overall\ efficiency}\ (\%) = -0.757 + 1.5698 * C + 1.8057 * O - 0.338 * M \quad (3)$$

The model is evaluated using the three biomass feedstocks in Table 1. The results as illustrated in Figure 4 demonstrate the excellence of HTL in the case of food waste that is high in moisture, fixed carbon and oxygen, gasification in the case of municipal solid waste (MSW) that is moderate in attributes, and pyrolysis in the case of waste plastic that is high in volatile matter and carbon content.

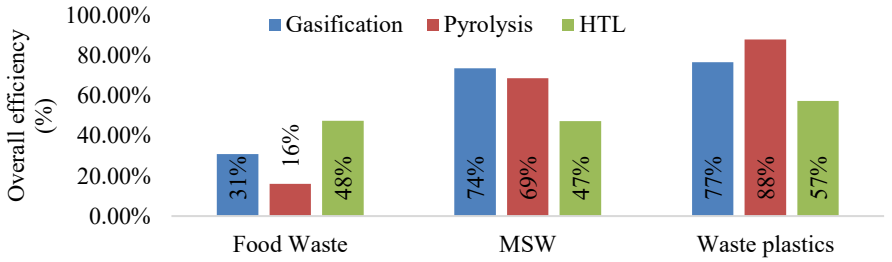


Figure 4: Results of model evaluation.

4. Conclusions

The study in hand presented a prediction model for the selection of optimal thermochemical processing pathway without the need for the costly experiments or time-consuming simulations. Whereby, the prediction model satisfies the technical, environmental and economic aspects of the selected optimal process based on the proximate and elemental attributes of different biomass resources. As such, this model is expected to contribute to the development of sustainable biomass-based energy production and expediate the ongoing research in biorefining field.

References

Al-Ansari, T., AlNouss, A., Al-Thani, N., Parthasarathy, P., ElKhalifa, S., Mckay, G., Alherbawi, M., 2020. Optimising Multi Biomass Feedstock Utilisation Considering a Multi Technology Approach, in: Pierucci, S., Manenti, F., Bozzano, G.L., Manca, D.B.T.-C.A.C.E. (Eds.), 30 European Symposium on Computer Aided Process Engineering. Elsevier, pp. 1633–1638.

Alherbawi, M., AlNouss, A., McKay, G., Al-Ansari, T., 2021a. Optimum sustainable utilisation of the whole fruit of Jatropha curcas: An energy, water and food nexus approach. Renew. Sustain. Energy Rev. 137, 110605.

Alherbawi, M., McKay, G., Mackey, H.R., Al-Ansari, T., 2021b. A novel integrated pathway for Jet Biofuel production from whole energy crops: A Jatropha curcas case study. Energy Convers. Manag. 229, 113662.

AlNouss, A., Mckay, G., Al-Ansari, T., 2021. Utilisation of Carbon Dioxide and Gasified Biomass for the Generation of Value Added Products, in: Türkay, M., Gani, R.B.T.-C.A.C.E. (Eds.), 31 European Symposium on Computer Aided Process Engineering. Elsevier, pp. 1567–1572.

Bhattacharya, S.C., Albina, D.O., Abdul Salam, P., 2002. Emission factors of wood and charcoal-fired cookstoves. Biomass and Bioenergy 23, 453–469.

Cheng, F., Luo, H., Colosi, L.M., 2020. Slow pyrolysis as a platform for negative emissions technology: An integration of machine learning models, life cycle assessment, and economic analysis. Energy Convers. Manag. 223, 113258.

Lentz, Z., Kolar, P., Classen, J., 2019. Valorization of Swine Manure into Hydrochars. Processes 7, 560.

Pedersen, T.H., Jensen, C.U., Sandström, L., Rosendahl, L.A., 2017. Full characterization of compounds obtained from fractional distillation and upgrading of a HTL biocrude. Appl. Energy 202, 408–419.

Spath, P.L., Mann, M.K., 2000. Life Cycle Assessment of a Natural Gas Combined-Cycle Power Generation System.

Swagathnath, G., Rangabhashiyam, S., Parthsarathi, K., Murugan, S., Balasubramanian, P., 2019. Modeling Biochar Yield and Syngas Production During the Pyrolysis of Agro-Residues BT - Green Buildings and Sustainable Engineering, in: Drück, H., Pillai, R.G., Tharian, M.G., Majeed, A.Z. (Eds.), . Springer Singapore, Singapore, pp. 325–336.

WBA, 2019. Global Bioenergy Statistics 2019.

Yap, H.Y., Nixon, J.D., 2015. A multi-criteria analysis of options for energy recovery from municipal solid waste in India and the UK. Waste Manag. 46, 265–277.

Zhong, C., Wei, X., 2004. A comparative experimental study on the liquefaction of wood. Energy 29, 1731–1741.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50159-7

Numerical Investigation of the Shear Rate Variation in Cooling Crystallization

Ken-Ichiro Sotowa[a*], Soranasataporn Pattana[a], Osamu Tonomura[a] and Sanghong Kim[b]

[a]*Dept. of Chemical Engineering, Kyoto University, Nishikyo-ku, Kyoto 615-8510 Japan*
[b]*Dept. of Applied Physics and Chemical Engineering, Tokyo University of Agriculture and Technology, Naka-cho, Koganei, Tokyo 184-8588 Japan*
sotowa@cheme.kyoto-u.ac.jp

Abstract

A recent numerical study on the mass transfer around a crystal showed that the mass transfer rate in a shear flow is greater than that in a uniform flow. In this study, we also demonstrated that this effect was greater for small crystals. This implies that under shear flow conditions, fine crystals grow faster, and a greater number of effective nuclei can be generated. In this study, the effect of shear flow on the size distribution of the product crystals was investigated. The evolution of the crystal size distribution in a batch cooling crystallizer was simulated by solving a population balance equation. Because batch crystallizers are operated under unsteady state conditions, the consequence of applying a shear in the early stage of the operation is expected to differ from that in the final stage. The numerical results obtained in this study showed that the crystal size distribution could be controlled by varying the profile of the shear rate.

Keywords: cooling crystallization, population balance equation, mass transfer, shear rate, dynamic operation

1. Introduction

Crystallization is an important class of unit operation for purification. The quality of product crystals is largely determined by their size distribution. There are several different operations that create supersaturation, which is the driving force behind crystal deposition. One of the typical operations is cooling crystallization, in which the solution temperature is lowered such that the saturation concentration decreases below the solute concentration. In cooling crystallization, the temperature profile is the only parameter that can be manipulated to control the crystal size distribution. The impact of manipulating the temperature profile on the crystal size distribution was demonstrated using the programmed cooling method (Mullin and Nývlt, 1971).

Recently, a new type of crystallizer, which utilizes the Taylor–Couette flow, has drawn considerable attention because it allows better control over the crystal size (Nguyen et al., 2017). A Taylor–Couette crystallizer comprises two concentric cylinders with a small gap. The inner cylinder rotates and produces a uniform shear flow in the solution in the gap. A previous study claimed that this type of crystallizer could produce finer crystals with a smaller coefficient of variation (Nguyen et al., 2017). The characteristics of the Taylor–Couette apparatus have been investigated experimentally (Nguyen et al, 2010, Khuu et al, 2016), but the physics behind the improved size distribution has not been clarified.

Notably, the fluid in a stirred tank, which is used in cooling crystallization, is not uniform with respect to the mixing intensity. The turbulent intensity is high and the fluid rapidly deforms near the impeller, whereas circulating motion with the fluid around the mixing axis is dominant when the fluid element and the impeller are distant. The fluid motion in the Taylor–Couette crystallizer is different from that in the mixing tanks. The fluid is subject to a uniform and high shear rate, which differentiates the Taylor–Couette crystallizer from the stirred-tank crystallizer.

A recent numerical study on the mass transfer around a crystal showed that the mass transfer rate in shear flow is greater than that in a uniform flow (Kitagawa, 2018). The study also showed that this effect was magnified for small crystals. This implies that under shear flow conditions, the fine crystals grow faster, and a greater number of effective nuclei can be generated.

This study focused on the effect of the shear rate on the size distribution of crystals obtained from a cooling crystallizer. A simple model was developed to express the growth-enhancing effects of the shear flow. The evolution of the crystal size distribution was calculated using the population balance equation. The results showed that an increase in the shear rate reduced the average size of the crystals. The simulation study also showed that by dynamically changing the shear rate, both the average size and coefficient of variation can be controlled.

2. Numerical model

The variation in the crystal size distribution in the cooling crystallizer was modeled using a population balance equation (Eq. 1).

$$\frac{\partial n}{\partial t} + \frac{\partial (Gn)}{\partial L} = 0 \qquad n|_{L=0} = \frac{B}{G|_{L=0}}, \tag{1}$$

where n is the population density (number weighted) [#·m^{-1}·kg^{-1}], t is the time [s], L is the particle diameter, G is the growth rate [m·s^{-1}], and B is the nucleation rate [#·s^{-1}·kg^{-1}]. The total mass balance of the solute is given by Eq. 2.

$$\frac{\mathrm{d}C}{\mathrm{d}t} = -3\rho_c k_\mathrm{v} \int_0^\infty nL^2 G \mathrm{d}L \tag{2}$$

Here, k_v and ρ_c are the volumetric shape factor [-] and crystal density [kg·m^{-3}], respectively. C [kg·(kg solvent)$^{-1}$] is the solute concentration in the bulk liquid. The solution of Eq. 2 is used to calculate the degree of supersaturation and subcooling, which are the driving forces of crystallization.

The kinetic parameters and solubility data for the numerical study were taken from a study on the cooling crystallization of potassium sulfate (Kobari et al., 2011, Kobari et al., 2012). The growth and nucleation rates are described as power functions of the degree of subcooling.

The kinetic data obtained from the literature were identified using data obtained from an experiment using a stirred tank. It is anticipated that each crystal in a stirred tank is subjected to a uniform flow of solution for most of the time. Kitagawa et al. (2018) numerically simulated mass transfer around a particle under uniform and shear flows. The results indicated that shear flow enhanced the mass transfer around the particles, in turn,

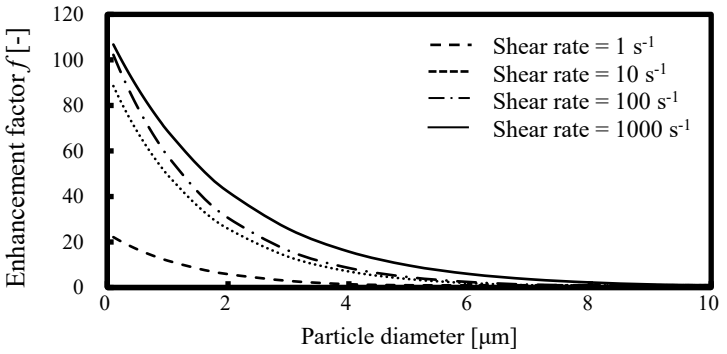


Figure 1 Enhancement factor as a function of the diameter and shear rate.

suggesting that the application of shear leads to increased nucleation and growth rates. The mass transfer data calculated by Kitagawa et al. (2018) were analyzed in this study, and the ratio of mass transfer rate under shear flow to that under uniform flow was expressed as a function of particle size and the shear rate (this ratio is referred to as the enhancement factor f [-] in the sequel).

To account for the effect of the shear rate on the crystallization, the growth rate was assumed to be greater than that obtained in a stirred tank experiment by a factor of f. This assumption holds when the rate-limiting step of crystal growth is the mass transfer of the solute from the bulk to the crystal surface. Another important assumption used in this study is that the mass transfer rate also results in an increase in the nucleation rate by a factor of f for L=0. In crystallization experiments, an increase in the stirring rate results in the formation of finer product crystals; thus, it is widely accepted that intensified mixing results in a higher nucleation rate. Because mixing is an act of enhancing mass transfer, this observation qualitatively justifies the assumption of the effect of shear rate on the nucleation rate.

The population balance equation (Eq. 1) was discretized using a backward difference scheme. A simulation program based on the Euler method was coded using the Python programming language. The simulation conditions were determined to reflect real crystallizer operations. No crystals existed at the beginning of cooling. The initial concentration of the solute was 0.199 kg·(kg solvent)$^{-1}$, and the initial temperature was the saturation temperature of the solution. The temperature of the solution was lowered by 20 °C at a constant rate.

3. Results and discussion

A simulation of the cooling crystallizer was conducted for different cooling and shear rates. The resulting number-based average size decreased with increasing shear rate (Fig. 2). The model used in this study was built on the assumption that shear enhances both the nucleation and growth rate. An increased nucleation rate reduces the size of the crystals, but a greater growth rate leads to increased crystal size. The simulation results indicate that the nucleation rate has a stronger effect on the crystal size than the growth rate under the current simulation conditions.

The effect of shear on the coefficient of variation (CV) is summarized in Fig. 3. The range of CV variation was approximately 0.28–0.38. With increasing cooling rate, CV

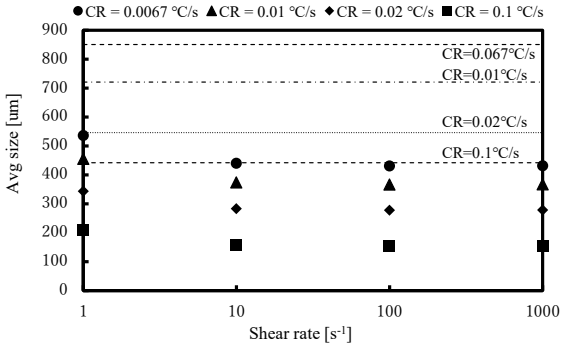


Figure 2 Effect of shear rate and cooling rate (CR) on the number-based average size. Plots show the results with shear, whereas lines indicate the average size of crystals when no shear was applied.

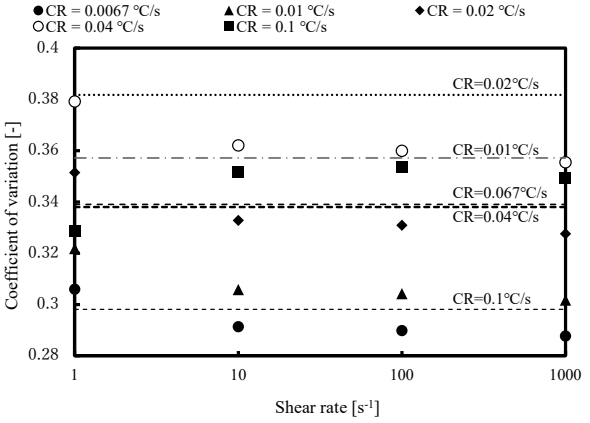


Figure 3 Effect of shear rate and cooling rate (CR) on the coefficient of variation. Plots show the results with shear, whereas lines indicate the average size of crystals when no shear was applied.

increased but reached a maximum when the cooling rate was 0.02 °C/s in the crystallizer operation without shear. The maximum CV was also recorded when the cooling rate was 0.04 °C/s in an operation with shear.

Because the cooling crystallizer is always in an unsteady state, the dynamic change in shear rate is expected to be a useful approach for controlling the crystal size distribution. Thus, we investigated the effect of stepwise changes in the shear rate on the crystal size distribution. In this simulation, the shear rate was zero at the beginning of the simulation and increased to 1000 s^{-1} at a certain time. Fig. 4 shows the variation in the number-based average size by the time of the step increase in the shear rate. The average size can be further reduced by applying a shear at the right time. When the shear started at the early stage of crystallization, the average size was insensitive to the time of the step increase. If the shear rate was changed near the end of the operation, the enhancement effect was small, and the average size approached that measured under the no shear condition. It was revealed that the maximum degree of subcooling was observed when approximately 1500

s lapsed. When shear was applied at this moment, numerous nuclei were formed, which, in turn, led to a reduction in the average size. However, when shear was applied after 1500 s, CV significantly increased, and then reached 0.7 when shear was applied at 1900 s. When the shear flow was applied too late, the nucleation process that was already ending started again, and this produced a second group of crystals.

The consequence of the step decrease in the shear rate was also examined. In this case, the shear rate was 1000 s^{-1} from the beginning of the simulation and reduced to 0 s^{-1} in a stepwise manner. The CV increased slightly and then decreased significantly as the step change time was delayed. The increase was caused by the change in shear rate, which resulted in the formation of two groups of crystals. However, as the high shear rate was applied longer, the number of crystals in the first group significantly increased because the degree of subcooling increased with time and greatly promoted the nucleation rate. At a certain point, this first group of particles, formed when the shear rate was high, will dominate over the group of crystals formed later. Consequently, the resulting crystal size distribution was nearly unimodal, and the overall CV decreased. However, if the shear flow was applied for too long, an excess number of particles would form and the average size would decrease, causing the CV to increase again.

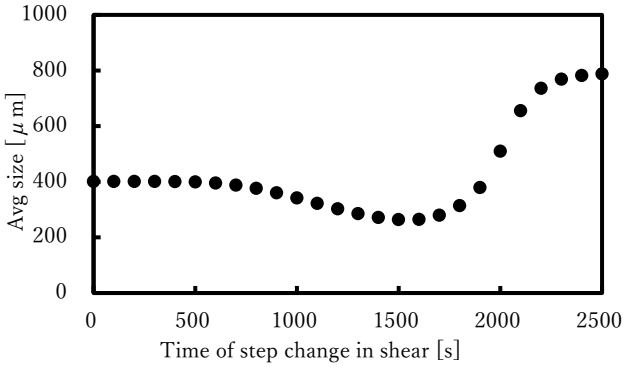


Figure 4 Time of step increase in the shear on the average size

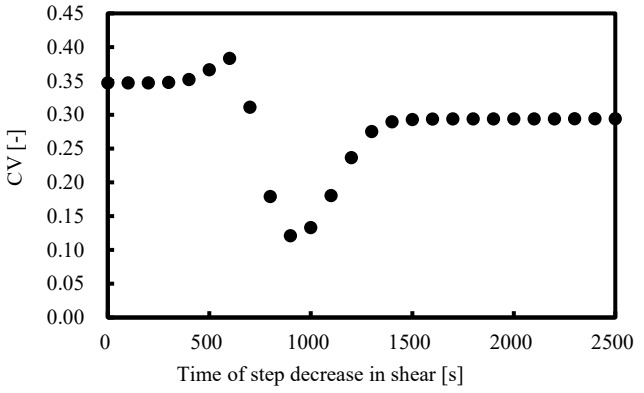


Figure 5 Time of step decrease in the shear on the average size

4. Conclusions

In this study, a model of a cooling batch crystallizer that reflects the effect of shear flow was developed. Various shear profiles were assumed, and their effects on crystal size distribution were investigated. It was demonstrated that shear flow could be utilized to reduce both the average size and coefficient of variation. It is anticipated that a further study on the optimal profile of shear rate will provide more insight into the operation of Taylor–Couette crystallizers.

Acknowledgements

This work was supported by JSPS KAKENHI (Grant Number 17H01339).

References

C.Q. Khuu, T.G. Dang, D.T. Nguyen, T.H. Le, T.T. Phan, T.T. Trinh and A.T. Nguyen, 2016, "Crystallization in Couette-Taylor crystallizer: Effect of Taylor vortices flow on the nucleation and reconstruction of L-glutamic acid in cooling crystallization," Journal of Science and Technology, **54**(5), pp. 625–631.

N. Kitagawa, K.-I. Sotowa, T. Horikawa and J.R. Alcantara-Avila, 2018, "The effect of shear flow on mass transfer rate around a crystal," The 31st International Symposium on Chemical Engineering, OF 07.

M. Kobari, N. Kubota and I. Hirasawa, 2011, "Computer simulation of metastable zone width for unseeded potassium sulfate aqueous solution," Journal of Crystal Growth, **317**(1), pp. 64–69.

M. Kobari, N. Kubota and I. Hirasawa, 2012, "Secondary nucleation-mediated effects of stirrer speed and growth rate on induction time for unseeded solution," CrystEngComm, **14**(16), pp. 5255–5261.

J.W. Mullin and J. Nývlt, 1971, "Programmed cooling of batch crystallizers," Chemical Engineering Science, **26**(3), pp. 369–377.

A.T. Nguyen, J. Kim, S. Chang and W. Kim, 2010, "Taylor vortex effect on phase transformation of guanosine 5-monophosphate in drowning-out crystallization," Industrial and Engineering Chemistry Research, **49**(10), pp. 4865–4872.

A.T. Nguyen, T. Yu and W. Kim, 2017, "Couette–Taylor crystallizer: Effective control of crystal size distribution and recovery of l-lysine in cooling crystallization," Journal of Crystal Growth, **469**, pp. 65–77.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50160-3

Application of machine learning model to optimization of the hydrogen liquefaction process

Seongwoong Min[a], Amjad Riaz[a], Muhammad Abdul Qyyum[b], Hansol Choe[a], Sang-gi Moon[a], Moonyong Lee[a*]

[a]*Process Systems Design & Control Laboratory, School of Chemical Engineering, Yeungnam University, Gyeongsan, 38541 Republic of Korea*
[b]*Department of Petroleum & Chemical Engineering, Sultan Qaboos University, Muscat, OMAN*
mynlee@yu.ac.kr

Abstract

Process optimization is an essential step for a feasible and cost-effective process design. However, problems arise due to more decision variables within the process for a complex system. In complex processes, many design and target variables have large nonlinear relationships that make it challenging to achieve the global optima. This study attempted to overcome the complexity of hydrogen liquefaction and nonlinear optimization processes. Optimize the simulation model to measure the time taken and compare it to the estimate made by the machine learning model. The basic process was simulated in Aspen HYSYS, while the machine learning model and optimization were carried out using the particle swarm optimization algorithm in MATLAB environment. Interestingly, the model optimization results of the ANN approach showed significantly less time and improved prediction.

Keywords: hydrogen liquefaction; machine learning; optimization; surrogate

1. Introduction

Liquefaction is one of the methods for storage and transportation of hydrogen (Durbin & Malardier-Jugroot, 2013), which is considered the future energy source (International Energy Agency, 2019). The liquefaction of hydrogen occurs at a very low temperature, 20 K, making it an energy-intensive process. Therefore, it is critical to minimize energy consumption through optimization in design, such as process simulation of the hydrogen liquefaction process. However, due to the fundamental problem of many optimization variables, direct optimization of the simulation process requires a considerable amount of time and effort, even with high-performance computers. In this situation, the applicability of machine learning techniques has great potential, as foretold by Qadeer et al. (2018) in their study for offshore LNG units.

Machine learning is a field of artificial intelligence technology that implements human intelligence through computers. Learning techniques include supervised learning, unsupervised learning, and reinforcement learning. The computers trained in these methods provide predictions or judgments for their models (Batta, 2020; Lecun et al., 2015). When a model is well-learned about a relationship, it is also called a surrogate model because it can provide the same results as a specific ‘real’ model that actually describes the relationship.

In this study, the relationship between design variables and performance parameters of the hydrogen liquefaction process was learned through machine learning, and the optimization results and computational load of two models optimized with the same algorithm were compared. It is expected to improve the computational load of a complex hydrogen liquefaction process using a well-trained surrogate model.

2. Hydrogen liquefaction Process

In this study, a triple mixed refrigerant (TMR) cycles hydrogen liquefaction process is considered from our previous works (Riaz et al., 2021), from now on called as Base Model (BM). BM process uses different mixed refrigerants (MR) for three refrigeration cycles; Precooling MR, Cooling MR, and Liquefaction MR's refrigeration. Figure 1 shows the process flowsheet diagram, whereas Table 1 shows feed conditions and mass flow rate of gaseous hydrogen (GH2) and liquid hydrogen (LH2) (Aasadnia & Mehrpooya, 2018).

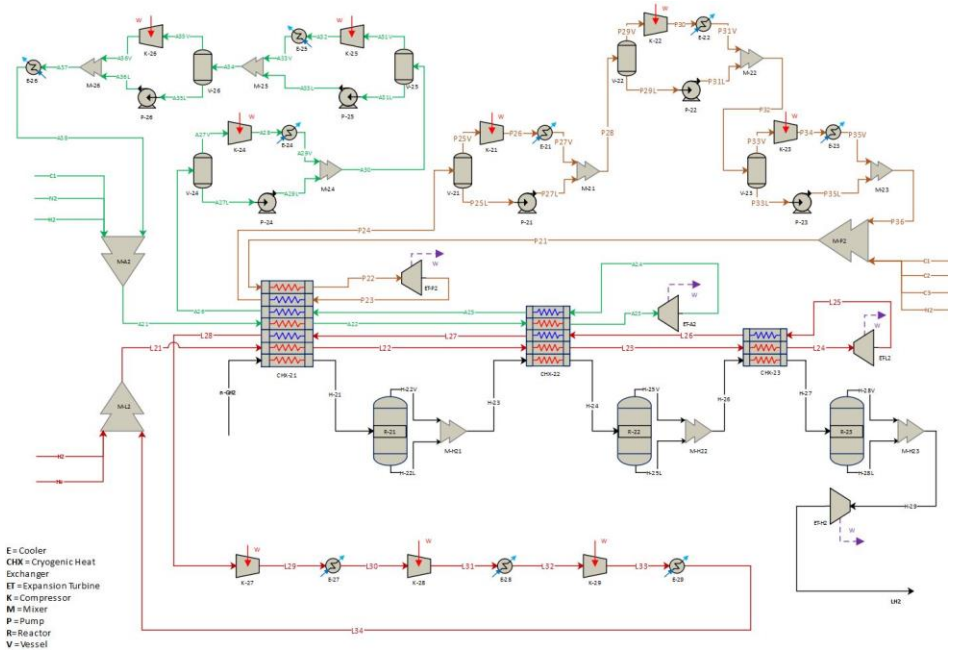


Figure 1 Process flowsheet of the BM (Riaz et al., 2021)

The hydrogen feed stream is precooled to 118 K and then sent to the ortho-para conversion (OPC) reactor. At 118 K, conversion reaches 13% and increases the share of para-hydrogen from 25 % to 35 %. At 34 K, the outlet stream flows to the next OPC reactor, increasing the para-hydrogen composition to 92.6 % with a percent conversion of 88.5 %. Similarly, the third reactor increases the molar percentage to 99.4 % by achieving 92 % conversion. As a result, a 100 % saturated liquid hydrogen (LH_2) is obtained as a product.

In each cycle, the hot pressurized refrigerant stream is cooled to the same temperature as the H_2 stream. The respective cold stream gives its pressure energy in the form of useful work in an expander. The resultant cold stream is returned through the cryogenic heat

exchanger (CHX) as cooling media. These streams enter the compression section at the hot end to close the loop.

Table 1 Process conditions of the gaseous and liquid hydrogen

Parameter		**GH2**	**LH2**
Temperature, K		298	21
Pressure, bar		21	1.3
Mass Flowrate, kg/s		1	1
Molar composition, %	Ortho-	74.92	0.59
	Para-	25.08	99.41

The design parameters and assumptions for the hydrogen liquefaction process presented in Table 1 and Table 2 are kept the same as BM. The low temperature of the cooling cycle is achieved because of a relatively high content of nitrogen and hydrogen in an optimum ratio. Also, it approaches the critical temperature of H_2 (33K). In terms of hydrogen, the high flow rate of H_2 increases the power requirement. In the refrigerant stream, Helium, one of the refrigerant candidates, in the refrigerant stream decreases the power needed by slightly increasing the flow rate. The presence of a high-boiling or low-boiling part as refrigerant results in little or no waste thermodynamically. For reducing energy consumption, it needs an optimal combination of refrigerants.

Table 2 Details of the BM refrigeration cycles (refrigerant flowrates and suction/discharge pressures

Component	**Precooling cycle**	**Cooling cycle**	**Liquefaction cycle**
C1 [kg/s]	4.683	2.280	
C2 [kg/s]	1.438		
C3 [kg/s]	18.291		
N2 [kg/s]	2.075	23.021	
H2 [kg/s]		3.185	1.640
He [kg/s]			0.173
Suction P [bar]	3.8	2.9	4.4
Discharge P [bar]	42	59	34

3. Machine learning application result

This study analyzed the results by utilizing the possibility of imitation of machine learning-based models for complex relationships between inputs and outputs. Furthermore, the results were compared by applying the optimization to the simulated base model (BM) and the trained machine learning model (MLM).

The base model is the TMR Cycles H_2 liquefaction process simulated in Aspen HYSYS, and the MLM is trained in MATLAB with a fully connected neural network structure (Figure 2). The learning of the MLM used about 1,600 randomly generated data from the BM case study. In machine learning, data were used for training/validation/testing at a ratio of 70:15:15. The result of machine learning is shown in Figure 3. As shown in Figure 3, R is greater than 0.999, and the mean squared error, which means the performance of the MLM, is also 0.01, so it can be seen that the MLM is well learned, and the BM is described with high accuracy.

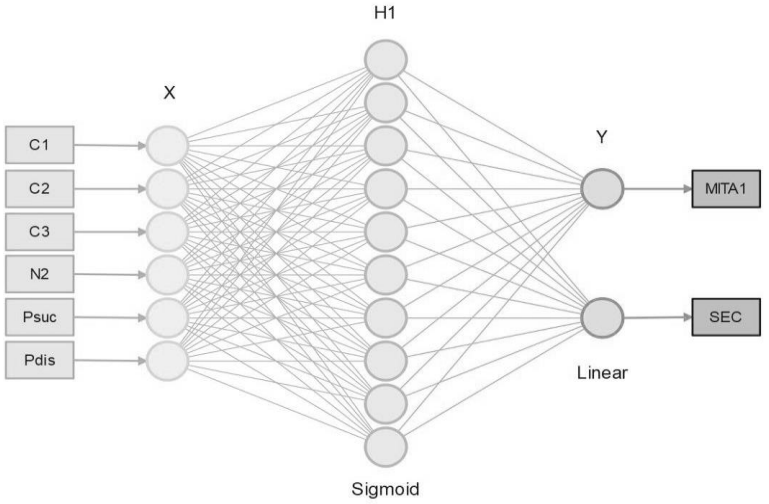


Figure 2 Artificial neural network structure of MLM

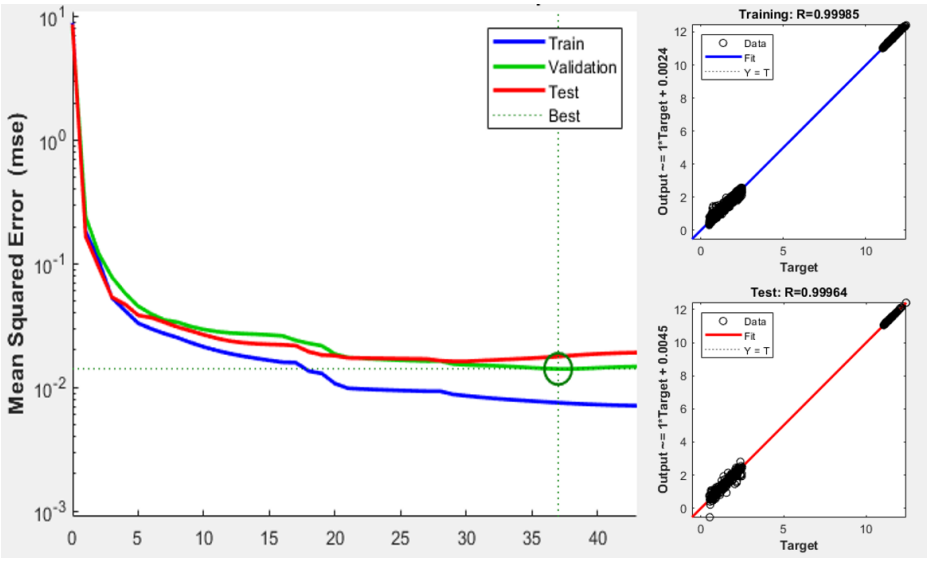


Figure 3 Training result of MLM

Both models were optimized using the Particle Swarm Optimization algorithm (PSO) (Zhou & Liao, 2013) programmed in MATLAB. The PSO algorithm is an algorithm that mathematically models the social information sharing behavior of individuals belonging to groups of ants, fish, and birds. From a probabilistic point of view, it is effective to find the global optimum. We tried to learn and optimize the MLM for all existing models. However, there were too many decision variables, so we focused on the precooling section of the hydrogen liquefaction process for preliminary research.

Therefore, there are 15 determinant variables in the base hydrogen liquefaction process. However, only six variables were considered: the precooling unit's suction and discharge pressures and the mass flow rate of the mixed refrigerant components. The optimization will proceed to minimize the specific energy consumption (SEC) of the hydrogen liquefaction process with these six variables. The constraints to be applied during optimization were kept such that the heat exchanger's minimum temperature approach (MITA) was not less than 1 ℃. Table 3 shows the PSO results and elapsed time of BM and MLM.

Table 3 PSO best position and optimal parameters of BM and MLM

Input variables	**BM**	**MLM**
N2 Mass Flow [kg/s]	1.746	1.8475
C1 Mass Flow [kg/s]	5.103	5.3193
C2 Mass Flow [kg/s]	1.604	0.72
C3 Mass Flow [kg/s]	20.170	20.2719
Discharge Pressure [kPa]	4,600.456	5,459.604
Suction Pressure [kPa]	522.245	570
Optimal output parameters		
MITA [℃]	0.99943	1.041078
SEC [kWh/kgH2]	11.040	11.01352
Elapsed time [s]	43,133.30073	3.243759

In Table 3, the values of the six decision variables and the optimal parameter values are slightly different, but it can be seen that they are within an acceptable range. On the other hand, the optimization time took 10,000 times more for BM than for MLM. This is due to the fundamental difference between the mathematical calculation models of BM and MLM. BM is a fundamental process simulation model, and it is a result value derived through material and energy balance and thermodynamic calculations. Various processing units between input and output delay the calculation time even more. However, on the other hand, MLM has a very light neural network structure to have many advantages in terms of time.

4. Conclusions

This study checked the machine learning model's performance as an imitation. We analyzed how well the self-learning model imitates the existing model and how different

the results are for certain tasks under the same conditions. Based on these results, for the analysis of the hydrogen value chain in various scenarios in the future hydrogen economy society, the limited system in this study can be expanded with a deeper and more advanced surrogate model. However, the system's expansion, such as the entire hydrogen liquefaction process, is accompanied by more determinants, making it difficult to collect and learn data. Therefore, it is necessary to actively utilize artificial intelligence and machine learning technology. It is essential to introduce big data processing techniques such as dimension reduction using principal components analysis.

Acknowledgment
This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (2021R1A2C1092152); Priority Research Centers Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2014R1A6A1031189); "Human Resources Program in Energy Technology" of the Korea Institute of Energy Technology Evaluation and Planning (KETEP), granted financial resource from the Ministry of Trade, Industry & Energy, Republic of Korea. (*No. 20204010600100*).

References

(IEA), I. E. A. (2019). *The Future of Hydrogen*. OECD. https://doi.org/10.1787/1e0514c4-en

Aasadnia, M., & Mehrpooya, M. (2018). Conceptual design and analysis of a novel process for hydrogen liquefaction assisted by absorption precooling system. *Journal of Cleaner Production*, *205*, 565–588. https://doi.org/10.1016/j.jclepro.2018.09.001

Batta, M. (2020). Machine Learning Algorithms - A Review. *International Journal of Science and Research (IJ*, *9*(1), 381-undefined. https://doi.org/10.21275/ART20203995

Durbin, D. J., & Malardier-Jugroot, C. (2013). Review of hydrogen storage techniques for on board vehicle applications. *International Journal of Hydrogen Energy*, *38*(34), 14595–14617. https://doi.org/10.1016/j.ijhydene.2013.07.058

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Qadeer, K., Qyyum, M. A., & Lee, M. (2018). Krill-Herd-Based Investigation for Energy Saving Opportunities in Offshore Liquefied Natural Gas Processes [Research-article]. *Industrial and Engineering Chemistry Research*, *57*(42), 14162–14172. https://doi.org/10.1021/acs.iecr.8b02616

Riaz, A., Qyyum, M. A., Min, S., Lee, S., & Lee, M. (2021). Performance improvement potential of harnessing LNG regasification for hydrogen liquefaction process: Energy and exergy perspectives. *Applied Energy*, *301*(April), 117471. https://doi.org/10.1016/j.apenergy.2021.117471

Zhou, F., & Liao, Z. (2013). A particle swarm optimization algorithm. *Applied Mechanics and Materials*, *303–306*, 1369–1372. https://doi.org/10.4028/www.scientific.net/AMM.303-306.1369

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50161-5

Density Functional Theory on the CO_2 Absorption Process with Ionic Liquids

Diego Román-Montalvo, Myrna H. Matus*

[a]*Instituto de Química Aplicada, Universidad Veracruzana, Av. Luis Castelazo Ayala S/N, Col. Industrial-Ánimas, Xalapa, Ver., 91190, Mexico*
myhernandez@uv.mx

Abstract

CO_2 is a pollutant which is commonly produced through industrial processes. In this work, Computational Chemistry tools are applied to seven choline based ionic liquids (ILs), which are used as an alternative to traditional organic solvents in the CO_2 extraction. The study was performed at the PBE-D3/6-311++G** level of calculation, including the solvent effect (water in this case), thus analyzing weak interactions. It was found that these ILs form hydrogen bond networks which favors the complex formation with CO_2, therefore this allows a more efficient absorption process. In addition, the complex formation process was found to be close to the thermodynamic equilibrium, which is key in the IL recovery, with the aim to be used in several CO_2 absorption cycles.

Keywords: Ionic liquids, CO_2 absorption, computational chemistry.

1. Introduction

Production of green-house gases is a worldwide problem linked to global warming and climate change. One of the main greenhouse gases is carbon dioxide (CO_2), which is generated as a byproduct in a variety of chemical reactions, such as combustion, which are part of all kinds of industrial processes (Lamb *et al.*, 2021). Therefore, research has focused on diverse strategies to diminish CO_2 production, which otherwise would be released to the environment. In this case, certain solvents are used to absorb and capture these green-house gases; however, organic solvents, such amines, are used for this purpose thus producing other waste which also generates pollution (Mazari, 2015).

For this reason and guided by Green Chemistry principles (Anastas & Eghbali, 2010), there is an interest in employing more environmentally friendly solvents, such as ionic liquids (ILs). ILs present many interesting properties, for example, low vapor pressure, high thermal stability, low toxicity, high solvation efficiency, and wide versatility; in addition, they can be recovered once the process has been carried out, thus being reutilized for several cycles. In that way, they are considered as green solvents and they may be a good choice for replacing traditional solvents (Plechkova & Seddon, 2008; Weingärtner, 2008).

To develop efficient absorption processes, it is necessary that ILs present a very good affinity to the CO_2. In previous theoretical studies, ILs with amino-acids as anions have been used for this purpose, for example, with amino-groups (Kasahara *et al.*, 2016), 1–butyl–3–methylimidazolium (Noorani & Mehrdad, 2020), or phosphonium-groups (Shaikh *et al.*, 2020) as the corresponding cations. In this work, Computational Chemistry

tools are proposed to carry out a thermochemical analysis for the CO_2 absorption with choline based and amino-acids ILs, as well as a non-covalent interactions analysis for the IL-CO_2 complexes under study. The selection of choline *versus* other cations is due to its non-toxicity, since it is considered an essential dietary amine (Rakkar & Hillier, 2007), therefore these ILs can be cataloged as green solvents (Anastas & Eghbali, 2010).

2. Methodology

In this study, the choline cation ([Ch^+]) was employed together with the organic anions: alaninate ([Ala^-]), butanoate [But^-], propanoate [$EtCO2^-$], glycinate [Gly^-], leucinate [Leu^-], lysinate [Lys^-], and salicylate [Sal^-].

Initial preoptimization of the ion structures was carried out through the PM3 semiempirical (Stewart, 1989), the lowest energy structures were calculated with the B3LYP functional (Andzelm & Wimmer, 1992; Becke, 1993; Stephens *et al*., 1994) and the DGDZVP2 basis set (Godbout *et al*., 1992), and again the lowest energy structures were selected to be reoptimized with the PBE functional (Perdew *et al*., 1996), including dispersion (D3; Grimme *et al*., 2010), and the basis set 6-311++G** (Hehre *et al.*, 1986). Once the ions of the ILs were obtained, ILs were built, employing several spacial distributions between anions and cations with the aid of the electrostatic potential surfaces of the corresponding ions.

The lowest energy structures of the ILs were thus employed to obtain the complexes with the CO_2 molecule, IL-CO_2, and the gas-phase lowest energy structures were calculated including the solvent effect of water through the COSMO implicit solvation model (Klamt & Schüürmann, 1993). Calculations for the IL formation were carried out through the program package Gaussian 09 (Frisch *et al*., 2016); whereas the complex formation calculations and addition of the solvent effect were performed using NWChem 6.6 (Valiev *et al*., 2009). Additionally, frequency calculations were included to ensure the minima on the potential energy surface.

The thermochemistry of the ILs and the IL-CO_2 complexes formation was obtained according to Eqs. (1, 2). For IL, the reactants were the corresponding anion and cation, and the product was the IL under study; in the case of the complexes, the reactants were the IL and the CO_2, and the product was the IL-CO_2 complex. Basis set superposition errors (BSSE) were included through the Counterpoise method (Boys & Bernardi, 1970).

$$Cation + Anion \rightarrow IL \tag{1}$$

$$IL + CO_2 \rightarrow IL - CO_2 \tag{2}$$

In addition, inter- and intramolecular interactions present in the ILs and the IL-CO_2 complexes were analyzed through the Atoms in Molecules (AIM; Bader, 1991) and Non-Covalent Interactions (NCI; Johnson *et al*., 2010) methodologies as implemented in GPUAM (Hernández-Esparza *et al*., 2014; Hernández-Esparza *et al*., 2019; Cruz *et al*., 2019).

3. Results and Discussion

3.1. Ionic liquids

Table 1 shows the thermochemistry for the formation of the choline based ILs under study. According to the results, the IL formation is a slightly exothermic process, as shown by the enthalpy; whereas the Gibbs energy presents a non-spontaneous or endergonic process, through positive values, except for the complex with leucinate as anion, which is slightly negative. Even though the formation process in not spontaneous at standard conditions, it may be achieved in the laboratory by changing the conditions of the experiment.

Table 1. Thermochemistry for the formation of the choline based ILs under study at standard conditions.

IL	Δ*H* [kJ/mol]	Δ*S* [J/mol*K]	Δ*G* [kJ/mol]
[Ch][Ala]	-35.80	-140.13	5.99
[Ch][But]	-38.81	-163.45	9.92
[Ch][EtCO2]	-43.25	-166.51	6.36
[Ch][Gly]	-43.12	-169.02	7.29
[Ch][Leu]	-56.94	-189.79	-0.33
[Ch][Lys]	-48.99	-199.92	10.63
[Ch][Sal]	-40.91	-189.87	15.74

An example from the AIM and NCI results is shown in Fig. 1, where [Ch][$EtCO_2$] presents interactions between the hydroxyl group of the choline cation and the carboxylate of the anion (pink lines); in this case, NCI results show a strong hydrogen bond through the red isosurface. In addition to the hydrogen bonds in the IL structure, van der Waals interactions (green isosurfaces) were also found, which is key to the stability.

In a similar way, intramolecular interactions between the cation and the anion are also depicted. Similar interactions are found in all the studied ILs, where O–H···O and C–H···O hydrogen bonds are formed between the cation, as hydrogen bond donor, and the anion, as hydrogen bond acceptor. A slightly different behavior is present for ILs containing [Ala^-], [Leu^-], and [Lys^-], since the carboxylate group is near the amine group; in a similar way, [Sal^-] presents an intramolecular O–H···O hydrogen bond with the carboxylate group.

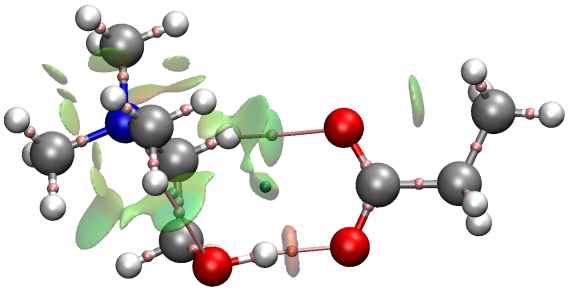


Figure 1. AIM and NCI for the [Ch][$EtCO_2$].

3.2. Ionic liquids and CO_2 complexes

Table 2 shows the thermochemistry for the formation of the IL-CO_2 complexes under study. The enthalpy for the formation process of the IL-CO_2 complex is less exothermic, in comparison with the formation of the sole IL. In a similar trend, the Gibbs energy is more endothermic in comparison with the IL, indicating a non-spontaneous process, which may be overcome through experiments at non-standard conditions. These values also indicate that IL-CO_2 complexes are more stable structures in comparison with the corresponding IL.

An interesting characteristic can be appreciated in the complex formation process, which in all cases is close to the thermodynamic equilibrium. This is very important, since this would allow the recovery of the IL; in this way, the ILs under study could be reused in several cycles of the absorption process.

Table 2. Thermochemistry for the formation of the IL-CO_2 complexes under study.

IL-CO_2 Complex	**ΔH [kJ/mol]**	**ΔS [J/mol*K]**	**ΔG [kJ/mol]**
[Ch][Ala]-CO_2	-20.10	-194.94	38.02
[Ch][But]-CO_2	-6.87	-107.64	25.25
[Ch][EtCO2]-CO_2	-8.83	-105.67	22.69
[Ch][Gly]-CO_2	-9.34	-65.65	10.26
[Ch][Leu]-CO_2	-8.92	-91.19	18.25
[Ch][Lys]-CO_2	-1.72	-104.67	29.52
[Ch][Sal]-CO_2	-4.02	-91.69	23.32

Fig. 2 shows an example related to Fig. 1, i.e., the AIM and NCI results for the [Ch][$EtCO_2$]-CO_2 complex. In this case, interactions between CO_2 and the ions of the IL are present; this can be observed through the NCI green isosurfaces, which show van der Waals interactions, as well as through the pink lines of the AIM results. In addition, possible hydrogen bonds were also found in the structure, especially those between the cation and the anion, where the smallest distance is 1.62 Å (see red isosurface). Similar results were found for the rest of the ILs.

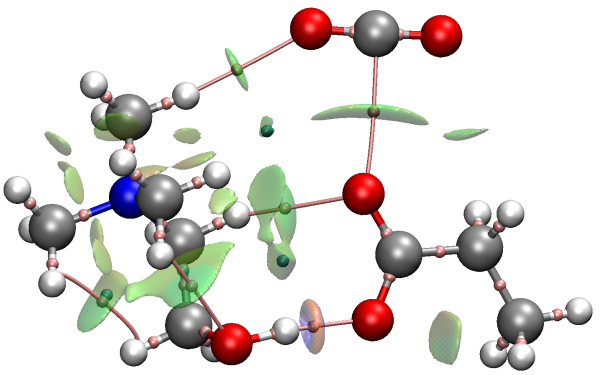


Figure 2. AIM and NCI for the [Ch][$EtCO_2$]-CO_2 complex.

4. Conclusions

In this work, the thermochemistry for the IL and IL-CO_2 complex formations were studied. Results show that hydrogen bonds between the carboxylates from the anions and cholinium cation are the main structural characteristic which provides stability to the ILs. In a similar way, IL-CO_2 complexes also present weak interactions between the ions and the CO_2, this consideration includes hydrogen bonds, where these interactions are important for the CO_2 absorption process.

With respect to the thermochemistry, energetics show that the IL-CO_2 complex formation is close to the thermodynamic equilibrium. This characteristic is very important since it is related to the possibility of the IL to be easily recovered and reused.

According to the results, all the studied ILs are good for the CO_2 absorption process, however, there is a slight preference for [Ch][Leu], since it presents an exergonic value for the Gibbs energy. The rest of the ILs are energetically close, therefore other considerations must be taken into account, such as cost and availability. In this way, CO_2 may be absorbed and further employed in other processes, such as biomass production.

Acknowledgments

MHM acknowledges Conacyt for grant CB-2016-284220. DRM acknowledges Conacyt for the scholarship No. 284220-28391-4307 through grant CB-2016-284220 and for the Master's degree No. 775129. Part of this work was carried out in the Yoltla Supercomputer at the *Laboratorio Nacional de Cómputo de Alto Desempeño* (LANCAD), Universidad Autónoma Metropolitana-Iztapalapa.

References

P. Anastas & N. Eghbali, 2010, Green Chemistry: Principles and Practice, Chem. Soc. Rev., 39, 1, 301-312.

R. F. W. Bader, 1991, A Quantum Theory of Molecular Structure and its Applications, Chem. Rev., 91, 893–928.

S. F. Boys & F. Bernardi, 1970, The Calculation of Small Molecular Interactions by the Differences of Separate Total Energies. Some Procedures with Reduced Errors, Mol. Phys., 19, 4, 553–566.

A. D. Becke, 1993, A New Mixing of Hartree–Fock and Local Density-Functional Theories, J. Chem. Phys., 98, 2, 1372–1377.

J. C. Cruz; R. Hernández-Esparza; A. Vázquez-Mayagoitia; R. Vargas; J. Garza, 2019, Implementation of the Molecular Electrostatic Potential Over Graphics Processing Units, J. Chem. Inf. Model, 59, 7, 3120–3127.

N. Godbout; D. R Salahub; J. Andzelm; E. Wimmer, 1992, Optimization of Gaussian-Type Basis Sets for Local Spin Density Functional Calculations. Part I. Boron Through Neon, Optimization Technique and Validation, Can. J. Chem., 70, 2, 560–571.

S. Grimme; J. Antony; S. Ehrlich; H. Krieg, 2010, A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu, J. Chem. Phys., 132, 15, 154104.

W. J. Hehre; L. Random; P. v. R. Schleyer; J. A. Pople, 1986, Ab Initio Molecular Orbital Theory, Wiley.

R. Hernández-Esparza; S. M. Mejía-Chica; A. D. Zapata-Escobar; A. Guevara-García; A. Martínez-Melchor; J. M. Hernández-Pérez; R. Vargas; J. Garza, 2014, Grid-based Algorithm to Search

Critical Points, in the Electron Density, Accelerated by Graphics Processing Units, J. Comput. Chem., 35, 31, 2272–2278.

R. Hernández-Esparza; A. Vázquez-Mayagoitia; L. A. Soriano-Agueda; R. Vargas; J. Garza, 2019, GPUs as Boosters to Analyze Scalar and Vector Fields in Quantum Chemistry, Int. J. Quantum Chem., 119, 2, e25671.

E. R. Johnson; S. Keinan; P. Mori-Sánchez; J. Contreras-García; A. J. Cohen; W. Yang, 2010, Revealing Noncovalent Interactions, J. Am. Chem. Soc., 132, 6498–6506.

S. Kasahara; E. Kamio; A. R. Shaikh; T. Matsuki; H. Matsuyama, 2016, Effect of the Amino-Group Densities of Functionalized Ionic Liquids on the Facilitated Transport Properties for CO_2 Separation, J. Membr. Sci., 503, 148–157.

A. Klamt; G. J. Schüürmann, 1993, COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and its Gradient, J. Chem. Soc., Perkin Trans., 2, 5, 799–805.

W. F. Lamb; T. Wiedmann; J. Pongratz; R. Andrew; M. Crippa; J. G. J. Olivier; D. Wiedenhofer; G. Mattioli; A. A. Khourdajie; J. House, 2021, An Review of Trends and Drivers of Greenhouse Gas Emissions by Sector from 1990 to 2018, Environ. Res. Lett., 16, 7, 073005.

S. A. Mazari; B. S. Ali; B. M. Jan; I. M. Saeed; S. Nizamuddin, 2015, An Overview of Solvent Management and Emissions of Amine-based CO_2 Capture Technology, Int. J. Greenh. Gas Control, 34, 129-140.

N. Noorani & A. Mehrdad, 2020, CO_2 Solubility in some Amino Acid-Based Ionic Liquids: Measurement, Correlations and DFT Studies, Fluid Phase Equilibr., 517, 112591.

R. G. Parr & W. Yang, 1989, Density Functional Theory of Atoms and Molecules, Oxford University Press, 989.

J. P. Perdew; K. Burke; M. Ernzerhof, 1996, Generalized Gradient Approximation Made Simple, Phys. Rev. Lett., 77, 18, 3865.

N. V. Plechkova, & K. R. Seddon, 2008, Applications of Ionic Liquids in the Chemical Industry, Chem. Soc. Rev., 37, 1, 123–150.

M. Rakkar & K. Hillier, 2007, “Choline” in xPharm: The Comprehensive Pharmacology Reference, S. J. Enna & D. B. Bylunds, Eds., Elsevier, 1–4.

A. R. Shaikh; M. Ashraf; T. AlMayef; M. Chawla; A. Poater; L. Cavallo, 2020, Amino Acid Ionic Liquids as Potential Candidates for CO_2 Capture: Combined Density Functional Theory and Molecular Dynamics Simulations, Chem. Phys. Lett., 745, 137239.

P. J. Stephens; F. J. Devlin; C. F. N. Chabalowski; M. J. Frisch, 1994, Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields, J. Phys. Chem., 98, 45, 11623–11627.

J. J. Stewart, 1989, Optimization of Parameters for Semiempirical Methods I. Method, J. Comput. Chem., 10, 2, 209–220.

M. Valiev; E. J. Bylaska; N. Govind; K. Kowalski.; T. P. Straatsma; H. J. Van Dam; D. Wang; J. Nieplocha; E. Apra; T. L. Windus; W. A. De Jong, 2009, NWChem: A Comprehensive and Scalable Open-Source Solution for Large Scale Molecular Simulations, Comput. Phys. Commun., 181, 9, 1477–1489.

H. Weingärtner, 2008, Understanding Ionic Liquids at the Molecular Level: Facts, Problems, and Controversies, Angew. Chem. Int. Ed., 47, 4, 654-670.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50162-7

Transport of CO_2/CH_4 through PEBA membranes: experiments and mass transfer modelling

Alejandro Solis-Jácome[a], Victor M. Rivera[a], Griselda Castruita de León[b], Miguel A. Morales-Cabrera[a*]

[a]*Facultad de Ciencias Químicas, Universidad Veracruzana, Xalapa, Veracruz, C.P. 91000, México.*
[b]*Centro de Investigación en Química Aplicada, Saltillo, Coahuila, C.P. 25294, México*
migmorales@uv.mx

Abstract

This work presents a study of the CO_2 transport process in Block Polyether Amide (PEBA) membranes, based on an experimental design and mathematical modelling, considering a fickian and non-fickian diffusion models. PEBA membranes were synthesized using the plate casting method and were thermally characterized by differential scanning calorimetry and thermogravimetric analysis. Subsequently, experimental measurements of the concentration of permeated gases were carried out, using different pressure and temperature conditions for a fixed CO_2/CH_4 mixture. The permeability, selectivity, diffusivity, and solubility of CO_2 were determined. It was found that the membranes are highly selective to CO_2 over the CH_4. The experimental data of the permeated CO_2 concentration throughout the operating time, the solubility and diffusivity were employed in the mathematical models to describe the solution-diffusion phenomenon. It was found that the fickian model was only able to reproduce the behaviour of the process at the steady state conditions. In contrast, the non-fickian model achieved an appropriate description of the process.

Keywords: CO_2 separation, dense membranes, solution-diffusion mechanism, mathematical modelling.

1. Introduction

Gas separation is one of the essential processes in the field of Chemical Engineering; specifically, the separation of acid gases (CO_2, H_2S, and SO_2) presents a considerable environmental and industrial importance, such as greenhouse gases capture and natural gas cleaning (Xu et al., 2019; Dai et al., 2016). For example, CO_2 accounts for approximately 60% of the total greenhouse gas emissions (Tan et al., 2019; Wang et al., 2019).

Natural gas, when burned, has a CO_2 emission factor of 26% and 41% lower than that emitted by oil and coal, respectively. Therefore, expanding natural gas consumption will be one of the effective measures to reduce CO_2 emissions (Li et al., 2018). However, nowadays, natural gas is extracted in large quantities and often contains an excess of CO_2, which must be removed to improve its heat power, reduce the volume of gas to be transported, avoid atmospheric pollution, and reduce corrosion in pipelines.

The most common methods of CO_2 separation are cryogenic distillation, adsorption, and absorption (Li et al., 2018; Kidnay et al., 2011). Among these methods, amine absorption is the most advanced commercial technology; however, this method implies

an increase of 50-90% in the energy cost of the plant (Ibrahim et al., 2018; Ji et al., 2010), as well as the corrosion of the equipment, and these are just some of the disadvantages of this technology (Ibrahim et al., 2018). A promising method to reduce the cost of CO_2 separation is the use of membranes, which, in comparison to chemical absorption, has advantages such as low cost and high energy efficiency (Ji et al., 2019; Li et al., 2016; Mahmoudi et al., 2015).

In the field of membrane technology, polymeric membranes have been widely used to capture CO_2 from different gas mixtures due to their selectivity and permeability, as well as their chemical and thermal resistance, low material costs, among others (Xu et al., 2019; Li et al., 2018; Dai et al., 2016). However, despite all the technological advances in the study of the CO_2 separation process using polymeric membranes, this is still an open field for research.

Another way to analyse the CO_2 separation process using membranes is through mathematical models and numerical simulations, validated with experimental data. However, most of the mathematical models reported in the literature for dense membranes use Fick's Law to describe the mass transport process. It may not be appropriate, since dense membranes are not a homogeneous medium and consequently predict the process far from reality. Therefore, this work proposes an analysis of the CO_2 transport in a dense membrane (PEBA) based on the determination of experimental data and a mathematical modelling that considers an effective diffusivity coefficient.

2. Methodology

Through sequential steps based on the elaboration and characterization of the membranes, properties such as permeability and selectivity of CO_2 were calculated using a gas stream consisting of methane including the variation of pressure and temperature.

The permeability of CO_2 and CH_4 was calculated through the following correlation.

$$\frac{\mathcal{P}}{\mathrm{L}} = \frac{\mathrm{QN}}{A(\mathrm{p_{in}} - \mathrm{p})} \tag{1}$$

Where $\mathcal{P}$ is the permeability, Q is the volumetric flow rate of the permeate gas, A is the active area of the membrane, $\mathrm{p_{in}}$ and p are the inlet and permeate partial pressures, respectively, and N is the permeate gas concentration.

For selectivity (α), it was calculated following the next expression.

$$\alpha_{\mathrm{CO_2/CH_4}} = \frac{\mathcal{P}_{\mathrm{CO_2}}}{\mathcal{P}_{\mathrm{CH_4}}} \tag{2}$$

It was also possible to determine certain parameters, such as molecular diffusion (D) and solubility (S).

$$\mathrm{D_{CO_2}} = \frac{\mathrm{L}^2}{6\theta} \tag{3}$$

$$\mathrm{S_{CO_2}} = \frac{\mathcal{P}_{\mathrm{CO_2}}}{\mathrm{D_{CO_2}}} \tag{4}$$

$\mathrm{D_{CO_2}}$ was calculated from the ratio obtained from the thickness (L) of the membrane and six times the period (θ) in which the concentration of the permeated gas reaches the steady state. Then, the $\mathrm{S_{CO_2}}$ coefficient was calculated from the relationship of the permeability of CO_2 and its diffusion coefficient through the membrane.

3. Mathematical Model

The solution-diffusion mechanism has been considered to study mass transfer in dense membranes, with diffusion inside the polymeric matrix and solubility at the limits of the membrane. The proposed mathematical model of this work focuses on determining the behaviour of CO_2 in the membrane, based on Fick's second law of diffusion, and considering an effective diffusivity.

$$\frac{\partial C_{CO_2}}{\partial t} = D_{eff} \frac{\partial^2 C_{CO_2}}{\partial x^2} \tag{5}$$

In Eq. (5), the coefficient D_{eff} is defined as a function of the molecular diffusion of CO_2 and a parameter ϕ, $D_{eff} = D_{CO_2}\phi$. A parametric estimation employing the Levenberg-Marquardt method was used to determine the parameter ϕ.

$$\mathbb{P}_{k,i+1} = \mathbb{P}_{k,i} + (J^T J + \lambda I)^{-1} J^T \left(y_i - y^k_{m,i}\right) \tag{6}$$

Where: $\mathbb{P}$ represents the parameter to be estimated; k and i are counters; J and J^T are the Jacobian and transposed Jacobian matrix, respectively; λ, the damping factor; I, an identity matrix; y_i, the real data; and y_m, the data based on the modelling. The CO_2 concentration data measured in the lower chamber of the membrane were used as real data in this work.

4. Results

4.1. Characterization of the membrane

Dense membranes with a uniform, flexible, and homogeneous appearance were synthesized from PEBA and characterized by Differential Scanning Calorimetry (DSC) and Thermogravimetric Analysis (TGA).

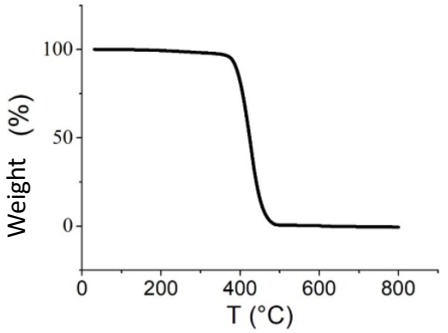


Figure 1. TGA analysis to membrane synthesized from PEBA.

The TGA was carried out to analyse the possible thermal decomposition of the PEBA membrane, as shown in Figure 1. A thermal decomposition for PEBA starts at 380 °C, where membrane degradation begins; this temperature is consistent with other studies reported in the literature (Soloukipour et al., 2017; Li et al., 2016).

4.2. Membrane permeability and selectivity

The permeabilities of CO_2 and CH_4 were calculated as a function of the pressure and the temperature (Figure 2). The CO_2 permeability values were in a range of 118.59 to 196.76 Barrer. These values are within the scope of 40 to 400 Barrer, reported in the

literature (Zhang et al., 2019; Zou et al., 2006). Regarding the permeability for CH_4, the values obtained were all smaller than 1 Barrer.

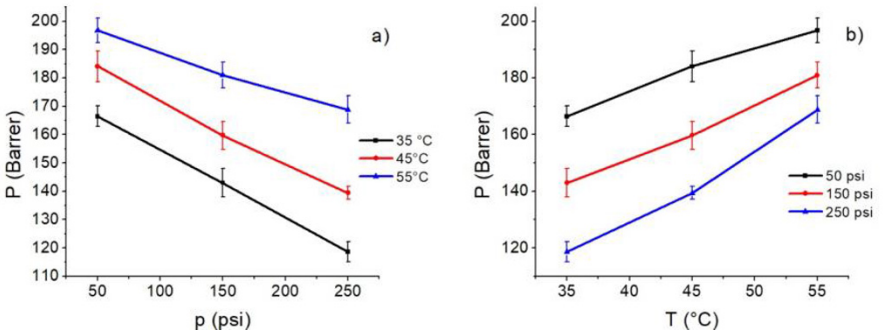


Figure 2. Effect of the pressure and the temperature on the permeability of PEBA membranes for CO_2 absorption. a) Effect of pressure. b) Effect of temperature.

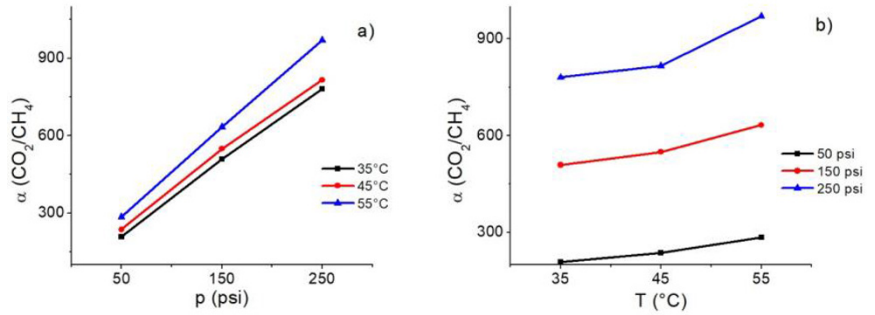


Figure 3. Effect of the pressure and the temperature on the selectivity of PEBA membranes for CO_2 absorption. a) Effect of pressure. b) Effect of temperature.

Figure 3 shows the values obtained for selectivity in a range of 207 to 970. Thus, these values are within 40 to 1200, as reported in the literature (Zhang et al., 2019; Kobayashi et al., 2019).

The calculated diffusion coefficients (D), with the effect of pressure and temperature, presented values between 2.5 to 6.5 x 10^{-6} cm^2/s. In the case of the solubility, values were found in a range of 2 to 6 cm^3 (STP) cm^{-3} cm Hg^{-1}, which are in good agreement with those reported in the literature (Kalantari et al., 2019; Hou et al., 2018).

4.3. Mathematical modelling

Prediction of the measured experimental CO_2 concentrations was carried out with both models, the classical fickian model and the non-fickian model proposed in this work. In addition, the parametric adjustment used molecular diffusivity and solubility.

Simulations were carried out with all operating conditions, as experimentally evaluated. For illustrative purposes, Figure 4 only shows the data corresponding to 35 °C and 50 psi. However, in all cases, the non-fickian model was the one with the best performance.

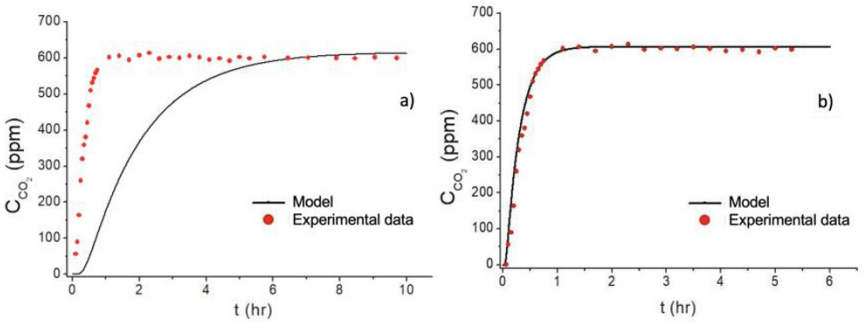


Figure 4. Dynamic behaviour of the CO_2 concentration in the membrane's lower chamber. a) Fickian model. b) Non-fickian model.

The non-fickian model required the adjustment of only one parameter, as included in the definition of the effective diffusivity coefficient ($D_{eff} = D_{CO_2}\phi$). The values of the parameter ϕ, determined by the Levenberg-Marquardt method, and the plasticization factor, reported by Minelli (2013) for several dense membranes, presented the same order of magnitude.

5. Conclusions

Experimental tests determined that the membranes under study are highly permeable and selective to CO_2, in comparison to CH_4, for all the evaluated pressures and temperatures. The permeation rate of CO_2 increased with the temperature and decreased with the pressure. The highest permeability value was 196.76 Barrer at a temperature of 55 °C and a pressure of 50 psi. In addition, pressure and temperature increments favoured the CO_2/CH_4 selectivity.

The mathematical modelling showed that, to describe the CO_2 transport process in dense membranes, a non-fickian model is required, where additional factors are included for the effective diffusion coefficient. It is important to remark that these factors must be a function of the operating conditions and the properties of the dense membranes.

References

Dai, Z., Noble, R. D., Gin, D. L., Zhang, X., & Deng, L., 2016, Combination of ionic liquids with membrane technology: A new approach for CO_2 separation. Journal of Membrane Science, 497, 1-20.

Hou, J., Li, X., Guo, R., Zhang, J., & Wang, Z., 2018, Mixed matrix membranes with fast and selective transport pathways for efficient CO_2 separation. Nanotechnology, 29(12), 125706.

Ibrahim, M. H., El-Naas, M. H., Zhang, Z., & Van der Bruggen, B., 2018, CO_2 capture using hollow fiber membranes: A review of membrane wetting. Energy & fuels, 32(2), 963-978.

Ji, Y., Zhang, M., Guan, K., Zhao, J., Liu, G., & Jin, W., 2019, High-performance CO_2 capture through polymer-based ultrathin membranes. Advanced Functional Materials, 29(33), 1900735.

Kalantari, S., Omidkhah, M., Amooghin, A. E., & Matsuura, T., 2019, Superior interfacial design in ternary mixed matrix membranes to enhance the CO_2 separation performance. Applied Materials Today, 100491.

Kidnay, A.J., Parrish, W.R., McCartney, D.G., 2011, Fundamentals of Natural Gas Processing. Taylor & Francis Group, Boca Raton.

Kobayashi, Y., Nakamitsu, Y., Zheng, Y., Takashima, Y., Yamaguchi, H., & Harada, A., 2019, Preparation of cyclodextrin-based porous polymeric membrane by bulk polymerization of ethyl acrylate in the presence of cyclodextrin. Polymer, 177, 208- 213.

Li, L., Wong-Ng, W., Huang, K., & Cook, L.P., 2018, Materials and processes for CO_2 capture, conversion, and sequestration. John Wiley & Sons.

Li, Y., Shen, J., Guan, K., Liu, G., Zhou, H., & Jin, W., 2016, PEBA/ceramic hollow fiber composite membrane for high-efficiency recovery of bio-butanol via pervaporation. Journal of Membrane Science, 510, 338-347.

Mahmoudi, A., Asghari, M., & Zargar, V., 2015, CO_2/CH_4 separation through a novel commercializable three-phase PEBA/PEG/NaX nanocomposite membrane. Journal of Industrial and Engineering Chemistry, 23, 238-242.

Minelli, M., & Sarti, G. C., 2013, Permeability and diffusivity of CO2 in glassy polymers with and without plasticization. Journal of membrane science, 435, 176-185.

Soloukipour, S., Saljoughi, E., Mousavi, S. M., & Pourafshari Chenar, M., 2017, PEBA/PVDF blend pervaporation membranes: preparation and performance. Polymers for Advanced Technologies, 28(1), 113-123.

Tan, X., Li, H., Guo, J., Gu, B., & Zeng, Y., 2019, Energy-saving and emission-reduction technology selection and CO_2 emission reduction potential of China's iron and steel industry under energy substitution policy. Journal of cleaner production, 222, 823-834.

Wang, S., Wang, J., Fang, C., & Li, S., 2019, Estimating the impacts of urban form on CO_2 emission efficiency in the Pearl River Delta, China. Cities, 85, 117-129.

Xu, B., & Lin, B., 2019, Can expanding natural gas consumption reduce China's CO_2 emissions? Energy Economics, 81, 393-407.

Zhang, J., Xin, Q., Li, X., Yun, M., Xu, R., Wang, S., & Zhang, Y., 2019, Mixed matrix membranes comprising aminosilane-functionalized graphene oxide for enhanced CO_2 separation. Journal of membrane science, 570, 343-354.

Zou, J., & Ho, W. W., 2006, CO_2-selective polymeric membranes containing amines in crosslinked poly (vinyl alcohol). Journal of Membrane Science, 286(1-2), 310-321.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50163-9

Mathematical modelling, simulation and optimisation of an indirect water bath heater at the Takoradi distribution station (TDS)

Richard Yentumi[a], Bogdan Dorneanu[b], Harvey Arellano-Garcia[b,*]

[a]*Engineering Department, Ghana National Gas Company, Atuabo, Ghana*
[b]*LS Prozess- und Anlagentechnik, Brandenburgische Technische Universität Cottbus-Senftenberg, Cottbus, D-03046, Germany*
arellano@b-tu.de

Abstract

In this contribution, a dynamic first principles model of an existing 3.01 MW natural gas fired water bath heater (WBH) in operation at the Takoradi Distribution Station (TDS) in Ghana is developed primarily to predict the outlet temperature of the natural gas stream being heated. The model is intended to be applied during operations to provide useful data to optimise material and energy consumption, as well as minimise CO_2 emissions. Due to the low thermal efficiencies of WBHs, even small improvements in efficiency can result in significant savings. The firetube and process coils are both modelled as one-dimensional (1D) thin-walled tubes and the entire model incorporates mass and energy conservation equations, heat transfer rate relations and rigorous thermodynamic p-V-T relations. In contrast to what commonly exists in literature, this model accurately estimates the enthalpy change of the natural gas stream being heated by accounting for its enthalpy departure correction term due to pressure, in addition to the ideal gas heat capacity relation which is a function of only temperature. The coupled ordinary differential and algebraic equations are implemented using gPROMS® ModelBuilder® V4.2.0, a commercial modelling and simulation software. Verification of the model results showed good agreement between the model predictions and actual on field measurements. With excess air at 15%, the simulation results closely approximate measured data with an absolute error of about 0.31 %. More importantly, the results show that significant savings of up to 30% per annum can be made through optimal operation of the water bath heater.

Keywords: modelling, optimisation, simulation, heater.

1. Introduction

In natural gas transportation systems, pressures must often be reduced from transmission pipeline to distribution pipeline values at regulating and metering (R&M) stations. This pressure reduction process with the accompanying temperature drops, a phenomenon known as the Joule-Thomson effect, is an isenthalpic expansion process. Consequently, if the natural gas chills to sufficiently low temperatures, owing to a huge pressure reduction across the throttling valve, for instance, some of the heavier components of the natural gas stream can condense out of the vapour phase into a liquid phase potentially forming hydrates in the presence of free water and causing corrosion problems. Hydrates can damage pipeline components or cause blockages that disrupt gas export operations. Therefore, to protect downstream consumers such as natural gas-fired turbine power

plants, indirect water bath heaters (WBH) are typically provided upstream of pressure regulating packages in R&M stations to superheat the natural gas stream. Normally, the natural gas is in a superheated state if it is at a temperature of at least 28 ^{0}C above the hydrocarbon dewpoint temperature at the specified operating pressure. Figure 1 shows a simplified schematic of an indirect WBH.

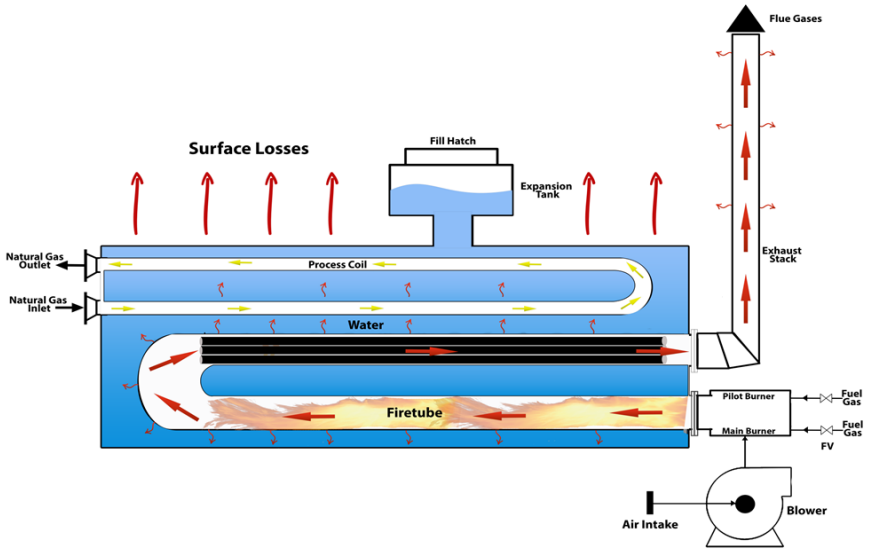


Figure 1: Indirect water bath heater at the TDS

As most process heaters, indirect WBHs are energy intensive units. However, they have relatively lower overall thermal efficiencies when compared to direct fired heaters, for instance. As a percentage of the total energy input (LHV basis), a large majority of previous research estimated that the stack losses were somewhere between 25 – 60% while the wall losses range from 0.5 to 2% of the total energy input. Stack losses were therefore, identified as a major source of energy wastage accounting for the low thermal efficiencies of indirect water bath heaters. Through a 12-month statistical data collection and analysis, Khalili et al. (2010) estimated the thermal efficiency of an indirect water bath heater to be between 41.6 and 52.0%. Similarly, Romocki et al. (2018), reported the thermal efficiency of a WBH to be around 46%. Clearly, such low thermal efficiency levels of the existing water bath heaters are incongruous with the looming transition into a low-carbon future which stipulates drastic reductions in carbon emissions and high energy efficiency. In this light, mathematical modelling and simulation can be a very powerful tool that can be used to gain deeper insight into the underlying phenomena involved in WBH design and operations and in effect, help to improve operating efficiencies. Accordingly, Azizi et al. (2014), proposed an energy integration approach by placing a heat exchanger upstream of the indirect water to extract heat energy from the flue gases released from the water bath heater, which hitherto, would have been released into the atmosphere to pre-heat the natural gas stream and thus, reduce the total energy input leading to a lower overall fuel consumption. Romocki et al. (2018), proposed the replacement of indirect WBH with an Immersion Tube Thermosyphon Heater (ITTH) as it offered a thermal efficiency of about 90% leading to a reduction in both fuel consumption and CO_2 emissions.

Interestingly, only a handful of these previous works were conducted to model and simulate the dynamic performance of these process heaters such as cold start-up operations. Moreover, one of the main weaknesses of the existing models was the estimation of enthalpy changes of the natural gas stream being heated based on the ideal gas heat capacity relation, which is a function of only temperature and does not account for enthalpy departure correction term due to pressure of the natural gas. The overall objective of this work is to achieve optimal operation of the TDS WBH by identifying optimal operating variables that minimise fuel gas use (increasing energy efficiency) for a given process heat duty while also minimising CO_2 emissions.

2. Model development

2.1. Process description

Three (3) indirect fired heaters are installed in a 3 x 50 % configuration (3 x 202.5 MMscf/D, design heating duty per heater is 3.01 MW) upstream of the pressure regulating package at the TDS. A cascaded temperature control loop is provided for controlling the temperature of the water bath. In addition, a cascaded temperature control loop is also provided for controlling the outlet temperature of the natural gas stream being heated. The water bath temperature is controlled and normally set at 60 ^oC (333.15 K). The design setpoint is 86 °C (359.15 K). Depending on the downstream customer requirement, the natural gas is typically heated to between 30 ^oC (303.15 K) and 52 ^oC (325.15 K); normal setpoint temperature is 40 ^oC (313.15 K). The complete model comprises firetube zones 1 and 2, the non-flow water bath, the process tube, and the natural gas stream. The various sub-models are then interconnected by energy flows shown in Figure 2.

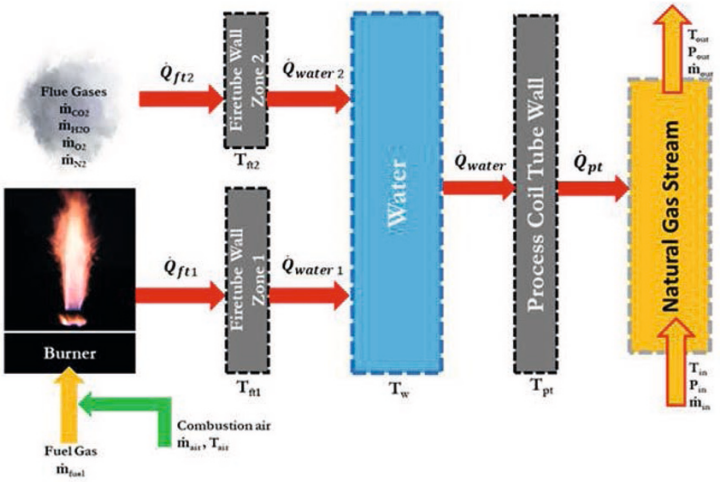


Figure 2: Model connectivity

The model considers the overall energy balance for the WBH control volume, as well as for the various sub-components, the combustion of a natural gas stream containing both nitrogen and carbon dioxide in the presence of excess air, the transfer of the heat to the first pass of the firetube through a combination of radiation and convection.

The outlet temperature of the natural gas stream, $T_{ng,2}$ is estimated from the relation:

$$q_{\dot{n},ng} \cdot d\bar{h} = \dot{Q}_{ng} \quad (1)$$

Changes in the specific molar enthalpy of the natural gas between two thermodynamic equilibrium states is estimated from the relation:

$$d\bar{h} = \bar{h}_2(T_{ng,2}, P_{ng,2}) - \bar{h}_1(T_{ng,1}, P_{ng,1}) \quad (2)$$

In general, the specific molar enthalpy at each state is calculated from the relation:

$$\bar{h}_i = \bar{h}_{ideal} + \bar{h}_{dep} \quad (3)$$

The first term on the RHS of Eqn. (3) represents the ideal gas enthalpy change for the individual components of the natural gas mixture at reference conditions of 298.15 K and 1 atm to the state temperature and at ideal gas conditions; the second term on the RHS represents the enthalpy departure (pressure correction term) involved in taking the fluid to the state pressure. The Lee-Kesler (LK) Equation of State (EOS) method (Lee and Kesler, 1975) was used to estimate this term as it is the most accurate enthalpy calculation model for gas mixtures (Changjun et al., 2011). The compressibility factor of the natural gas stream at each state is calculated using the Peng-Robinson EOS. (Peng and Robinson, 1976). Physical property data and binary interaction parameters were obtained from ASPEN® HYSYS® V11. Since the flue gas flow through the firetube is calculated to fall under forced convection, the Dittus-Boelter correlation as discussed in Incropera et al., (2007) is used for estimating the convection heat transfer coefficients in all cases. All external free convection heat transfer coefficients are estimated using the GPSA Engineering Data Book.

2.2. Optimisation model

The complete optimisation problem is posed in the form below:

Minimise:

$$f = n_{days} \cdot 24 \cdot c_{fuel} \cdot q_{\dot{m},fuel-MB} \cdot LHV_{fuel}$$

Subject to:

$$50\frac{kg}{hr} \leq q_{\dot{m},fuel-MB} \leq 620\frac{kg}{hr}$$

$$333.15\ K \leq T_w$$

$$313.15\ K \leq T_{ng,2}$$

The objective function is the total cost of fuel gas burned, in $ per annum. The % valve opening, and ambient air temperature are set as decision variables in the gPROMS® optimisation settings. The water bath (T_w) and the hot natural gas outlet ($T_{ng,2}$) temperatures, which are external discontinuities, are set as constraints. The fuel gas flow rate ($q_{\dot{m},fuel-MB}$) flowing through the modulating flow control valve is set as an additional constraint.

2.3. Parameter specifications

Firetube Zone 1: $D_{i,ft1}$ = 0.89534 m; $D_{o,ft1}$ = 0.9144 m; $A_{o,ft1}$ = 24.0264 m^2; $A_{i,ft1}$ = 23.5256 m^2; m_{ft1} = 2739.50 m^2; cp_{steel} = 0.4610 kJ/kg K; k_{steel} = 0.0450 kW/m K;

Firetube Zone 2 - Single Firetube in Zone 2: $D_{i,ft2}$ = 0.06268 m; $D_{o,ft2}$ = 0.073 m; $A_{o,ft2}$ = 1.4147 m^2; $A_{i,ft2}$ = 1.2147 m^2; m_{ft2} = 2739.50 m^2; n_{tubes} = 52;

Firetube Zone 2 – Total Number of Firetubes in Zone 2: $A_{o,ft2}$ = 73.5621 m^2; m_{ft2} = 2764.00 m

2.4. Variable specifications

$\varepsilon = 1.15$; $\eta_{MB} = 0.95$; $q_{m,fuel}$ = Min (Valve_Flow, 620) kg/hr; $LHV_{ng} = 47{,}990.08$ kJ/kg; $\alpha_{ft1} = 0.90$; $e_f = 0.90$; $e_{ft1} = 0.50$; $T_{amb,air} = 298.15$ K; $h_{ws} = 0.120$ kW/m^2 K; $h_{al\text{-}air} = 0.00250$ kW/m^2 K; MW_ng = 18.6529 kg/kmol; w = 0.0259; $q_{std,v} = 85.23$ MMscf/D; $T_1 = 26.10$ ^oC ($T_{ng,1} = 299.25$ K); $P_1 = 68.00$ barg; $v_{m,1} = 0.8221$ m^3/kmol; $P_2 = 67.75$ barg; $v_{m,2} = 0.2964$ m^3/kmol;

The calculated Lee-Kesler pseudo-critical properties for the natural gas composition are: $T_c = -63.53$ ^oC and $v_c = 0.1063$ m^3/kmol; $F_1 = 0.05$ and $F_2 = 0.85$.

3. Results and discussion

Per the cold start-up operating procedure of the WBH, motor operated valves (MOVs) are opened to allow for natural gas flow and heating only once the setpoint of the water bath is reached. This discontinuity in the process was captured in the model execution in gPROMS®. A summary of the cold start-up simulation results for a water bath set point temperature of 60 ^oC (333.15 K) and a natural gas outlet temperature setpoint of 40 ^oC (313.15 K) is presented in Table 1. The initial cold conditions were 27.1 ^oC (300.25 K) for the water bath and 26.1 ^oC (299.25 K) for the cold inlet natural gas stream.

Table 1: Summary of TDS WBH Cold Start-up Simulation Results for 3,050 seconds

Valve % Travel	%	47	50	-6.00%
Time Taken to Reach Water Bath Setpoint Temperature	s	2,152	2,192	-1.82%
Outlet Temperature of Natural Gas Stream	K	314.28	313.30	0.31%

The temperature profiles of the water bath and the natural gas stream for the cold start-up simulation period are illustrated by Figures 3.

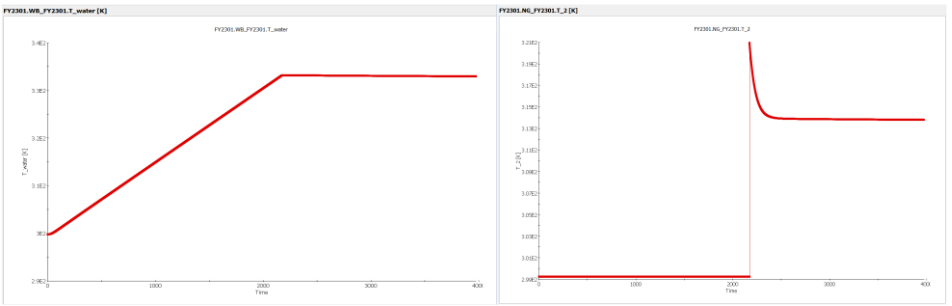


Figure 3: Temperature profiles of the water bath (left) and the natural gas (right)

The results obtained demonstrate that the TDS WBH model closely approximates the actual behaviour of the key operating variables such as water bath and outlet natural gas temperature. At the same time, the model deviates quite significantly in predicting other variables such as thermal efficiency which the manufacturer specifies as a minimum of 75%.

A model-based dynamic optimisation operation was conducted in gPROMS to determine the operating variables that minimise fuel gas purchasing cost. Considering a 24 h over 340 days operation in a year and a natural gas purchase cost of $0.004265/MJ, the results obtained indicate that the annualised purchasing cost of natural gas for the base case is $674,065 from cold start up to steady state, a total time of 2,296 seconds (38.26 minutes). The optimised case shows significant savings between 10 – 30% can be realised for a total time of 3,600 seconds (60 minutes) from cold start up to steady state.

The objective function was particularly sensitive to the % opening of the modulating flow control valve on the main burner fuel gas supply line.

4. Conclusions

The simulation and optimisation study results clearly demonstrate that the dynamic model proposed in this work is a powerful predictive tool for process improvement and to aid decision making at TDS as it closely approximates the behaviour of the actual water bath heater, FY2301A/B/C under varied operating conditions. The model, thus, affords an insight into the dynamics of the process that did not exist previously. Further work should be done to validate the model with experimental data and verify the optimisation results. Moreover, a robust process controller can be designed to accurately control and optimise the operations of the WBH.

References

E. Khalili, et al., 2010, Efficiency and heat losses of an indirect water bath heater installed in natural gas pressure reduction station; Iran, shahrekord, pp. 1-9

S. Romocki, et al., 2018, An indirect heating solution to reduce CO_2 emissions and improve efficiency of gas distribution networks, Energy Reports, Elsevier Ltd.

S. Azizi, et al., 2014, Study of preheating natural gas in gas pressure reduction stations by the flue gas of indirect water bath heaters, International Journal of Science and Engineering Investigations 3 (27)

B.I. Lee, et al., 1975, A generalized thermodynamic correlation based on three-paramter corresponding states, AIChE Journal 21 (3)

L. Changjun, et al., 2012, Application of Lee-Kessler equation of state to calculating compressibility factors of high-pressure condensate gas, 2nd International conference on advances in Energy Engineering, Energy Procedia 14, pp. 115-120

D. Peng, D.B. Robinson, 1976, A new two-constant equation of state, Ind. Eng. Chem., Fundam. 15 (1), pp. 59-64

F.P. Dewitt, et al., 2007, Introduction to heat transfer, 5th Edition, John Wiley & Sons

Gas Processors Suppliers Association (GPSA), 2012, Engineering Data Book, 13th Edition

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50164-0

Marine SOx Scrubber: Mass transfer Analysis, Design, Simulation and Experiment

Dongyoung Lee[a*], Van D. L. Nguyen[a,1], Gwangsik Kim[a,1], Myungjin Kim[b], Choongyong Kwag[b], Youngmok Lee[b], Sungwon Lee[b], Moonyong Lee[a*]

[a]*School of chemical Engineering, Yeungnam University, Gyeongsan 712-749, South Korea*
[b]*Hanbal Masstech Ltd, golden root complex, Gimhae, South Korea*
mynlee@ynu.ac.kr

Abstract

The International Marine Organization (IMO) established regulations on SO2 emissions in the MARPOL Annex VI, which is applied from 1st January 2020. The emissions standard was 3.5 % in weight, but now, sulphur emissions for oceangoing vessels must be equivalent to a sulphur content in fuel lower than 0.5 % in weight worldwide. Especially, in some coastal regions named “Sulphur Emission Control Areas”, SECAs sulphur emissions have to be lower than 0.1 % in weight. Due to economical compliance to the environment regulation, wet flue gas desulfurization (FGD) process is preferred. In this work, experiments are performed using scrubber, which has square-based shape and fresh water with addition of sodium hydroxide as absorbent, to treat flue gas from a marine diesel engine (720 kW) that has to comply with current regulation. The experimental tests in spray column indicate that the absorption efficiency above 95 % can be achieved. The experimented performances match well with the simulated results

Keywords: Process design; Marine flue gas desulfurization(FGD); The International Marine Orgarnization(IMO)

1. Introduction

In the worldwide, trade by sea consists more than 90 % of the total trade amount. In transport activities, large ships often use cheap heavy fuel oil, resulting in sulfur oxides (SOx)exhaust, which can negatively affect human health and marine communities.(Nielsen,2014) IMO and several governments have put into use rules to control the damages from sulfur emission. In particular, a maximum limit for equal sulfur emission of less than 0.1 % in weight was established for some coastal areas termed as sulfur emission control areas (January 1, 2015). In addition, from January 1, 2020, sulfur emission for oceangoing ships worldwide must be equal to those given-off by fuel with sulfur content lower than 0.5% in weight.(Flagiello, 2019) These rules have forced ship owners to search for solutions to meet the needed things, not only for new ships, but more importantly for the current fleet of ships.(Strandberg, 2017)

For controlling sulfur emissions effectively, FGD is a viable option, which the wet processes consists around 87 % of those used worldwide.(Srivastava, 2017) Shipments can apply open loop with seawater, closed loop with sodium hydroxide (NaOH), or hybrid systems which be able to switch the modes.(Oikawa, 2003) Seawater’s natural alkalinity and large availability make absorption ability. However, open loop system has several limitation such as it requires large amount of seawater which cause corrosion on process

equipment and several countries have banned any discharge from open loop system. (Eason, 2018) Thus, the closed-loop mode using NaOH is preffered alternative. In this paper, a new square-shaped closed-loop scrubber with spray has been proposed for marine applications. By applying the square shape, the spray scrubber can reduce its volume, area and weight, which are essentially considered for installation on board. We performed design, construction, experiment, and simulation to evaluate the proposed FGD square-shaped closed-loop scrubber with spray.

2. Closed-Loop Square Scrubber With Spray

2.1. Proposed Methodology

As the first step, a literature survey on the availability of components in Aspen Plus, solubility or equilibrium data, reaction parameters, requirements, and constraints of a marine FGD was performed. Subsequently, a rigorous simulation and a sensitivity analysis were performed to evaluate the scrubber and identify the variables affecting its performance with actual feed conditions. Finally, the simulation and experimental data were compared to validate the proposed systematic methodology.

2.2. Process Description

Existing FGD systems are coastal and commercial marine scrubbers that usually consist of spray towers fed with pure seawater (open loop) or water doped with an alkaline neutralizing agent (closed loop). NaOH is commonly used as an alkaline agent(MAN,2020) to neutralize the sulfuric acid in the scrubber water. When SO2 is absorbed into solutions, the diffusion of SO2 molecules from the current gas phase core to the gas/liquid interface and the dissolution in the washing agent should be initially considered, as described by Eq. (1).(Wang,2015) The dissolution process is considered to obey Henry's law for low concentrations of SO2 in effluent gases.

$$SO_2(g) \leftrightarrow SO_2(l) \tag{1}$$

Some dissolved acid gas molecules dissociate according to reaction (2), and others directly react with hydroxide ions according to reaction (3). In turn, hydrogen sulfite reacts with the additional hydroxide ions to form sulfite, as shown in reaction (4).

$$SO_2 + H_2O \leftrightarrow HSO_3^- + H^+ \tag{2}$$

$$SO_2 + OH^- \leftrightarrow HSO_3^- \tag{3}$$

$$HSO_3^- + OH^- \leftrightarrow SO_3^{2-} + H_2O \tag{4}$$

Reaction (3) is very fast while reaction (4) has a considerably higher rate constant than reaction (3). Thus, both reactions can be considered as instantaneous.(Hikita,1977) The above processes are related with the dissociation of water (Eq. 5) and NaOH (Eq. 6):

$$H_2O \leftrightarrow H^+ + OH^- \tag{5}$$

$$NaOH \leftrightarrow Na^+ + OH^- \tag{6}$$

$$H_2SO_3 + 2NaOH + 1/2O_2 \leftrightarrow Na_2SO_4 + 2H_2O \quad (7)$$

$$SO_2(g) + 2NaOH(aq) + 1/2O_2(g) \leftrightarrow 2Na^+ + SO_4^{2-} + H_2O \quad (8)$$

$$CO_2(g) \leftrightarrow CO_2(l) \quad (9)$$

$$CO_2 + H_2O \leftrightarrow HCO_3^- + H^+ \quad (10)$$

$$CO_2 + OH^- \leftrightarrow HCO_3^- \quad (11)$$

$$HCO_3^- + OH^- \leftrightarrow CO_3^{2-} + H_2O \quad (12)$$

Figure 1 shows a simplified closed-loop FGD process in which water is recirculated through the scrubber with no discharge at sea.

2.3. Design of Square Spray Column

Spray towers have traditionally been used in several applications such as low pressure drop is essential, a high degree of separation is not required, and there are solid particles and precipitating solvents.(Yeh,2003) They have advantages such as low pressure drop, lightness, simple construction and operation, and low investment, operating, and maintenance costs.(Bandyopadhyay,2012)

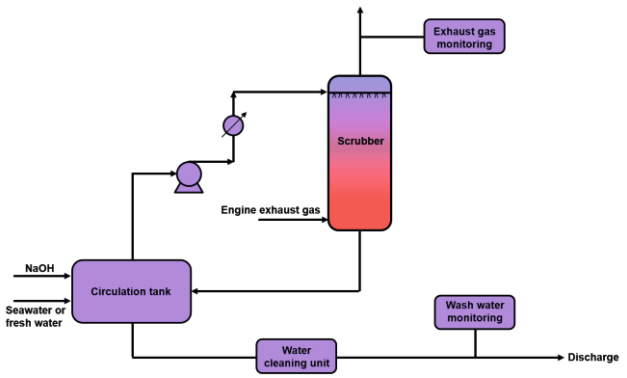


Figure 1 Schematic diagram of the closed-loop FGD process.

3. Materials And Methods

3.1. Materials

Fresh water with the addition of NaOH was used as an alkaline neutralizing agent. NaOH 5 % was purchased from Hanwha Solutions.

3.2. Diesel Engine

In the scrubber experiment, a Shinko marine diesel engine (Figure 2a) with a power of 720 kW was operated with high sulfur fuel oil containing 3.5 % in weight. Figure 2b shows the effect of the engine speed on the flow rate of flue gas. During the scrubber experiment, an engine speed of 800 rpm was fixed.

(a) (b)

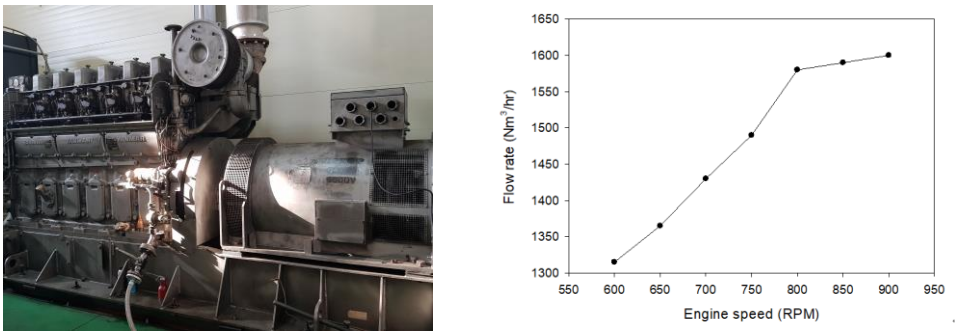


Figure 2 (a) Marine engine and (b) influence of engine speed on the flow rate of flue gas.

3.3. Spray Column

The square spray tower, with a cross section of 800 × 800 mm, length of 11,400 mm, and operated at 35 °C and 1 atm, was constructed after the design phase. The liquid flow rate, L, was supplied at the top by a centrifugal pump controlled by a globe valve. For spray scrubbing, a good atomizer should produce a fairly uniform spray with drop diameters sufficiently small to generate a large interfacial area of contact and sufficiently large to prevent excessive entrainment.(Bandyopadhyay,2012)

3.4. Analytical Methods

The Testo 350 exhaust gas analyzer equipped with an electrochemical sensor was used to measure SO2 based on the principle of ion selective potentiometry. The experimental SO2 removal efficiency (η_{SO_2}) was calculated as follows:

$$\eta_{SO_2} = \frac{C^{o}_{SO_2(g)} - C_{SO_2(g)}}{C^{o}_{SO_2(g)}} \tag{13}$$

where $C^{o}_{SO_2}(g)$ and $C_{SO_2}(g)$ are the inlet and outlet SO2 concentrations, respectively.

4. Results and Discussion

4.1. Simulation

The simulation was performed using the simulator Aspen Plus V10. The Electrolyte NRTL activity coefficient model was used for the prediction of the vapor–liquid equilibrium of these simulations. In this study, during the experiment and simulation, the pH value of the solvent was maintained at approximately 8.

Figure 3 shows the influence of the gas velocity, flue gas temperature, and liquid/gas ratio (L/G) on the removal efficiency. The results indicate that the removal efficiency reduces from approximately 97 % to 92 % when the flue gas velocity increases from 0.5 m/s to 2.5 m/s (Figure 3a). In addition, the influence of flue gas temperature within 30–90 °C is negligible (Figure 3b). Furthermore, the removal efficiency increases with increasing L/G (Figure 3c).

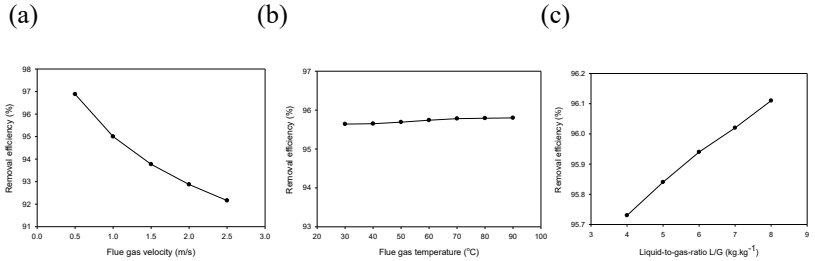


Figure 3 Influence of the (a) flue gas velocity, (b) flue gas temperature, and (c) L/G on the removal efficiency of SO2

4.2. Scrubber Operation

The flue gas from the engine was quenched in a venturi scrubber before entering a cyclone to remove particulate matter. 100 ppm SO2 was involved in feed gas. The industrial-scale scrubber unit constructed at the Hanbal Masstech company was used to scrub SO2 from the flue gas. When the column was wetted, the engine was activated to generate flue gas, which was fed into the scrubber unit. To maintain the operating temperature at 35 °C, fresh water was pumped to the air cooler system.

The analyses of the flue gas inlet and outlet were performed using the Testo 350 exhaust gas analyzer. The results showed that under the designed operating conditions, the outlet SO2 was 5 ppm, resulting in an outlet SO2(ppm)/CO2(%) ratio of 0.64. It was possible to achieve an absorption efficiency higher than 95% by using a liquid-to-gas mass ratio of approximately 4.32 kg.kg-1. The results also showed that the unit had small pressure drops (0.5 mbar).

4.3. Comparison Between Simulation And Experimental Results

The experimental outlet SO2 ppm was compared with the simulation results obtained using the Aspen Plus software, which were based on actual conditions. Excellent agreement was observed between the experimental and simulated performances. In particular, the real operation indicated an outlet SO2 of 5 ppm, which is similar to the simulated outlet SO2 of 4.3 ppm (Figure 4a).

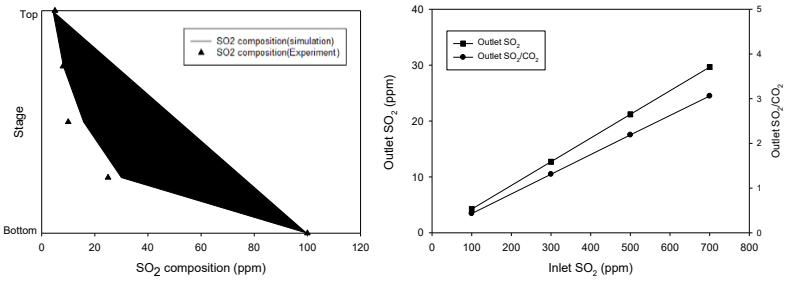


Figure 4 (a)SO2 profile in scrubber. (b)Influence of the inlet SO2 on the scrubber performance.

The results indicated that the higher the inlet SO2, the higher the outlet SO2 and the SO2(ppm)/CO2(%) ratio (Figure 4b). This implies that, in this range, the designed scrubber can comply with the IMO regulation.

5. Conclusions

A compact, light, and effective square scrubber equipped with spray was proposed for SOx removal in this study. Through the proposed systematic methodology, the design, construction, simulation, sensitivity analyses, and experiment were successfully accomplished. The results indicated a good agreement between experimental and simulation results. The proposed square scrubber system can achieve high SO2 removal efficiency with low water consumption and low pressure drop. In particular, an SO2 removal efficiency higher than 95 % was achieved using a liquid-to-gas mass ratio of approximately 4.32 kg.kg-1. Most part of mass and heat transfers occurred in the bottom section of the scrubber. In addition, the results indicated that the unit exhibited comparatively low pressure drop.

6. Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT)(2021R1A2C1092152) and by Priority Research Centers Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2014R1A6A1031189).

References

[1] R.F.Nielsen, F.Haglind, U.Larsen, 2014, Design and modeling of an advanced marine machinery system including waste heat recovery and removal of sulphur oxides, Energy Convers, 85, 687–693.

[2] D.Flagiello, A.Parisi, A.Lancia, C.Carotenuto, A.Erto, F.Di Natale, 2019, Seawater desulphurization scrubbing in spray and packed columns for a 4.35 MW marine diesel engine, Chem. Eng. Res. Des., 148, 56-67.

[3] P.Strandberg, 2017, Marine exhaust gas scrubber, US 9,776,125 B2.

[4] R.K.Srivastava, W.Jozewicz, C.Singer, 2001, SO2 scrubbing technologies: A review, Environ. Prog, 20, 219-228.

[5] K.Oikawa, C.Yongsiri, K.Takeda, T.Harimoto, 2003, Seawater flue gas desulfurization: its technical implications and performance results, Environ. Prog, 22, 67-73.

[6] C.Eason, 2018, Announcement: Scrubber maker response to Singapore ban on open loop scrubbers, https://fathom.world/announcement-scrubber-maker-response-to-ro-singapore-ban-on-open-loop-scrubbers/

[7] MAN, 2020, Emission project guide, MAN B&W Two-stroke marine engines, https://indico.cern.ch/event/659434/attachments/1528657/2391369/MAN_BW_Two-stroke_Marine_Engines.pdf

[8] Z.Wang, Y.Peng, X.Ren, S.Gui, G.Zhang, 2015, Absorption of sulfur dioxide with sodium hydroxide solution in spray columns, Ind. Eng. Chem. Res, 54, 8670−8677.

[9] H.Hikita, S.Asai, T.Tsuji, 1977, Absorption of sulfur dioxide into aqueous sodium hydroxide and sodium sulfite solutions, AIChE J, 23 (4), 538.

[10] N.K.Yeh, G.T.Rochelle, 2003, Liquid-phase mass transfer in spray contactors, AIChE J, 49, 2363-2373.

[11] A.Bandyopadhyay, M.N.Biswas, 2012, CO2 capture in a spray column using a critical flow atomizer. Sep. Purif. Tech, 94, 104-114.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50165-2

Connecting the Simulation Model to the Digital Twin to help drive Sustainability

Julien de Beer[a*], Mihaela Hahne[b]

[a]*AVEVA, 5 square Felix Nadar, 94300 Vincennes, France*
[b]*AVEVA, 26561 Rancho Parkway South, Lake Forest, CA 92630, USA*
[*]*julien.debeer@aveva.com*

Abstract

The chemicals industry is facing a highly dynamic environment. Demand is continually fluctuating, and the pressure for new sustainable processes and products is rising. New environmental regulations are always expanding and becoming less harmonized globally, and customers are demanding sustainable products that are friendly to the environment. Owner operators in all industries are maximizing the value that can be extracted from their projects and have been seeking out new and innovative ways to drive greater efficiency and productivity. Efficiency, safety and quality must be guaranteed and therefore more stringent control over compliance and regulation requirements has to be at the heart of every aspect of the design process. Profitability, safety and efficiency must be intrinsically linked.

Today, companies need to find ways to be more efficient in the execution of the engineering projects. One way is to compress engineering cycles and adapt processes and products to comply with sustainable KPIs and new demands. The Unified Engineering methodology is one option enabled by the latest technologies and tools available. For decades, engineering to design and build industrial plants has been developed through projects with a complex and highly iterative workflow, using siloed solutions. The entire process is very time consuming, and the final deliverable to the owner of the operation is a set of documents with disperse data, with no guarantee of consistency.

In order to compress engineering cycles, a data-centric approach must be used. With this approach, documents and applications are always kept up to date with the latest validated data. The data-centric approach is the first step to apply the Unified Engineering methodology, which will evolve later to the plant Digital Twin. The Unified Engineering methodology uses a single source of information that is available for all the teams involved in the project. Engineers become more efficient and work with reliable information as documents and models are updated in a controlled way as soon as any change is made. For even greater sustainability and effectiveness, the best way is to use the Unified Engineering methodology in the cloud.

Unified Engineering reduces capital project costs, risks, and delays, enabling shorter engineering cycles required to deliver new sustainable projects. By minimizing engineering errors and accelerating project execution, companies can get 50% faster FEED stage, 30% increase in engineering efficiency, and 5% reduction in TIC (Total Installed Cost).

Keywords: sustainability; simulation; design; digital twin; predictability

1. Introduction

As an industrial software provider, we support chemical companies achieve superior performance in their quest to make sustainable products, align with the circular economy and demonstrate product stewardship throughout the product life cycle. By digitally connecting assets, process, and people, our solutions empower companies to run safe and responsible operations, mitigating EHS risks, and moving toward more circular systems, while remaining profitable.

With more than 50 years of industrial software innovation, AVEVA enables 13 of top 15 chemical companies and most of the world's petrochemical crackers, with the most comprehensive portfolio that ties profitability to sustainability goals.
Research typically identifies the following Critical Sustainability Drivers:

- Demand higher transparency on a company's environmental, social & governance (ESG) performance by stakeholders
- Manage a complex environment and provide safety to employees, processes, products and local communities
- Minimize energy and utilities consumption, so as emissions
- Manage liquid effluents, waste and suppliers, seeking the circular economy
- Commitments to global and local regulations
- Portfolio management towards innovation and sustainable products

Today´s technology allows suppliers to develop better solutions to those markets that are in constant change. The Digital Twin technologies, initially adopted mainly by the automotive and aerospace industries, are now promoting big changes in how chemical plants are operated and managed. This type of technology can change the decision-making process since more reliable information is available in real time.
A significant step was taken recently in terms of process simulation. Now it is possible for the engineering and operating companies to build the Digital Twin of the process plant. The Digital Twin is built on a simulation platform that will support the entire plant lifecycle, from design to operation.

2. The Digital Twin

The new generation of process simulation uses a platform approach that evolves the simulation model from the conceptual engineering to the operation optimization. It allows a new approach to be implemented so companies can transition from the conventional scenario to the use of the process simulation Digital Twin, expanding benefits to the entire plant lifecycle. The same platform is used for process simulation and process utilities (cooling water, flare, steam and others), allowing engineers to further evaluate how each system impacts the other. Heat and material balances can be re-evaluated after equipment and pipeline sizing, since that information is in the simulation from the beginning, as a result, little or no extra engineering effort is required. Once sizing is validated, the simulation is switched to dynamic mode, in which control loops are included to the simulation model to validate the process control strategy. As it is seamless to shift the simulation to dynamic mode, rather than build a completely new model using the conventional approach or converting a model that cannot be taken back to the steady state mode, dynamic studies are performed earlier in the project lifecycle. This promotes savings in equipment acquisition and in operating costs, since control logic responses are

evaluated in earlier stages. Plus, when something doesn´t respond as expected, simulation is taken back to steady state mode, for re-evaluation of heat and material balance and re-sizing. The ability to go back and forth between steady state and dynamic modes is critical to increase efficiency in the project lifecycle, leading to huge savings in engineering effort.

3. Unified Engineering

Companies are beginning to make progress on their digitalization journey, finding the right applications for digital transformation and seeing increasingly better returns on their investment. While the age-old market environment challenges (such as supply and demand, cost and price) haven't gone away, competitive pressures are making the digital transformation opportunity more pressing than ever. Many have already started to leverage the latest data-centric technology and work processes for their workforce to collaborate and take control of their data, reducing the risk for errors, delays and increased project cost throughout the asset lifecycle. By doing so they are in a stronger position to become more competitive, increase their margins and win new business.

As outlined by an AVEVA whitepaper (2019), Unified Engineering is a new proposition to break down the silos between FEED and Detailed Design to minimize risk and maximize return on Capital Investment.
Unified Engineering enables global multi-discipline teams to work concurrently in a common data-centric environment, controlling and managing change across the entire project. This breaks down the silos between FEED and detailed design. The simulation data created in FEED is readily available for use in detailed design and is checked and validated in real-time, increasing efficiency, minimizing risk, and maximizing return on investment on your Capital Projects.

Never have the stakes been higher for companies when it comes to making improvements to their engineering work processes to maximize ROI on Capital Projects. Productivity has not developed in decades – the average Capital Project schedule lags by 20 months and goes over budget by 80%. These are results that have been shared by Jayanth (2017) during the Rice Global E&C Forum Roundtable.
In many of today's Capital Projects, there is a disconnect between FEED and Detailed Design. In response, AVEVA are the first industrial software provider to pioneer a new solution to break down the silos between these engineering disciplines.

Unified Engineering consists of two main components, the Unified Lifecycle Simulation Platform (one model), and Integrated Engineering and Design (one database). The two are combined to form a robust process model and an engineering database that is able to synchronize through bi-directional flow of all 1D, 2D and 3D data on one platform. The bi-directional integration of a steady state and dynamic process model with an engineering database makes the process seamless and eliminates the need for MS Excel or other intermediate steps to transfer information between tools.

With the Unified Engineering model, you can have:

- Verification that the plant will operate as expected, and that controls are properly configured
- Verification that equipment and piping are properly sized
- One single version of the truth that remains up to date

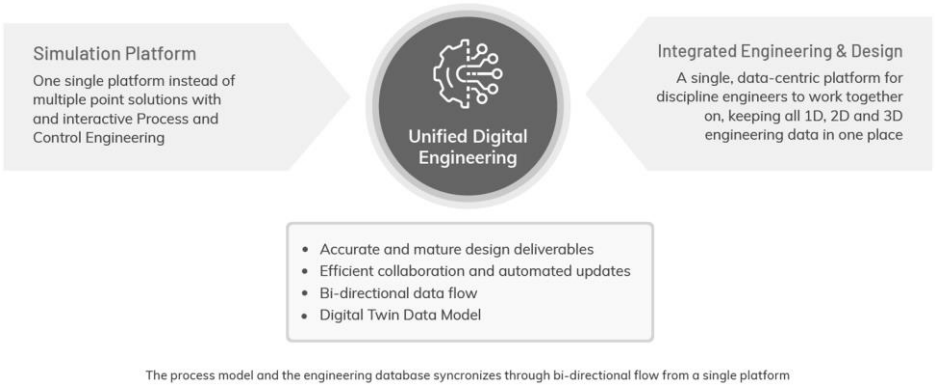


Figure 1 - Unified Engineering principles

4. Process Simulation lifecycle

Process simulators are irreplaceable tools for every process engineer. Since the nineteen seventies, process simulators have found widespread adoption within operating companies in oil & gas, refining and chemical industries, as well as the engineering companies and equipment manufacturers that service these industries. The tools available in the market today have incrementally improved over the years to provide more features and functionality. However, they trace their origins to legacy architectures, operating systems and aftermarket user interfaces, which create inherent limitations:

- They cannot support the full plant lifecycle as they are limited by their single-purpose architecture such as steady state process simulation, dynamic simulation, optimization, or flow network analysis for which they were originally designed
- Extending their functionality can be performed by a very small number of software developers with chemical engineering knowledge, software programming skills, and/or knowledge of that particular specialized program
- They are often based on decades old programming code that cannot leverage the more recent technological developments within the software industry

Today's simulators typically only support a single phase of the lifecycle and are often based on thermodynamics of different simulation vendors and different calculation methods. This not only leads to lack of trust in the results but causes substantial rework by having to build a new simulation model in each new tool. And the results are hard to compare.

The technological limitations of incumbent process simulation tools forestall improvements in engineering workflows. Collaboration only occurs outside the simulation because the software cannot accommodate it. Engineering departments send analysis and optimization questions to outside specialists because they cannot easily perform advanced simulations with their in-house tools and software expertise. These complications trap engineering workflows in a waterfall project management paradigm where development is forced into a linear process. Iteration is to be avoided because

legacy software makes it cumbersome, error-prone, and tedious. Willetts and Depew (2020) describe in detail the current challenges and how a Process Digital Twin will significantly help improve efficiency and drive increased sustainability.

Global competition, pricing pressure and energy alternatives are now driving the need for a new approach. The oil & gas industry has seen high volatility and the lower price level of today is seen as the "new normal". The chemicals industry has a continuous need to innovate for greater agility and lower costs.

The next generation of workers also expects a modern, scalable and easy to use solution with technology they now take for granted – high speed internet access, mobile devices, touch screens and virtual reality. New concepts like the Industrial Internet of Things (IIoT), Industry 4.0, and Artificial Intelligence have created greater opportunities with a new next generation platform that provides a "Digital Twin" of the plant through the process lifecycle that cannot be provided with today's tools.

A next generation process simulation platform means that one process model is extended throughout the entire lifecycle of the plant, from concept through to operations. This requires a process design mode, a fluid flow/rating mode and a dynamic mode, in combination with the ability to toggle back and forth between modes. Optimization may be provided to any mode.

A single, easy-to-use simulation platform will allow engineers to move seamlessly between questions of design, analysis, and optimization. Engineers will be able to assess the impact of design and specification changes quickly and with a holistic view of multiple disciplines. Intensive collaboration becomes commonplace. In this environment, organizations will be able to adopt agile engineering workflows based on smaller pieces of work with continuous integrated testing to reduce development cost while eliminating surprises at the end of the project.

5. Conclusion

Lifecycle process simulation has been a vision for process simulation providers and their customers for a long time. Today's simulators cannot leverage the rapid developments occurring in the software industry due to legacy architecture.

Looking at the industry's increasing demand for higher transparency, this can only be achieved using a data-centric Digital Twin approach. This data-centricity enables an ideal platform for new product and process development to create new models and include the management of complex environments bringing together both steady-state and dynamic simulation with constant iteration and constant solving capabilities.

The integration of the process analysis and simulation with other Engineering & Design disciplines also allows to breakdown the silos that were typically existing previously. Connecting and remotely controlling previously unconnected processes will increase sustainable operations and improve business efficiency in a sustainable environment. Over the long term the impact will drive resilience and sustainable performance through technologies.

References

Portfolio Marketing group, AVEVA, 2019, Unified Engineering: A new proposition to break down the silos between FEED and Detailed Design to minimize risk and maximize return on Capital Investment – Whitepaper

T.G. Jayanth, Expert – Capital Projects & Infrastructure, McKinsey & Co, 2017, Reinventing Construction - a route to higher productivity, McKinsey, Rice Global E&C Forum Roundtable

Ian Willetts, Vice President Process and Simulation, Cal Depew, Product Owner, AVEVA, 2020, Building a process Digital Twin for your plant with Unified Lifecycle Simulation – Whitepaper

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50166-4

Development of predictive model for the size of gas and liquid slugs formed in millimeter scaled T-junctions

Fabian Lechtenberg[a,b], Osamu Tonomura[b,*], Satoshi Taniguchi[b], Shinji Hasebe[b]

[a]*Aachener Verfahrenstechnik, RWTH Aachen University, Forkenbeckstraße 51, Aachen, 52074, Germany*
[b]*Department of Chemical Engineering, Kyoto University, Kyoto 615-8510, Japan*
**Corresponding Author's E-mail: tonomura@cheme.kyoto-u.ac.jp*

Abstract

Gas-liquid slug flow in micro/millimeter-scale channels is attractive in a broad range of applications. The sizes of slugs need to be known precisely, since the mass transfer characteristic depends on the slug length. To estimate the slug length from design and operating conditions, a number of models have been developed so far for T-junctions, but parameters of such models need to be empirically determined through many experimental trials. To reduce the experimental efforts, the physically sound model has been developed but are limited to micrometer-scale T-junctions. In this study, the applicability of the existing physically sound model is experimentally investigated in the case of millimeter-scale T-junctions. The results show that the volume added to the slug during the squeezing period is strongly affected by the condensed phase or liquid volumetric velocity. Taking this result, the combination of physical and empirical models is newly developed in this study. Our developed model will ensure the high accuracy in the design of the millimeter-scale T-junctions with gas-liquid slug flows.

Keywords: Millimeter-scale T-junctions; Gas-liquid slug flow; Slug generation mechanism; Slug size; Process modelling.

1. Introduction

Gas-liquid slug flows are attractive in a broad range of applications, such as in chemical, bio-chemical and material synthesis, drug discovery, medical diagnostics (Suryawanshi et al., 2018). The slug generation mechanism can be divided in different regimes such as squeezing, dripping and jetting, based on the capillary number (Zhu and Wang, 2016). The focus of the present research lies on T-junctions operating in the squeezing regime (cf. Fig. 1). The squeezing regime is governed by surface tension forces, making the pinch-off mechanism independent from fluid parameters like viscosity. Extensive work

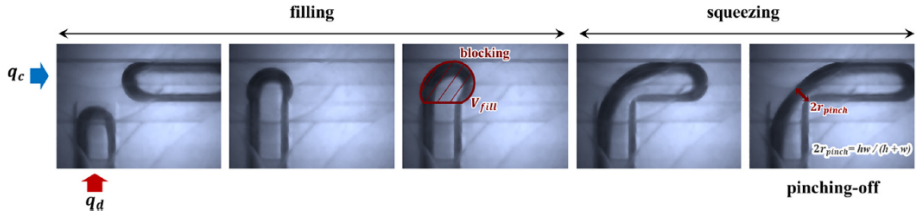


Figure 1. Slug generation mechanism in a T-junction with liquid flowrate Q_c and gas flowrate Q_d.

in this area has been done which resulted in models trying to predict the slug size from feed flows and channel geometry. In the present research, experimental validation of the proposed model by van Steijn et al (2010) is performed in millimetre scaled channels and channel geometries that have not been tested for this model. Based on the results, the shortcomings of the model are explained and tried to overcome.

2. Method

This chapter introduces the applied methods. The model by van Steijn et al. (2010) and the optical sensor and micrograph measurements is briefly explained.

2.1. Steijn's Model

The proposed model by van Steijn et al. (2010) enables the prediction of slug sizes within the squeezing regime at low capillary numbers $Ca < 0.01$. In this area, the slug formation mechanism is dominated by surface forces rather than viscous forces. It is a physically sound expression based on continuity, geometrical assumptions and the pinch-off mechanism. The slug formation mechanism can be divided into a filling and a squeezing part (cf. Fig. 1). The filling period ends when the bubble confines the whole main channel, leading to the accumulation of continuous phase fluid behind the bubble. This accumulating fluid squeezes the neck of the bubble during the squeezing period until it collapses. The upcoming dispersed phase then starts to fill the main channel again. This cyclic mechanism creates a segmented flow within the main channel.

In Steijn's model, it is possible to determine the filling shape of a bubble from simple geometrical calculations. During the squeezing period, the continuous phase flow is obstructed by the bubble which confines the main channel. Continuity states that the incoming continuous phase accumulates behind the bubble and a part of it bypasses the bubble as gutter flow. Herein, the ratio of gutter flowrate (Q_{gut}) to continuous phase fluid flowrate (Q_c) is assumed to be constant 0.1 (van Steijn et al., 2009).

2.2. Optical Signal Measurements

Slug sizes can be measured by using pairs of optical sensors, which need to be placed along the channel, capturing the altering light transmission due to scattering and absorption. Using the signals from two optical probes placed a certain interval, the two signals can be cross-correlated to determine the time lag and then yield instantaneous superficial velocity of slug flow and slug lengths of gas and liquid (Ide et al., 2009).

2.3. Micrograph Measurements

A method to capture the gutter flow has been developed and implemented in Python. During the experiments, the T-junction needs to be aligned with the camera's field of vision. The video's contrast should be high and capturing the whole pinch-off area for accurate performance. The procedure to acquire the gutter flow is automatized.

3. Experimental

This chapter describes the experimental setup and procedure.

3.1. Setup

The experimental setup composed of three sections can be seen in Fig. 2. The first one being the green underlined liquid feed section. The continuous liquid phase (ethanol) is supplied by a syringe pump. Additionally, the pressure in the liquid feed line is

observed to determine the start of steady state operation. The second section is the yellow underlined gas feeding section. The dispersed gas phase (nitrogen) comes from a gas cylinder. The flowrate is regulated by a mass flow controller (MFC). A subsequent mass flow meter (MFM) is used to check for steady state operation. The gas then flows through a barrier channel (BC) into the T-junction. The barrier channel is used to enable steady state operation by reducing the effects of pressure swings within the channel on MFC. The third section is the blue underlined channel section. A variety of channels has been used. In all of them, the continuous liquid phase intersects with the dispersed gas phase coming from the perpendicular oriented feed channel. At this point, the segmented flow starts in the main channel. The T-junction itself is observed by a camera. Further downstream, the pair of optical sensors is placed, observing the segmented gas-liquid flow. The fluids are then led into a beaker at constant liquid level to ensure constant outlet pressure during the operation.

Five different device geometries have been used. Their dimensions are given in Table 1. The devices were fabricated both in polymethyl methacrylate (PMMA) and glass. The relative production accuracy is 5% of the stated lengths. This information was confirmed with the available microscope and a micro scale.

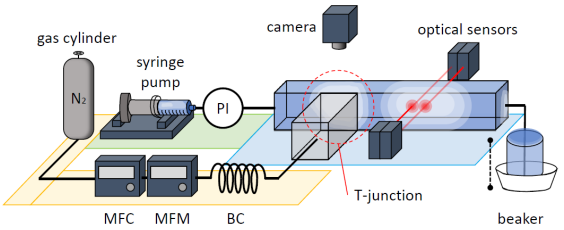


Figure 2. Schematic illustration of experimental setup.

Table 1. Millimeter-scale devices.

device	h [mm]	w [mm]	w_{in} [mm]
A-1	1.0	1.0	1.0
B-1	1.5	1.5	1.5
C-1	2.0	2.0	2.0
E-1	1.0	2.0	2.0
E-3	1.0	2.0	1.0

3.2. Procedure

For each device, the connection to the feeding lines and beaker were reconnected. The surface of the T-junction should be cleaned for optimal image quality. The temperature of the ethanol supplied to the syringe pump was adjusted to room temperature. During start-up procedure, the gas must be supplied first to prevent backflow of liquid in the gas feed line. The setpoints for the feed flowrates are inserted into the syringe pump and MFC. The pressure signal from the liquid feed line and the display from the MFM are observed to determine reach of steady state operation. Then, the sensor signals are recorded for 60s at a sampling rate of 1ms^{-1}. During this time, the video data was captured for usually 15s. After the data has been saved for one configuration, the feed rates were adjusted again.

4. Results and discussion

This chapter summarizes and discusses the main results of the conducted experiments. In the first section, the prediction accuracy of Steijn's model is evaluated. After that, the experimental validation of the gutter flow evolution is given. The following section shows how additional process information influences the prediction accuracy of Steijn's model. Based on this finding, an additional series of experiments was conducted with the goal to derive a gutter flow prediction submodel. The derivation and prediction results are described in final section.

Table 2. Prediction accuracy of Steijn's model for dimensionless gas slug volume.

(a) PMMA

Device	Max [%]	Min [%]	Mean [%]
A-1	-43	-12	29
B-1	-46	-21	31
E-1	31	0	8
E-3	-28	-5	17

(b) Glass

Device	Max [%]	Min [%]	Mean [%]
A-1	-41	-7	22
B-1	-37	-2	16
E-1	34	3	19
E-3	24	-2	8

4.1. Steijn's Model Evaluation

The measured dimensionless gas slug volume from the experiments has been compared to the predictions by Steijn's model. Table 2 shows the boundaries and mean prediction errors. It can be seen that there is a general underestimation of the gas slug volumes. The underlying reasons are discussed in the following section.

4.2. Gutter Flow Ratio

Q_{gut} is one of influencing parameters in Steijn's model as described in section 2.1. The gutter flow during the squeezing period was captured for every conducted experiment. Q_{gut} can be determined by the Python script from the micrograph video data. Figure 3 shows a scatter plot of all gutter flow ratios over linear flow velocities for all conducted experiments. From this data it is clear that the assumption of a global constant gutter flow ratio does not hold for the scope of the performed research. It is confirmed that bigger channels inhibit higher gutter flow and flat channels E-1 and E-3 have lower gutter flows than quadratic channels. There is a slight material dependency of the gutter flow ratio. For glass devices the measured ratio was generally lower. This could be due to the corner roundness of the glass devices. The sharp corners of PMMA devices may lead to a bigger gutter area, allowing for more fluid to pass the bubble. A general trend of decreasing gutter flow ratio with increasing flow velocity can be seen. While the flat devices inhibit gutter flow ratios in the vicinity of Steijn's assumption, the other devices have considerably higher ratios. In the following, it is tested if the use of additional information about the ratio in Steijn's model improves the prediction accuracy and how the gutter flow ratio can be modeled.

4.3. Incorporated Gutter Flow

Using the additional information about the gutter flow, it is possible to calculate adjusted gas slug sizes from Steijn's model. For each experiment the information about the gutter flow ratio has been inserted into the prediction model resulting in new prediction errors. As a result, the prediction accuracy improved. Figure 4 visualizes the error distribution of PMMA and glass devices before and after the incorporation of gutter flow data. It can be seen, that in most cases the width of the error region narrows and gets shifted towards zero. The mean error of the PMMA predictions drops from 21.6% to 11.9%. For glass devices the error drops from 16.8% to 9.7%.

4.4. Gutter Flow Model Construction

From the previous experiments, it was shown that using the actual measured gutter flow ratio in the prediction model leads to improved prediction accuracy. For system design purpose, it is necessary to have a descriptive model for the gutter flow ratio. Here, a possible model derived from experimental data is constructed and evaluated.

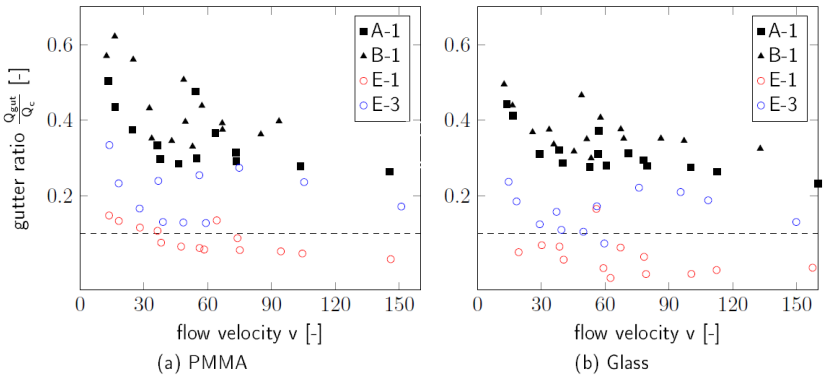


Figure 3. Gutter flow to continuous phase ratio of all conducted experiments for PMMA and glass devices. The dashed line shows Steijn's model assumption of a constant ratio of 0.1.

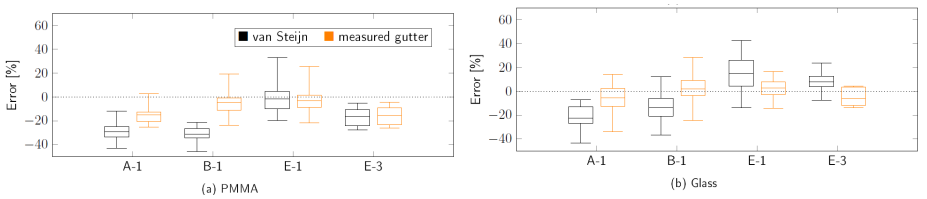


Figure 4. Prediction error of Steijn's model with conventional assumption (black) and with incorporation of gutter flow data (see Fig. 3) (orange).

The main influencing factors for the gutter flow ratio appeared to be the channel geometry, the flow velocity v and the ratio of feed flowrates Q_d/Q_c. To reduce the complexity and experimental effort, in this research the geometrical influence was not included in the model but the extrapolation ability of the model has been tested. A set of experiments has been designed to cover the variable space ranging from v = 4-70 mm/s and Q_d/Q_c = 0.5-4. In total 27 samples have been acquired. It turned out that with increasing v and Q_d/Q_c, Q_{gut}/Q_c decreases. As there is no physical description of this mechanism, it is tried to fit a function candidate to the data. Plotting the measured data points according to Eq.(1) leads to a scatter that resembles a hyperbolic dependency for both variables. For this reason the following model has been selected:

$$\frac{Q_{gut}}{Q_c} = a \cdot v^{-b} \cdot \left(\frac{Q_d}{Q_c}\right)^{-c} \tag{1}$$

The three empirical parameters a, b, c have been determined from least squares optimization. A cross-validation procedure was chosen in which four random samples have been chosen from the training set. The remaining samples were used for the parameter fitting. This procedure was repeated ten times to get an impression of the model sensitivity towards the selected training samples. The optimal parameters fall within a short range around a = 0.9, b = 0.2, c = 0.1 and the cross-validation shows high robustness towards training sample selection. It can be concluded that the constructed gutter flow model performs well within its training range for glass device C-1. To test the models extrapolation ability, it was used to predict Q_{gut}/Q_c for the other devices. The results can be seen in Table 3(a).

Table 3. Extrapolation ability of the constructed gutter flow model.

(a)

Device	C-1	A-1	B-1	E-1	E-2
RMSE [-]	0.011	0.084	0.041	0.356	0.240

(b)

Device	C-1	A-1	B-1	E-1	E-2
RMSE [-]	0.034	0.060	0.031	0.328	0.215

It is evident that the constructed model is only valid within its training region for device C-1. The closest prediction accuracy is achieved for device B-1 which has the same aspect ratios only scaled down. The accuracy for A-1 is even worse and the prediction for E-1 and E-3 barely compare to the measured data. The gutter flow ratio behaves similar for devices with same aspect ratios (A, B and C-1) as can be seen from Fig. 3. To obtain a model that can describe the gutter flow for channels with aspect ratios $h = w = 1$ and $w_{in} = w = 1$, it was tried in a next step to fit the parameters a, b and c to the whole database for these devices. In this case, the training set consisted of 57 samples and a "10-fold leave 6-out" cross validation approach was applied. The best model in this training procedure slightly differs from previous one. The newly obtained parameters are $a = 0.94$, $b = 0.23$ and $c = 0.099$. The prediction accuracy for this model can be seen in Table 3(b). Using the broader training data leads to a more uniform prediction error for the quadratic devices. The accuracy for B-1 and A-1 increased slightly on the cost of C-1 accuracy. For the flat devices E-1 and E-3 no qualitative difference can be identified.

5. Conclusions

In this research, an experimental evaluation of a model for slug size prediction in T-junctions has been performed. The channel sizes in this study were one order of magnitude bigger than those from the original study and made from PMMA and glass. The chemical system was a gas-liquid system of ethanol as continuous and nitrogen as dispersed phase. For these conditions, mean prediction error of 21.6% and 16.8% for PMMA and glass devices respectively could be achieved with Steijn's model. Methods were developed to calculate the actual slug volume from light sensor data and capturing the gutter flow volume from micrograph data. Incorporating the measured gutter flow data improved the prediction accuracy to 11.9% and 9.7%. An empirical gutter flow model was derived and trained to the acquired database. This model was found to describe channels of same aspect ratio (A-1, B-1, C-1) with a reasonable accuracy while the flat channels (E-1, E-3) cannot be described by the constructed model.

References

H. Ide, R. Kimura, M. Kawaji, 2009, Optical measurement of void fraction and bubble size distributions in a microchannel, Heat Transfer Eng. 28, 713-719.

P.L. Suryawanshi, S.P. Gumfekar, B.A. Bhanvase, S.H. Sonawane, M.S. Pimplapure. 2018, A review on microreactors: Reactor fabrication, design, and cutting-edge applications, Chemical Engineering Science, 189, 431-448.

P. Zhu, L. Wang, 2016, Passive and active droplet generation with microfluidics: a review. Lab Chip, 17, 1, 34-75.

V. van Steijn, C.R. Kleijn, M.T. Kreutzer, 2010, Predictive model for the size of bubbles and droplets created in microfluidic T-junctions. Lab Chip, 10, 19, 2513-2518.

V. van Steijn, C.R. Kleijn, M.T. Kreutzer, 2009, Flows around confined bubbles and their importance in triggering pinch-off. Phys Rev Lett, 103, 21, 214501.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50167-6

Enviro-economic assessment of DME synthesis using carbon capture and hydrogen from methane pyrolysis

Andrea Bernardi[a,b], Fatima Bello[a], Antonio Valente[c], David Chadwick[a], Gonzalo Guillen-Gonzalbez[c], Benoit Chachuat[a,b,*]

[a] *Sargent Centre for Process Systems Engineering, Imperial College London, UK*
[b] *Department of Chemical Engineering, Imperial College London, UK*
[c] *Department of Chemistry and Applied Bioscience, ETH Zurich, Switzerland*
Corresponding Author's E-mail: benoit.chachuat@imperial.ac.uk

Abstract

The catalytic conversion of captured CO_2 and H_2 into fuels is recognised as an interesting option to decarbonise the transport sector in the short-midterm future. DME has been identified as an ideal diesel-substitute for heavy-duty vehicles due to its high cetane number and excellent combustion properties, but to be competitive with diesel a low-cost and low-carbon H_2 production route is a key enabler. Recent developments indicate that methane pyrolysis has the potential to produce H_2 at a similar cost compared to steam methane reforming, the main H_2 production route nowadays, yet with no direct CO_2 emissions. This paper presents an enviro-economic assessment of 12 life-cycle pathways for DME production. Our results show that DME produced using H_2 from methane pyrolysis could be competitive with diesel, both economically and environmentally, but is highly dependent upon the utilisation of the carbon by-product.

Keywords: Dimethyl ether; Process simulation; Life cycle assessment; Enviro-economic assessment.

1. Introduction

Decarbonisation of the transport sector is a long-standing challenge for our modern societies and an important step towards a sustainable future. Fuels that can be produced from CO_2 hydrogenation (methanol, dimethyl ether (DME), oxymethylene ethers (OME), Fisher-Tropsch fuels) are gaining significant attention. DME, in particular, has been identified as an ideal diesel-substitute for heavy-duty vehicles (HDVs). The combustion of DME compared to diesel offers several advantages (no particulate matter or SO_x emissions, reduced NO_x and CO_2 emissions), and major truck producers such as Volvo and Mack have already developed and tested DME-fuelled trucks with positive results (Szybist *et al.* 2014). While the DME production process has been extensively studied in the literature, the economic feasibility and environmental benefits of DME over its whole life cycle are yet to be proven (Royal Society, 2019).

In a recent paper, Matzen and Demirel (2016) carried out a life cycle assessment of DME from methanol dehydration, considering CO_2 and H_2 from water electrolysis powered by wind energy as raw materials and found that DME outperforms diesel from an environmental perspective. However, a techno-economic analysis focusing on the same route showed that DME is not cost competitive in the current market conditions (Michalios *et al.* 2019). The main cost driver is H_2 production, which is 4 to 6 times more

expensive when produced by renewable energy compared to methane reforming, the main H_2 production route nowadays (Al-Quatani *et al.* 2021). This is in line with the results of similar studies focusing on other alternative fuels such as methanol (Gonzales-Garay *et al.* 2019) and OME (Rodríguez-Vallejo *et al.* 2021). Recent developments indicate that methane pyrolysis, a thermal decomposition of methane into H_2 and solid carbon in a non-oxidative environment, produces H_2 with no direct CO_2 emissions and at a cost between \$0.4 and \$1.9 /kg_{H2} which compares favourably against steam methane reforming, for which the estimated cost is \$1.2 /kg_{H2} (Parkinson *et al.* 2018, 2019).

The objective of this work is to assess DME production pathways that use H_2 from methane pyrolysis and compare them with diesel, when both are used to power HDVs. A distinctive feature of our analysis is that it relies on detailed kinetic models, calibrated against experimental data to simulate the DME synthesis process and generate the foreground inventories used in both the techno-economic analysis (TEA) and life-cycle assessment (LCA). A total of 12 life-cycle pathways for DME production are compared with diesel from and economic and environmental point of view using a well-to-wheel (WTW) approach.

2. Material and methods

Figure 1 summarises the DME pathways considered in this work and the following subsections provide additional information on the modelling assumptions and approaches used to carry out the TEA and LCA.

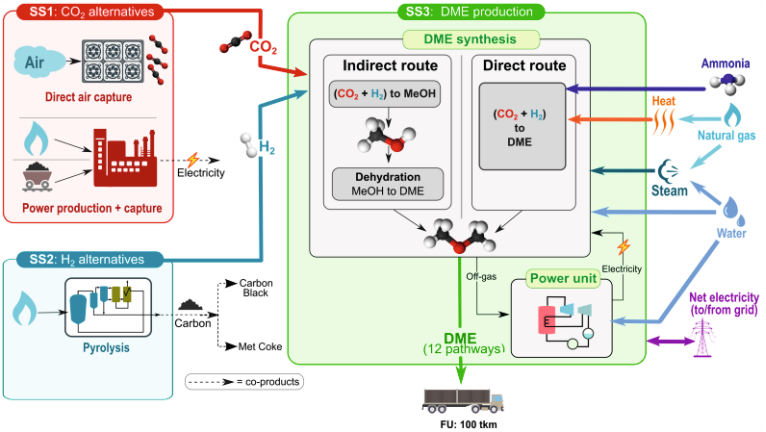


Figure 1: scope of the DME infrastructure divided in CO_2 procurement (SS1), H_2 production (SS2), DME production (SS3) and DME utilization in HDV.

2.1. Process simulation and economic analysis

Our focus is on two DME production processes: the direct synthesis where CO_2 and H_2 are converted to DME in a one-step process; and the indirect synthesis that first produces methanol that is subsequently dehydrated to form DME. Detailed process simulations relying on kinetic models calibrated against experimental data were implemented in Aspen HYSYS (version 11). The modified Peng-Robinson fluid package (PRSV) was used to simulate the methanol synthesis, while the UNIQUAC with the Redlick-Kwong activity model was used for methanol dehydration and direct DME synthesis.

Feed composition. Both DME synthesis processes are fed with a mixture of CO_2 and H_2 with a stoichiometric molar ratio of 1:3. The CO_2 is assumed to be available at 1 bar with

2.5 $\%_{mol}$ of water and nitrogen, regardless the source. The H_2 is available at 30 bar, with 2.5 $\%_{mol}$ of methane.

Indirect DME synthesis. The indirect DME plant is based on previous work by Van Dal *et al.* (2013) for methanol synthesis and by Luyben (2017) for methanol dehydration. Methanol is produced in an adiabatic reactor operating at 75 bar with a feed temperature of 210 °C, in accordance with Michalios *et al.* (2019). The unreacted gases are separated from the liquid products (methanol and water) in two consecutive flash drums then a distillation column separates the methanol from the water. A small (2 %) fraction of the unconverted feed is purged and burnt to produce electricity, while the rest is recycled back to the reactor. The electricity production unit is simulated as a syngas-fired gas turbine integrated with a steam turbine. The gas turbine cycle comprises a compressor to pressurize the air at 15 bar, a combustion chamber represented as a conversion reactor in Aspen HYSYS, and an expander. The hot flue-gas is used to produce additional electricity with a Rankine cycle. Operating variables, such as the pressure of the combustion chamber, and parameters, such as the compressors and turbines efficiencies, are taken from Narvaez *et al.* (2019). The methanol is fed at 240 °C to the DME reactor where it is dehydrated under 12 bar. The reactor is cooled to keep the internal temperature lower than 400 °C to avoid catalyst deactivation, and the outlet stream exits at 360 °C. Two distillation columns are used to separate the products. DME is extracted from the top of the first column at the 99.5 $\%_{mol}$ purity, while methanol is separated from the water in the second column and recycled to the methanol reactor.

Direct DME synthesis. In the direct DME process the mixture of H_2 and CO_2 is fed to an isothermal reactor operated at 55 bar and 260 °C (Bernardi *et al.*2020). The outlet stream contains the unreacted syngas, DME, methanol and water. It is cooled to 35 °C before entering an absorption column where a mixture of water and methanol is used as the absorbing agent. The gas stream leaving the absorption column has most of the unreacted H_2, CO, and CO_2 with traces of DME. 2 % of the unreacted syngas is used to produce electricity with the combined gas and steam turbine described before, while the rest is recycled to the reactor. DME, methanol and water exit the absorption column in the liquid stream and are fed to a distillation train. The first column separates water and methanol from the DME and gaseous impurities. Water and methanol are in part recycled to the absorption column and in part fed to a second distillation column for their separation. The top stream of the first column is fed to a third column operating under 10 bar where DME is recovered from the bottom at 99.5 $\%_{mol}$ purity. A small (3 %) fraction of DME is lost from the top of this column in order to maintain a temperature above –20 °C in the condenser. A refrigeration cycle with ammonia as working fluid is used to operate this cryogenic distillation column.

Economic analysis. The capital and operating costs of the DME production plants are estimated using Aspen Economic Analyzer and considering that the plant is operating in the UK. An annual capital charge of 0.16 is used to annualize the capital cost. The CO_2 and H_2 prices are taken from Gonzales-Garay *et al.* (2019) and Parkinson *et al.* (2018) respectively. The H_2 price depends on the selling price of the carbon by-product, and values of $150 /ton and $500 /ton are assumed for the metallurgical coke (MC) as carbon black (CB). The diesel production cost considered here is the average price excluding VAT and duty between 2017-2021 reported in BEIS (2021).

2.2. Life-cycle analysis

Goal and scope: The functional unit (FU) for the LCA analysis is chosen as "*1 ton transported over 100 km with a heavy-duty vehicle*" (i.e.: FU = 1 tkm). The assessment

of DME entails the analysis of the fuel use, as well as three background subsystems as shown in Figure 1. Subsystem SS1 involves the procurement of CO_2 and three alternatives are considered in this work: direct air capture (DAC), carbon capture from a natural gas power plant (NG), and carbon capture from a coal power plant (C). Subsystem SS2 is concerned with the H_2 production via methane pyrolysis. Two alternative uses of the carbon by-product are considered: carbon black (CB) and metallurgical coke (MC). Subsystem SS3 is the DME production step using either the direct or indirect process. Apart from DAC the other CO_2 and H_2 production pathways are multi-product and require dealing with *multi-functionality*. Following the recommendation by von der Assen *et al.* (2013) *system expansion via substitution* was adopted in this work. It is important to note that CB is a more valuable product compared to MC, but CB market demand is two orders of magnitude lower than MC and would not be enough if a large share of the HDV fleet was to be converted from diesel- to DME-fueled trucks.

Data collection and life cycle inventory: The life-cycle inventories (LCIs) for SS3 are obtained by the simulations described in Section 2.1, while the inventories for the other subsystems are gathered from the literature:

- SS1: LCIs are taken from Gonzales-Garay *et al*. (2019). The electricity produced by the power plants is assumed to substitute electricity from a conventional power plant located in UK and using the same fuel;
- SS2: The inventory data for methane pyrolysis are derived from Parkinson *et al.* (2018), and the credits for the by-products are taken from the ecoinvent database;
- Fuel use: fuel consumption and associated direct emissions are estimated for diesel and DME using the software GREET (version 1.3), considering heavy-duty trucks.

Environmental impact assessment. ReCiPe2016 is the impact assessment methodology used in this work (Huijbregts *et al.* 2017). The methodology provides midpoint indicators, which quantify the effects of resource utilization and emissions on a specific environmental category (e.g.: global warming); and endpoint indicators, which represent the three areas of protection: human health, ecosystems quality, and resources. In our analysis we converted the endpoint indicators into a common monetary basis according to the economic penalties proposed by Weidema (2015), a process known as monetization of the environmental externalities. The environmental assessment is conducted in Simapro (version 9.1.1.1), using ecoinvent 3.6 for the background process inventories.

3. Results and discussion

3.1. Midpoint environmental analysis - emissions

Figure 2a compares the GWP of DME with diesel in our WTW analysis. The labels indicate: the DME process, (indirect synthesis, iDME; direct synthesis, dDME); the use of carbon by-product (metallurgical coke, MC; carbon black. CB), and the CO_2 source (direct air capture, DAC; carbon capture from natural gas power plant, NG; and from coal power plant, C).

We can observe that all the 12 pathways lead to a reduction of the GWP_{WTW}: the pump-to-wheel (PTW) impacts are similar to those of diesel, but the well-to-pump (WTP) impacts are negative for 11 out of 12 pathways. This is the consequence of the negative GWP of the raw materials: the CO_2 from DAC is accounted for as a negative emission, while the other multifunctional processes are associated with environmental credits for the avoided by-products. These credits exceed the direct emissions from the DME production stage, except for dDME-MC-C. We can also observe that iDME slightly

outperforms dDME when the same CO_2 and H_2 sources are considered. DME also outperforms diesel in terms of life-cycle PM, NO_x, and SO_2 emissions.

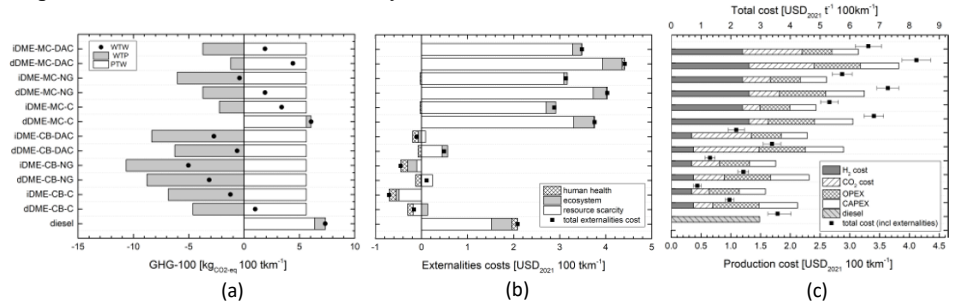


Figure 2: Enviro-economic assessment of DME- and diesel-powered HDV considering "*1 ton transported for 100 km*" as functional unit: (a) GWP; (b) externalities cost; (c) production cost. The error bars in (c) represent the minimum and maximum production costs considering a ±20% variation in the cost of the CO_2 and H_2 and historical prices for diesel in the period 2017-2021.

3.2. Endpoint analysis and total cost

Figure 2b shows the monetized values of the endpoint environmental impacts. We can observe that CB is the by-product of methane pyrolysis DME outperforms diesel, otherwise the total externalities cost of DME is higher than diesel. The damage to resources always represents the largest share. In the other two areas of protection DME is superior to diesel, except for impacts on ecosystems by dDME-MC-C and dDME-MC-DAC that are comparable to diesel. As for the GWP, iDME compares favorably against dDME also from an endpoint perspective.

Turning to economic considerations, finally, figure 2c shows the production and total cost per functional unit. We can observe that: (i) iDME outperforms dDME, and (ii) iDME is cost competitive with diesel only if the CO_2 is captured from power plants and CB is the by-product of pyrolysis. iDME outperforming dDME is the result of an easier separation between products and unreacted reagents, that leads to a lower amount of CO_2 and H_2 lost in the purge streams. CO_2 from DAC and H_2 when MC is the by-product are the most expensive alternatives, and if used together they account for approximately 60% of the total production cost. In terms of total cost (including externalities) the alternative selling CB as by-product achieves a lower total cost compared to diesel as a result of the significantly lower monetized externalities.

Conclusions

This paper has presented an enviro-economic comparative assessment of different DME production pathways using H_2 from methane pyrolysis and considering two alternatives for the solid carbon by-product. Using "1 ton transported by 100 km by HDV" as the basis of comparison, our results show that direct and indirect DME production routes using H_2 from methane pyrolysis have a similar enviro-economic performances, and that they both have the potential to be cost-effective and to present lower environmental impacts compared to diesel, provided that the carbon by-product of the methane pyrolysis is sold in the market as carbon black. If the solid carbon is sold as metallurgical coke, DME might still offer benefits at midpoint level in terms of GWP, PM, NO_x and SO_2 emissions, but at endpoint level DME has higher monetized externalities than diesel due to high damages to resources. Future work will aim to assess the robustness of our results incorporating a sensitivity analysis to account for uncertainty in the life-cycle inventories

and investigate the impact of key assumptions of the LCA study. Moreover, different low-carbon H_2 production routes, such as electrocatalytic routes using renewable electricity, will be included in the comparison to assess the role of methane pyrolysis in the context of sustainable H_2 production.

Acknowledgements
This paper is based upon work supported by the Engineering and Physical Sciences Research Council (EPSRC) under Grants EP/P016650/1 and EP/V011863/1.

References

A. Al-Qahtani, B. Parkinson, K. Hellgardt, N. Shah, G. Guillen-Gosalbez, 2021. Uncovering the true cost of hydrogen production routes using life cycle monetisation. Applied Energy, 281, 115958.

BEIS, 2021. https://www.gov.uk/government/statistical-data-sets/oil-and-petroleum-products-monthly-statistics. [last accessed 5 Nov 2021]

A. Bernardi, Y. Chen, D. Chadwick, & B. Chachuat, 2020. Direct DME synthesis from syngas: a techno-economic model-based investigation. Computer Aided Chemical Engineering 48, 655-660.

A. González-Garay, M.S. Frei, A. Al-Qahtani, C. Mondelli, G. Guillén-Gosálbez J. Pérez-Ramírez, 2019. Plant-to-planet analysis of CO2-based methanol processes. Energy & Environmental Science, 12, 3425

M.A. Huijbregts, Z. Steinmann, P. Elshout, G. Stam, F. Verones, M. Vieira, & R. Van Zelm, 2017. ReCiPe2016: a harmonised life cycle impact assessment method at midpoint and endpoint level. The International Journal of Life Cycle Assessment, 22(2), 138-147.

M. Matzen, & Y. Demirel, 2016. Methanol and dimethyl ether from renewable hydrogen and carbon dioxide: Alternative fuels production and life-cycle assessment. Journal of cleaner production, 139, 1068-1077.

S. Michailos, S. McCord, V. Sick, G. Stokes, & P. Styring, 2019. Dimethyl ether synthesis via captured CO2 hydrogenation within the power to liquids concept: A techno-economic assessment. Energy Conversion and Management, 184, 262-276.

A. Narvaez, D. Chadwick, & L. Kershenbaum, 2019. Performance of small-medium scale polygeneration systems for dimethyl ether and power production. Energy, 188, 116058.

W.L. Luyben, 2017. Improving the conventional reactor/separation/recycle DME process. Computers & Chemical Engineering, 106, 17-22.

B. Parkinson, M. Tabatabaei, D. Upham, B. Ballinger, C. Greig, S. Smart, E. McFarland, 2018, Hydrogen production using methane: Techno-economics of decarbonizing fuels and chemicals. International Journal of Hydrogen Energy, 43(5), 2540-2555.

B. Parkinson, P. Balcombe, J. F. Speirs, A. Hawkes, K. Hellgardt, 2019, Levelized cost of CO2 mitigation from hydrogen production routes. Energy & environmental science, 12(1), 19-40.

D. Rodríguez-Vallejo, A. Valente, G. Guillén-Gosálbez, & B. Chachuat, 2021. Economic and life-cycle assessment of OME 3–5 as transport fuel: a comparison of production pathways. Sustainable Energy & Fuels, 5(9), 2504-2516.

The Royal Society, 2019, Sustainable synthetic carbon-based fuels for transport: policy briefing, London, UK Royal Society

J. P. Szybist, S. McLaughlin, & S. Iyer, 2014. Emissions and performance benchmarking of a prototype dimethyl ether fueled heavy-duty truck. Papers of the american chemical society, 248, 1155.

E.S. Van-Dal, & C. Bouallou, 2013. Design and simulation of a methanol production plant from CO2 hydrogenation. Journal of Cleaner Production, 57, 38-45.

N. von der Assen, J. Jung, A. Bardow, 2013, Life-cycle assessment of carbon dioxide capture and utilization: avoiding the pitfalls. Energy & Environmental Science, 6(9), 2721-2734.

B. P. Weidema, 2015, Comparing three life cycle impact assessment methods from an endpoint perspective. Journal of Industrial Ecology, 19(1), 20-26.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50168-8

Operational Envelopes of Cost-effective Sour Gas Desulfurization Processes

Chinmoy B. Mukta, Selen Cremaschi and Mario R. Eden*

Department of Chemical Engineering, Auburn University, Auburn, AL 36849, USA
edenmar@auburn.edu

Abstract

Cost-effective desulfurization of sour gas is of paramount importance not only in terms of safety, but also for the effective utilization of natural gas resources, as a significant amount of the gas content of a shale oil well is currently burned or emitted, which is detrimental for both the environment and the economy. Therefore, we need robust and cost-effective desulfurization technologies that can handle a wide range of gas flowrates and hydrogen sulfide concentrations. There has been significant progress in developing new desulfurization technologies, however determining their cost-effectiveness at different scales remains a challenge. Process modeling of these emerging desulfurization processes can help elucidate the parameters/uncertainties that impact their performance and thus help identify the best solution for a given system configuration. In this work, three natural gas desulfurization processes, i.e., oxidative sulfur removal (OSR) SourCat™ (Deng *et al.*, 2020), iron-chelate redox systems LO-CAT® (Faramawy *et al.*, 2016) , and Triazine based scavenger (Lozano & Trauffer, 2000) are simulated for natural gas feeds ranging from 1 to hundred thousand MSCFD and inlet H_2S concentrations from 500 to 2,500 ppm.

Keywords: Desulfurization; Process Simulation;

1. Introduction

Fossil fuels are still an indispensable part of modern living for human society, keeping cars running and houses heated, for example. Due to the renewed emphasis on renewables and rigorous environmental scrutiny on other dirtier fossil fuels (coal, oil), there is an increasing demand for natural gas as a cleaner energy resource. Moreover, with disinvestment happening around the energy industry, it has become imperative to utilize natural gas resources effectively. Sour gas is a natural gas resource that contains a significant fraction of hydrogen sulfide (H_2S) and carbon dioxide. The high hydrogen sulfide content severely limits the technical, economic, and environmental viability of certain natural gas resources. Hydrogen sulfide is toxic and corrodes pipelines if untreated. Sour gas from the gas well is processed by separating the H_2S to give a sweet natural gas with specifications appropriate for transport and use by customers (maximum of 4 ppm H_2S).

Natural gas produced in lesser but not insignificant quantities or at remote locations is termed as stranded gas. These natural gas resources are typically vented or flared because of high transportation and processing costs. Processing stranded sour gas resources requires the development of scalable and easily deployable gas sweetening processes.

Modularization of chemical processes is a process intensification option that provides significant economic and safety benefits because of flexibility and robustness in the

transportation, deployment, and operation of a processing facility (Baldea *et al*., 2017; Bielenberg *et al*., 2019). Modularized desulfurization of stranded gas is highly beneficial for better utilization of natural gas and prevention of environmental emission.

In this work, three natural gas desulfurization processes, i.e., oxidative sulfur removal (OSR) SourCat^{TM}, iron-chelate redox systems LO-CAT®, and Triazine based scavenger (Hugo *et al*, 2000) are simulated for a natural gas feed at various scales (1 to hundred thousand MSCFD) and inlet H_2S concentrations (500 to 2,500 ppm). For each set of operating conditions, the desulfurization cost is calculated for each process. The calculated cost is depicted and analyzed through heat maps. The process with the lowest desulfurization cost is identified for each operating condition.

2. Process Simulations and Desulfurization Cost Calculations

2.1. Triazine-based scavenger process

Scavenger-based absorption is used hydrogen sulfide removal method for small sour gas cleaning. The process does not employ solvent regeneration and can be discarded to the environment as the spent liquid are readily biodegradable (Taylor *et al*., 2019). A process flow diagram (PFD) for the triazine-based absorption process is given in Figure 1. Sour gas (10-70°C) is injected into an absorber with a 52 wt% triazine-based scavenger, and the liquid-gas mixture is separated in a tower where sweet gas is collected from the top and spent liquid scavenger from the bottom (Lozano & Trauffer, 2000). To avoid solids formation, the inlet triazine is kept at 20% excess. Finally, the gas phase is maintained with sulfur content lower than 4 ppm, and the liquid phase is the reacted Triazine which will be disposed of.

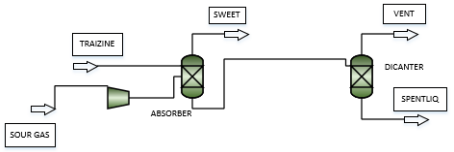


Figure 1. Triazine-based PFD

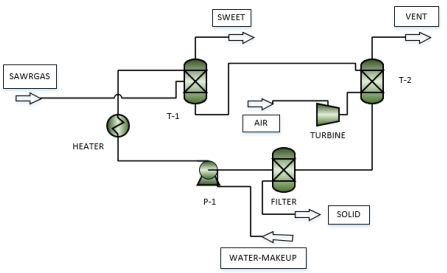


Figure 2. LOCAT® PFD

2.2. Liquid redox process

In the LO-CAT® process, H_2S is oxidized to elemental sulfur using a chelated iron sodium ferric ethylenediaminetetraacetate (EDTA FeNa) solution (Faramawy *et al*., 2016). This process is primarily used for small scale desulfurization of natural gas. The process flow diagram of the LO-CAT® process is given in Figure 2 (Faramawy *et al*., 2016; Speight, 2018). In the absorber vessel (T-1), H_2S from sour gas is absorbed by the EDTA solution, where ferric ions oxidize H_2S to form solid sulfur. Subsequently, in vessel T-2, ferrous ions are oxidized back to ferric ions when reacted with oxygen in air. Sweet gas with the required specification (<4 ppm H_2S) leaves the top of the absorber. The ferric ion solution is recycled back to the absorber. Fresh EDTA solution is added to the recycle loop to compensate for losses related to the wet sulfur cake. To prevent iron precipitation and accelerate H_2S absorption, NaOH or KOH is used to maintain the solution pH between 8 and 9.

2.3. Chemical oxidation (SourCat^{TM}) process

The SourCat^{TM} process is a newly patented solid catalyst-based oxidative sulfur removal (OSR) desulfurization process developed by IntraMicron, Inc., an Auburn University spin-off company (Deng *et al*., 2020). A process flow diagram of the simulation is given in Figure 3. Sour gas and air are injected into the solid fixed-bed catalytic reactor, where H_2S is converted to solid sulfur and a small amount of SO_2 with high selectivity to elemental sulfur (over 90%). The reactor effluent is washed and cooled with water to remove solid elemental sulfur and most SO_2. The solid sulfur is collected by a filter and discarded as sulfur cake. The small amount of SO_2 produced is carried by the gas stream and separated from the sweet gas by an adsorption bed (V-2).

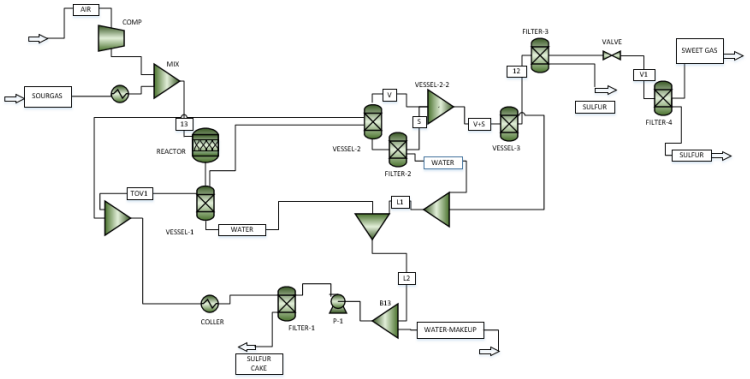


Figure 3: SourCat^{TM} PFD

2.4. Cost calculation

Desulfurization costs are defined as the cost of sweetening one thousand standard cubic feet (MSCF) natural gas. Aspen Plus^{TM} process simulations are used to calculate the raw material/utility requirements and equipment sizes for each process. Given this information, the capital expenses (CAPEX) and operating expenses (OPEX) are calculated using standard cost calculation coefficients (Turton *et al*., 2008). Next, the desulfurization cost before and after paying back the capital investment is calculated using Equations (1) and (2), respectively. In Eq. (1), the annual interest rate, *r*, is assumed at 10%, and the payoff time, *t*, is set as 5 years. The variable capacity represents the sour gas flow rate in thousand standard cubic feet per day (MSCFD).

$$Desulfurization\ Cost\ before\ payback\ = \frac{CAPEX \times \frac{r(1+r)^t}{(1+r)^t - 1} + OPEX}{Capacity} \quad (1)$$

$$Desulfurization\ Cost\ after\ payback\ = \frac{OPEX}{Capacity} \quad (2)$$

2.5. Modularization criteria

In this work, a simple modularization criterion is used to limit the size of principal equipment in each process to allow for easy transportation. In sizing the conventional process, each piece of equipment was scaled freely according to the process requirements. On the other hand, to identify the size of the modularized plants, we constrained the size of the equipment based on federal transportation guidelines (Morris, 2003). The sizing algorithm for the modularization is illustrated in Figure 4 (Shao *et al.*, 2020).

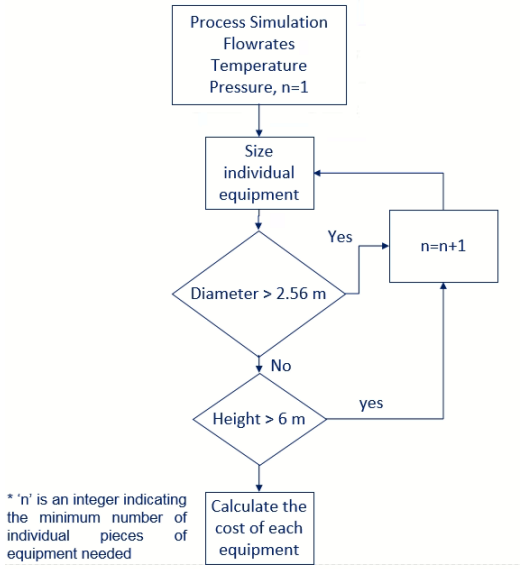


Figure 4. Method and criteria used for modularization.

3. Results & Discussion

After the cost calculations to compare three different desulfurization processes, a plot is generated comparing desulfurization cost before and after the payback of the equipment. In Figure 7 (left), the cost before payback is shown, and Figure 7 (right) shows after payback. Both plots show the cost for SourCat™ process is minimum at ten thousand MSCFD, for before and after payback time, respectively.

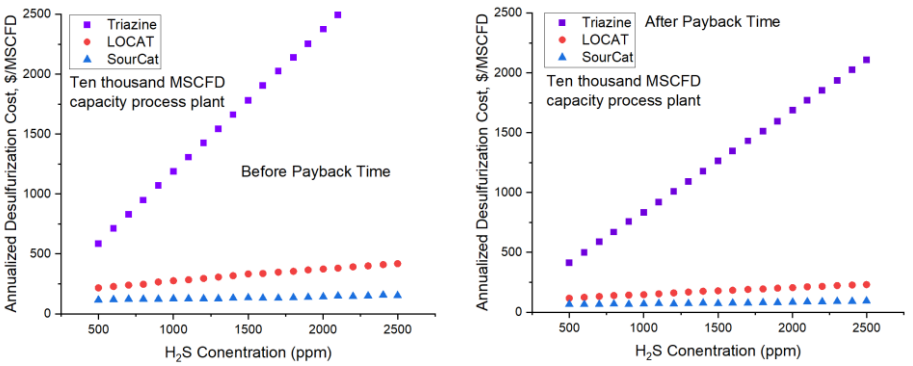


Figure 7: Desulfurization cost at 10,000 MSCFD before (left) and after (right) payback.

In Figure 8 (left), the cost before payback is shown, and Figure 8 (right) shows after payback for 100,000 MSCFD. Both plots show that the cost of the SourCat™ process is the lowest at 100,000 MSCFD, both before and after payback. LOCAT could only compete at very low H_2S concentration. The Triazine-based process cost has been excluded from the plots as it is significantly more expensive than the others at 100,000 MSCFD.

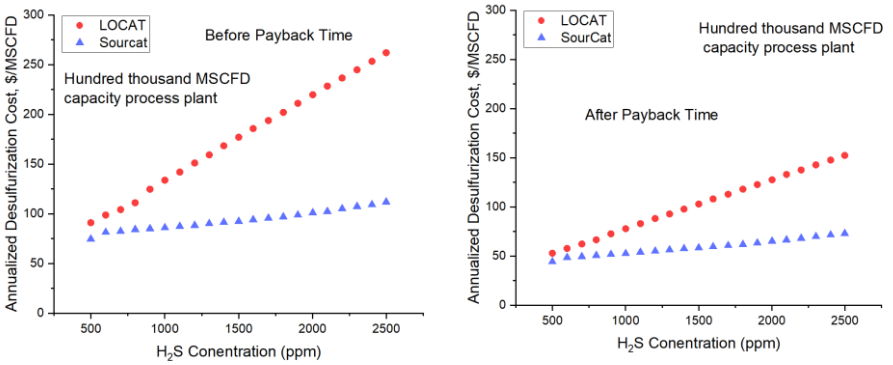


Figure 8: Desulfurization cost at 100,000 MSCFD before (left) and after (right) payback.

Based on the cost calculation results, heat maps illustrating the competing technologies for different combinations of flowrate and H_2S concentration have been developed and shown in Figure 9. Both plots indicate that at higher natural gas flowrates, SourCat™ may be the most cost-effective technology, but at very low natural gas flowrates and particularly at lower hydrogen sulfide concentrations, the Triazine-based scavenger process can be cost-effective.

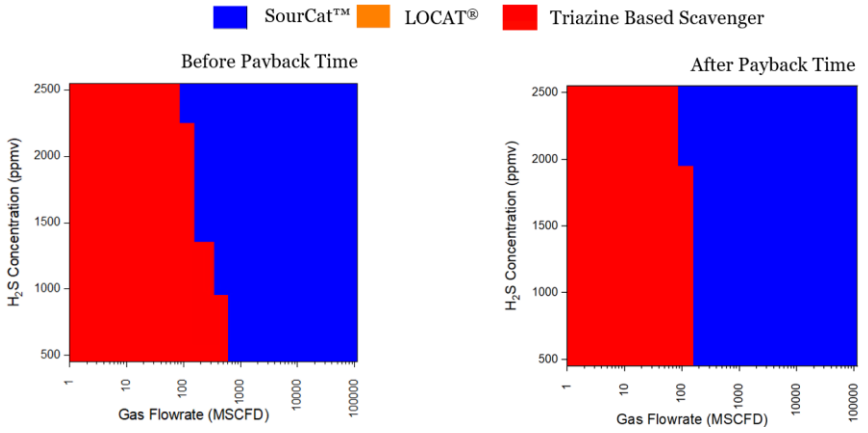


Figure 9: Desulfurization cost heat map.

Similarly, heat maps illustrating the comparison of the three technologies for the modularized case have been developed and are shown in Figure 10. It shows that at high flow rates of 100,000 MSCFD and low H_2S concentration, LOCAT can be cost-competitive, however for most cases investigated here, the best choices are Triazine for low flow rates, and SourCat™ for higher flowrates (>100 MSCFD).

4. Conclusions

In this paper, we have investigated three natural gas desulfurization technologies at a range of gas flowrates and hydrogen sulphide concentrations. Conventional designs, as well as modularized configurations, have been compared. For all combinations of gas flowrate and H_2S concentration, the technology with the lowest sulphur removal cost has been identified. The new SourCat™ process appears to be the most cost-effective and

economically favorable process for flowrates above 100 MSCFD regardless of H_2S concentration, while the Triazine-based Scavenger process is more suitable for small capacity/lower flowrate systems. The LOCAT® process only shows promise at a very high flowrates and very low H_2S concentration, even in the modularized configuration. Therefore, while the Triazine-based process has potential for very small systems, the flexibility and scalability of the SourCat™ process lends itself more to process intensification and deployment at medium to larger size systems.

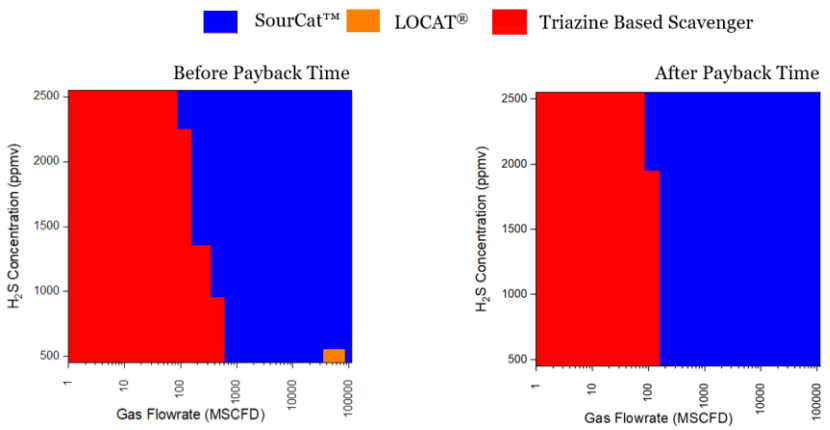


Figure 10. Desulfurization cost heat map for modularized process.

References

Baldea, M., Edgar, T. F., Stanley, B. L., & Kiss, A. A. (2017). Modular manufacturing processes: Status, challenges, and opportunities. AIChE journal, 63(10), 4262-4272.

Bielenberg, J., & Palou-Rivera, I. (2019). The RAPID Manufacturing Institute–Reenergizing US efforts in process intensification and modular chemical processing. Chemical Engineering and Processing-Process Intensification, 138, 49-54.

Deng, Y., Xu, S., Webb, K., Wright, H., Dimick, P. S., Cremaschi, S., & Eden, M. R. (2020). Sensitivity Analysis of Desulfurization Costs for Small-Scale Natural Gas Sweetening Units. In Computer Aided Chemical Engineering (Vol. 48, pp. 973-978). Elsevier.

Faramawy, S., Zaki, T., & Sakr, A. E. (2016). Natural gas origin, composition, and processing: A review. Journal of Natural Gas Science and Engineering, 34, 34-54.

Lozano, H., & Trauffer, E. A. (2000). Re-generable H_2S Scavenger. In Proceedings of the Laurance Reid Gas Conditioning Conference, pp. 47-78.

Morris, J. R. (2003). TRB Special Report: Regulation of Weights, Lengths, and Widths of Commercial Motor Vehicles. TR News, (228).

Speight, J. G. (2018). Natural gas: a basic handbook. Gulf Professional Publishing.

Taylor, G., Smith-Gonzalez, M., Wylde, J., & Oliveira, A. P. (2019, March). H2S Scavenger Development During the Oil and Gas Industry Search for an MEA Triazine Replacement in Hydrogen Sulfide Mitigation and Enhanced Monitoring Techniques Employed During Their Evaluation. In SPE International Conference on Oilfield Chemistry. OnePetro.

Turton, R., Bailie, R. C., Whiting, W. B., & Shaeiwitz, J. A. (2008). Analysis, synthesis and design of chemical processes. Pearson Education.

Shao, Y., & Zavala, V. M. (2020). Modularity measures: Concepts, computation, and applications to manufacturing systems. AIChE Journal, 66(6), e16965.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50169-X

Process Alternatives for the Co-Production of Hydrogen and Methanol using Fuel Switch and Energy Mix Systems

Usama Ahmed [a,b], Umer Zahid [a, c], Nabeel Ahmad [d], Nauman Ahmad [e]

[a] *Chemical Engineering Department, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia*
[b] *Interdisciplinary Research Center for Hydrogen and Energy Storage, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia*
[c] *Interdisciplinary Research Center for Membranes & Water Security, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia*
[d] *Department of chemical engineering, COMSATS University Islamabad, Lahore campus 54000, Pakistan*
[e] *Minhaj Univeristy Lahore, 54000, Pakistan*
usama.ahmed@kfupm.edu.sa

Abstract

With the increase in the global energy requirements, the utilization of fossil fuels have been also increased which not only increased greenhouse gas (GHG) emissions but also caused global warming. The sharp reduction in the GHG emissions demands for utilizing the energy mix and fuel switch technologies to develop the cleaner fuels. In this study, process simulation framework has been proposed to simultaneously produce two fuels (methanol and $H_{2)}$ in the co-generation process integration framework. Two process models have been developed in Aspen Plus ® V12 followed by the techno-economic assessment to analyze the process feasibility. Case 1 is considered as the base case process, where the process initiates with the coal-biomass gasification to produce the synthesis gas. The synthesis gas is then treated in the acid gas removal units to remove the CO_2 and H_2S. The synthesis gas is then fed to the methanol synthesis unit to produce methanol, whereas, the unconverted CO is treated in the water-gas shift (WGS) reactors to convert CO into the H_2 and CO_2 followed by CO_2 removal in the AGR section. Case 2 design represents the novel process for co-production of methanol and H_2, where the coal-biomass gasification technology is sequentially integrated with the methane reforming technology to maximize the heat utilization without any energy penalties. Both the models are techno-economically compared in terms of methanol and H_2 production rates, specific energy requirements, carbon conversion, CO_2 specific emissions, overall process efficiencies, fuel production costs and project feasibilities. The results showed that the case 2 design offers higher methanol and H_2 production rates with the low energy requirements. Also, the process efficiency of case 2 is 3.8% higher than the case 1 design, where, the specific energy requirements of case 2 is almost 16% less compared to the case 1 design. While performing some sensitivity studies, it has been analyzed that the case 2 process offers better process feasibility in terms of process performance and economics.

Keywords: gasification, reforming, hydrogen, methanol.

1. Introduction

The global energy demand has increased manifolds during the last few decades due to rapid industrialization and continuously relying on the conventional processes for heat

and power generation. Also, the recent report issued by Global Energy Outlook (GEO, 2020) also forecasted an increase in energy demand up to 30% in the near future. To minimize the difference between energy demand and supply, conventional and economical fuels have been used for power generation, where, coal and natural gas are among the most widely used feedstock's. However, the recent increase in the fuel prices has put more pressure to use the cleaner fuels and reduce the carbon emissions. Since, the coal fee stocks are widely available and most of the infrastructure is developed based on the coal based processes, it is expected that the coal will be used in the coming decades. Therefore, an attention has been made to convert the coal into clear fuels i.e. syngas, H_2, methanol, FT chemicals, etc.

Natural gas to methanol and hydrogen (NGTMH) technology has been widely used to produce hydrogen and methanol. Steam methane reforming (SMR) is an intermediate step to convert the natural gas into syngas with the HCR of 2.0-2.5. The syngas can be further transformed into various products including H2, methanol, and FT chemicals. On the other hand, hydrogen and methanol can be produced from the coal by converting it into syngas using gasification techniques, which, can be further transformed into methanol and H_2. Coal to methanol and hydrogen (CBH) processes have received a lot of attention due to lower coal prices, however, hydrogen to carbon (HCR) from the coal-based processes lies in the range of 1:2, which is not suitable for methanol production. Therefore, water gas shift (WGS) and CO_2 removal systems are deployed to enhance the H_2 production and minimizing the CO_2 emissions. On the other hand, an integrated approach to utilize the natural gas reforming and coal gasification is the single process can be also used to enhance the overall hydrogen production. This approach utilizes the heat from the gasification unit in the natural gas reforming unit to minimize the overall energy needs while producing additional syngas. Mixing the gasifier and reformer derived syngas; the overall HCR can be increased to make it suitable for the methanol and hydrogen production. Yi et al. (2015) reported that the 2.6 ton of CO2 is emitted for each ton of methanol production while using the CTM technologies. On the other hand, SMR processes showed a potential of lower carbon emissions due the higher HCR ratio in the synthesis gas. Recently, Blumberg et al. (2019) also performed the exergy analysis and developed the alternative designs by integrating different reforming technologies to enhance the H_2 production that can improve the overall methanol production. Ahmed et al. (2021, 2021, and 2021) also integrated the syngas production processes to enhance the overall H_2 production rates that can influence the downstream methanol production.

This study utilizes coal and biomass feedstocks, which are gasified to generate the syngas. Moreover, the gasification unit is integrated with the methane-reforming unit, where, the heat from the gasification unit is utilized in the reforming unit. The key idea is to use the multiple feedstocks that can involve the biomass to reduce the overall carbon footprint, while, enhancing the syngas production. The technical and economic analysis is performed to determine the overall process efficiency, carbon emissions fuel production costs.

2. Modelling and Simulation

Aspen Plus (V12) has been used in this study to develop the simulation models for the production of methanol and hydrogen. The Peng Robinson with Boston Mathias (PR–BM) is used as the effective thermodynamic package, where, coal and biomass are defined based on proximate and ultimate analysis. Table 1 represents the process conditions and some of the main units of the overall process:

Process	Conditions
Reformer (RGibbs)	Temperature: 900°C, Pressure: 32 bar NG flow rate: 5.50kg/s, H_2O:CH_4 = 3:1, Ni-based catalyst
Gasification (RGibbs)	Temperature: 1350-1370°C, Pressure: 56 bar Coal Flow Rate: 62.2kg/s Biomass flow rate: 6.2kg/s
Methanol Reactor(RGibbs)	Temp/Pressure = 200°C / 5.5 MPa
Heat Exchangers (MHeatX)	ΔTmin=10°C
Water Gas Shift Rector (REquil Reactor)	Adiabatic reactors, Steam/CO: ~2.2

3. Process Description

Two process models have been developed in Aspen Plus (V12). Figure 1 shows the base case model, where, coal and biomass are used to produce syngas in the gasification unit.

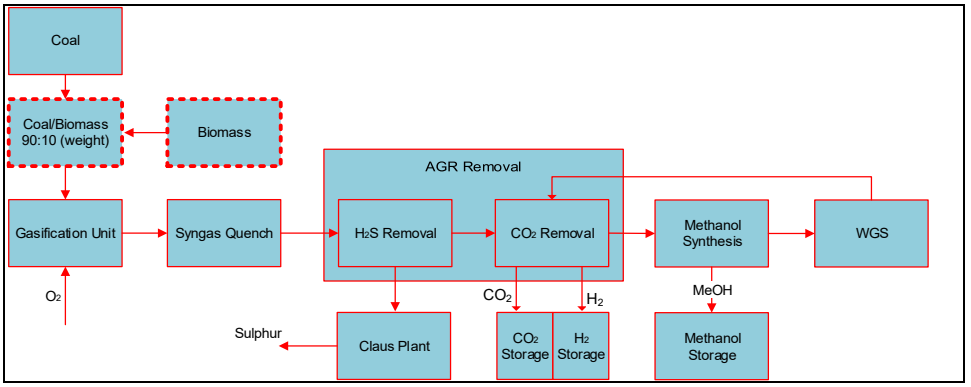


Figure 1: Dual Methanol and H_2 production from Coal and biomass – Case 1

The temperature of the syngas is then reduced in the radiant coolers to generate high-pressure steam. The syngas is then passed through the acid gas removal unit to remove the H_2S and CO_2 from the syngas. The synthesis gas is then allowed to enter in the methanol synthesis unit, in the presence of copper-based $CuO/Al_2O_3/ZnO$ catalyst, to produce the methanol. The unconverted syngas from the methanol unit is then fed to the WGS unit to convert the CO in the syngas to CO_2 and H_2. The mixture of CO_2 and H_2 is then fed to the CO_2 removal unit to capture excess of the CO_2 and to get the pure H_2, which is sent to the storage section. Figure 2 represents the case 2 process model, which uses biomass, natural gas and coal feedstocks for the production of methanol and H_2. Case 2 mainly integrates the coal/biomass gasification process with the natural gas reforming technology, where, the energy from the gasification unit is utilized in the reforming unit. The overall aim of this integration is to minimize the energy needs of the process while increase the production of syngas and H_2.

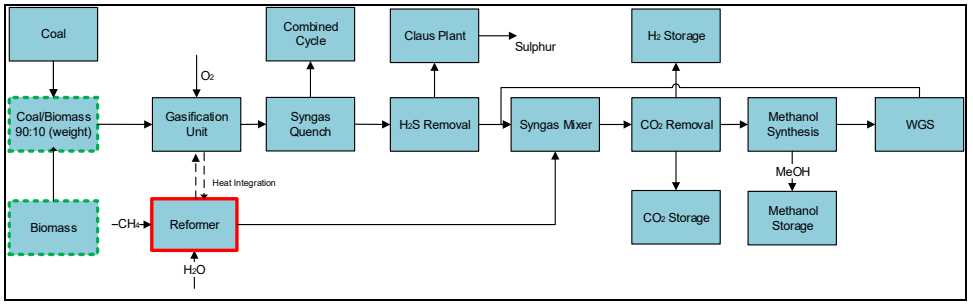


Figure 2: Production of H_2 and Methanol from Coal, Biomass and Methane in an integrated process – Case 2

4. Results and Discussion

4.1 Methanol and H_2 Production Rates and Overall Process Performance

The results from the case 1 and case 2 are compared to analyze the overall process efficiency and to evaluate the production rates of the methanol and H_2 from both cases. The results showed that the case 2 design has higher syngas production rates with the higher HCR in the syngas. The results are also compared in terms of higher heating value (HHV) of the syngas, where, the HHV of the syngas in case 1 and case 2 is calculated as 16.24MJ/kg and 18.84MJ/kg, respectively. The production rates of methanol for case 1 and case 2 have been calculated as 31.17kg/s and 45.44kg/s, respectively. Similarly, the H_2 production rates for case 1 and case 2 have been calculated as 3.1kg/s and 2.58kg/s, respectively, as shown in Figure 3.

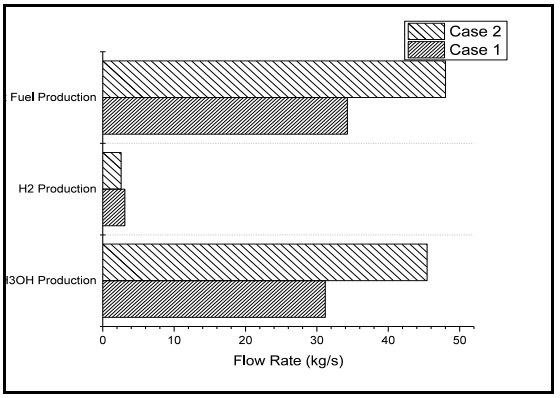


Figure 3: Methanol and Hydrogen Production Rates

The specific energy requirement is an important indicator to find out the energy required for each unit of product formed. It has been seen from the results that methanol specific energy requirement for case 1 and case 2 is 20.51kg/GJ and 26.15 kg/GJ. Also, the H_2 specific energy requirement for case 1 and case 2 is 2.04 kg/GJ and 1.48kg/GJ, respectively. It is also important to estimate the overall process performance of case 1 and case 2 for comparative analysis. The efficiency for both the cases have been calculated based on thermal energy of feedstocks and produced fuels (methanol and H_2) as represented in the equation 1.

$$\text{Process Efficiency} = \frac{\text{Methanol thermal energy [MWth]} + \text{H}_2 \text{ thermal energy [MWth]}}{\text{Feed stock thermal energy [MWth]} + \text{Energy consumed [MWth]}} \text{x } 100\% \quad (1)$$

The results showed that the overall process efficiency of case 1 and case 2 is 65.3% and 69.9% respectively. Moreover, it has been seen from results that the specific production energy requirement for case 1 and case 2 is 44.33GJ/MT and 36.18GJ/MT of energy, respectively. Moreover, the carbon conversion efficiency for case 1 and case 2 is calculated around 30% and 40%, respectively. In terms of emissions, case 1 and case 2 generated 0.31 kg and 0.19 kg of uncaptured CO_2 for each kg of fuel (methanol and H_2) produced.

5. Economic Analysis and Project Feasibility

Economic analysis includes the estimation of the capital expenditure (CAPEX) and the operational expenditure (OPEX) to estimate the cost of the final product. The CAPEX mainly involves the cost the equipment, piping and installation, instrumentation, civil work etc. On the other hand, OPEX is calculated on the base of cost of raw material or feedstocks, utilities, catalyst replacement, maintenance, etc. The CAPEX is calculated by sizing all the equipment depending upon the operational conditions and the flow rates of the stream through each unit. The CAPEX is calculated using equation 2, where the $CEPCI_{old}$, $Capacity_{old}$ and $Cost_{old}$ values are taken from the literature for the corresponding units along and the new CEPCI of 618 for the year 2021 has been chosen to estimate the cost of new units along with new production capacities taken from the model. The value of x is taken as 0.6 ($6/10^{th}$ rule) for keeping the comparative analysis simple.

$$\text{Cost}_{\text{New}} = \text{Cost}_{\text{Old}} \times \left(\frac{\text{Capacity}_{\text{New}}}{\text{Capacity}_{\text{Old}}}\right)^{\text{x}} \times \frac{\text{CEPCI}_{\text{New}}}{\text{CEPCI}_{\text{Old}}} \quad (2)$$

The results from the economic analysis is represented in the Table 2. The results showed that CAPEX/MT required for case 1 and case 2 is 75.37 $/MT and 66.51$/MT, respectively. Moreover, the OPEX/MT for the case 1 and case 2 is calculated as 250.82 $/MT and 217.02 $/MT, respectively. The CAPEX and OPEX calculations are used to determine the total production cost (TPC) and minimum selling price (MSP). The results showed that the selling price of the methanol for case 1 and case 2 is estimated as 322.73 $/MT and 256.93 $/MT, respectively. Moreover, the selling price of the H_2 for case 1 and case 2 is estimated as 36.09 $/MT and 54.96 $/MT, respectively.

Table 2: Economic Analysis

	Units	**Case 1**	**Case2**
CAPEX	USD/MT	75.37	66.51
OPEX	USD /MT	250.82	217.02
Total (CAPEX+OPEX)	USD /MT	326.20	283.53
10% Profit Rate	USD /MT	32.62	28.35
MSP od Fuel (H_2+CH_3OH)	USD /MT	358.82	311.88
MSP of Methanol	USD /MT	322.73	256.93
MSP of H_2	USD /MT	36.09	54.96

The cash flow analysis for both the cases have been also done to analyze the project feasibility, where, the lifetime of the plant in both cases is taken as thirty three (33) years as represented in the Figure 4. The results showed that the case 2 design offers higher rate on return on the investment as compared to the case 1 design.

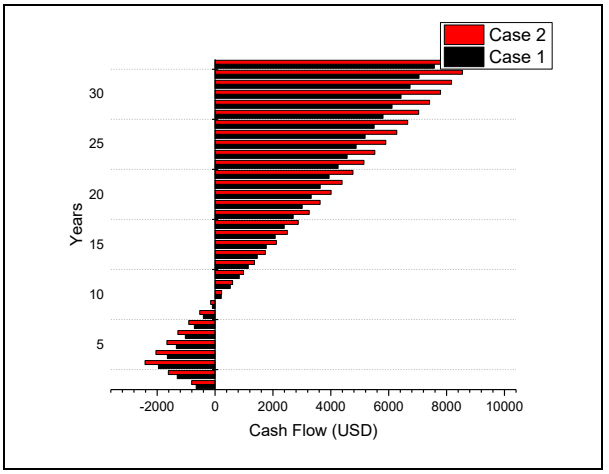


Figure 4: Cumulative Cash Flow Analysis

6. Conclusions

This study represented the technical and economic analysis of two process models for the production of hydrogen and methanol. Both the models were developed in Aspen Plus, where, the base case model used coal and biomass for the production of hydrogen and methanol. On the other hand, the alternative model used the natural gas along with the coal and biomass, where, the sequential integration between the gasification unit and reforming unit is done to minimize the energy needs. The two models were compared in terms of process efficiencies, fuel production rates and carbon conversion. The results showed that the case 2 performs better compared to the case 1 design in terms of both process performance and economics.

Acknowledgements

The authors would like to acknowledge the support provided by the Deanship of Scientific Research (DSR) at King Fahd University of Petroleum & Minerals (KFUPM) for funding this work through project No. SB201019.

References

Ahmed, U., Zahid, U., Onaizi, S. A., Abdul Jameel, A. G., Ahmad, N., Ahmad, N., & AlMohamadi, H. (2021). Co-production of hydrogen and methanol using fuel mix systems: Technical and economic assessment. Applied Sciences, 11(14), 6577.

Ahmed, U., Hussain, M. A., Bilal, M., Zeb, H., Zahid, U., Onaizi, S. A., & Jameel, A. G. A. (2021). Utilization of Low-Rank Coals for Producing Syngas to Meet the Future Energy Needs: Technical and Economic Analysis. Sustainability, 13(19), 10724.

Ahmed, U. (2021). Techno-economic analysis of dual methanol and hydrogen production using energy mix systems with CO2 capture. Energy Conversion and Management, 228, 113663.

Blumberg, T., Morosuk, T., & Tsatsaronis, G. (2019). CO2-utilization in the synthesis of methanol: Potential analysis and exergetic assessment. Energy, 175, 730-744.

Newell, R., Raimi, D., Villanueva, S., & Prest, B. (2020). Global energy outlook 2020: Energy transition or energy addition?

Yi, Q., Li, W., Feng, J., & Xie, K. (2015). Carbon cycle in advanced coal chemical engineering. Chemical Society Reviews, 44(15), 5409-5445.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50170-6

Documenting Models Comprehensively Using a Minimal Graphical Language

Heinz A Preisig

Department of Chemical Engineering, NTNU, N-7491 Trondheim, Norway
Heinz.Preisig@chemeng.ntnu.no

Abstract

A small graphical language provides the means to document and discuss process models in details without engaging into a programming environment. It proved to be a powerful tool to discuss model on the back of an envelope as well as for defining a graphical user interface for ontology-based modelling suite ProMo.

Keywords: Modelling ontologies, simulation, control, design, operations

1.1 Background

Mathematical models are the core of any computational engineering activity and models are used in various contexts. Traditionally modelling tools attempt to construct process models from basic building blocks. The nature of the building blocks is often not transparent to the user. The tools are constructed with expecting only basic knowledge of the basic building blocks. The actual realisation is usually left to an expert and remains mostly hidden from the process model builder. This approach works fine as long as the building blocks are matching the behaviour the user expects. If the match is not good enough, the user has no hands on learning about the detailed implementation and the expert has to be called on the scene.

Independent on the type of user, it is good practice to provide a complete documentation of the building blocks and the resulting process model. It is often essential to get the insight on what the person constructing the model assumed and how it was "seen". Modelling tools are also mostly integrated with a solver environment, which in the past often used to be visible and accessible. Today, a myriad of solvers are available and it is often desirable to move from one solver environment to an other, which brings the issue of generating code compatible into the limelight. One quickly also gets interested using models with different solvers and thus one formalised the problem of transferring models between different environments. The CAPE-OPEN project (2020) is the most outstanding in chemical engineering, while the Modelica community designed functional mock-up interfaces (FIM 2021). In both cases, one retains the actual code and resorts to wrapper technology (for example SWIG 2021). We have taken a different approach by lifting model documentation and behaviour to a higher level.

2. Foundation: Reductionism

Chemical engineering software builds on implementing basic principles of physics, chemistry and biology, enriched, when necessary, with black-box models. Latter to capture behaviours that can for one or the other reason not be captured by a mechanistic description. Either it is too complicated, which is probably the most common cause, or

it is merely not sufficiently well known what is going on between the excitation and the response of the part in question.

Most simulators have a graphical input language, just because it is easier for people to generate models using a graphical interface in contrast to a textual interface. So why then a graphical language for documenting/describing models?

Graphical tools for capturing a plant's behaviour may or may not have models attached to the visual object representing them. Some systems are purely focusing on a graphical representation (example: ConceptDraw, 2021). Simulators, in contrast, do have models linked to the items shown on the screen. Examples are gProms™ (2021), Aspen™ (2021), etc., but also Matlab's Simulink to mention a product with another background.

Flowsheet simulators require experts to build the building blocks. The blocks typically describe the behaviour of a process entity, like a unit operation. The individual blocks are hand written pieces of code including all the necessary mathematical relations that describe their behaviour. Thus these facilities do include mathematical models, but the new development of constructing models based on ontologies, opens a whole new approach.

Ontology-based model construction has the flavour of the system theory's concept of minimal realisation. It aims, and succeeds, in defining based entities from which models are being constructed. The approach is based on using reductionism. Reductionism is a commonly applied technology for analysing and describing objects. It takes the object of interest and recursively subdivides it into smaller and smaller pieces until a granularity is reached that is rich enough to represent the process. We use reductionism to define the smallest domain-specific entities required to represent the considered classes of processes. The base entities are captured in an ontology, which is used to define the mathematical behaviour of the base entities. A small language serves to define the equations, and the analysing parser is combined with a template machine generating different output codes suitable for different programming environments.

Once the base entities are associated with a graphical representation, one can use them in a visual tool to construct process models on a canvas. This process we have described before {references}.

The graphical language can, though, also be used as a manual design tool. Choosing the base entities carefully, keeping in mind a minimal set of objects for the given application domain, one generates a powerful graphical language that serves a multitude of purposes: Paper and pencil model design, model documentation, and graphical input to ProMo's model composer software.

The graphical language captures frame information, thus time and space. The time scale characteristics, being constant, dynamic and event-dynamic and the space characteristics being the distribution nature: 0D for lumped systems, where the relevant intensive properties are not a function of the spatial coordinates. 1-3 D where they are a function of the named number of spatial coordinates. Defining a set of arcs depicting the interactions between the base entities, captures how they interact and what is being transported.

3. Language as paper-and-pencil tool and documentation

We shall first use the language to demonstrate the use of the graphical language as a discussion and documentation tool.

3.1. Example espresso maker

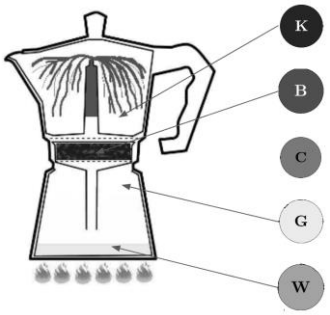


Figure 1: A typical Italian espresso maker

The traditional traditional espresso maker has a lower part, where the water is being filled in after unscrewing the upper part. In the middle is a filter section, where the coffee powder is being added. The filter section is connected to the fountain of the top section when screwing the top part onto the lower.

When heating at the bottom, the water temperature rises until it the pressure in the lower gas phase is large enough to push the hot water up the immersed pipe into the ground coffee and further up to the fountain to the top coffee reservoir. Once the water level is low and the upwards pipe is not anymore in the water, steam is passing through the pipe, the filter cake and spits out the remaining fluid through the fountain. Voila – espresso is ready for consumption.

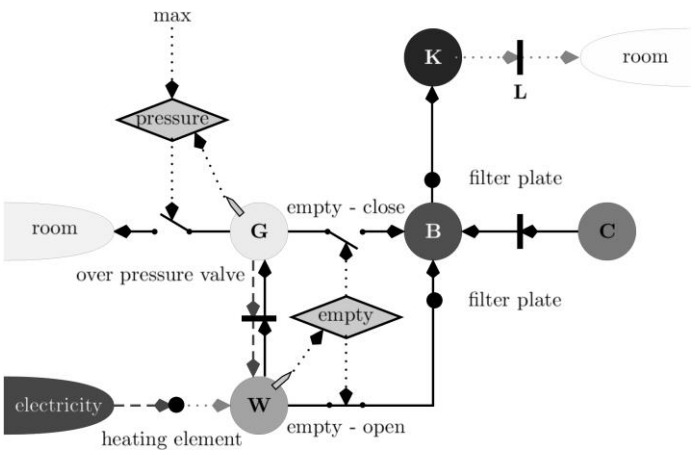


Figure 2: A possible topology for the espresso maker

The topology shows the room and the electricity source as reservoirs. The gas phase (G), the water (W), the fluid in the filter container (B), the ground coffee (C), and the product (K) are shown as lumped systems. The operation associated with having first hot water and in the end steam passing upwards is shown as a control element with two switches. This is a typical example, where the MODEL is controlled, in contrast where there is a physical controller acting on some flow. The over pressure valve is a physical element and can be seen as a controller that is built into the valve. We use circles for lumped systems, and semi-open olives for reservoirs. The arrows stand for the transfer of extensive quantity, like mass or energy in one or the other form. Phase interfaces appear as bars. For the control we introduce an observer picking information from the inside of a system and the rhomboids are control decision elements. Control signals are dotted lines. Arrows are in all cases defining a reference coordinate for the represented flow.

The direction of the arrows define a reference coordinate system for the flow. Flow in the direction of the arrow is positive, while the flow in the opposite direction is negative. The full line arrows show mass flow. Notice that the coffee being leached is only one arrow, though the water is flowing in the negative direction and the leached product into the positive direction. The dashed arrows are associated with work. The arrow from the electricity reservoir to the heating element is an electrical work flow, while the same type of arrows between the gas phase, the interface and the water stand for volumetric work, as the gas phase increases in volume while the liquid phase shrinks. The heating element is shown as and ideal converter of electrical energy into heat, latter flowing into the water. Some heat is lost through the lid into the room. The capacity effect of the construction is neglected, as no element is shown in the topology that represents the construction.

3.2 Example: A melting process

Melting of a solid in a heated device is characterised by going through three stages: (i) heat the solid (ii) melt the solid (iii) heat the liquid. The first and last stage have only one phase, while the second stage is characterised by having two phases. The model is being switched with the triggered by the formation of a first liquid phase and the completed disappearance of the solid phase.

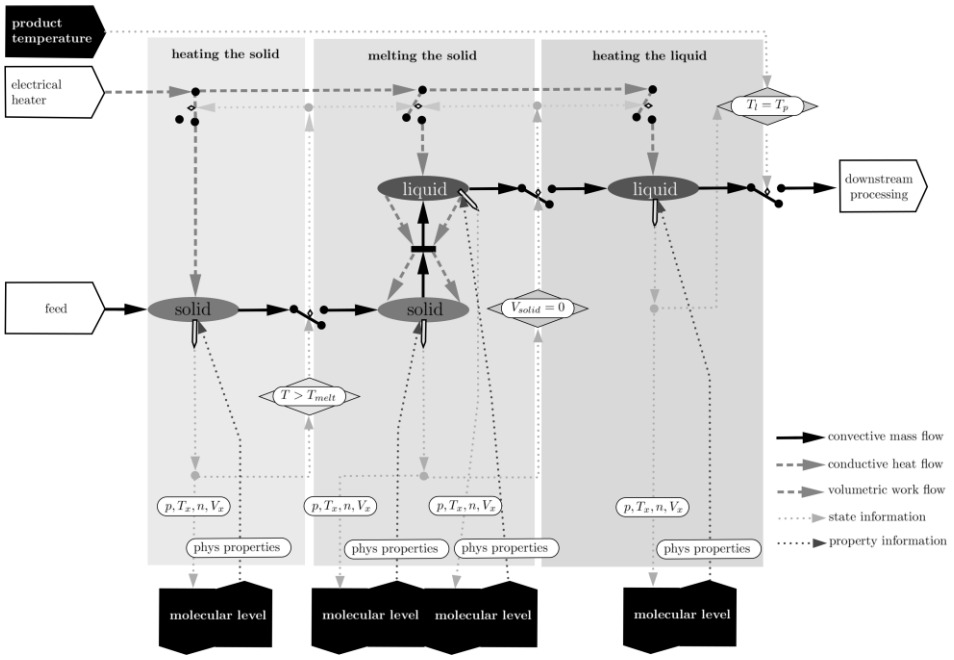


Figure 3: The three stages of melting a solid

The melting of solid topology has three control stuctures. Two are controlling the model, namely the two triggering the switching between the stages and one that is a physical controller opening the outflow tap once the desired temperature in the liquid is achieved. Notice that the model will not work if the product temperature is below the melting temperature. The ontology models the two phases as distributed systems, thus assuming that the temperature is a function of the position within the respective phase.

Notice that that the graphical elements are enriched with olives for the representation of distributed systems and connectors to other parts of the model.

3.3. Example molecular modelling

The top-level of the molecular-level model has a couple of interesting control structures.

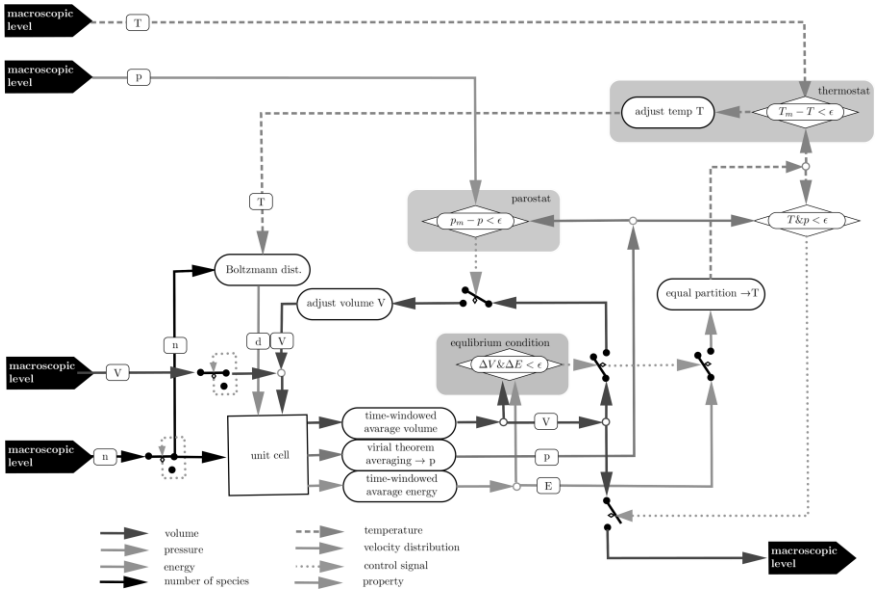


Figure 4: Molecular modelling level of properties

The overall process is iterative. The state of the molecules in the unit cell is adjusted by changing the temperature and the volume until the temperature and the pressure is close enough to the required temperature and pressure. The two involved control loops include a thermostate and a barostate, which are triggered by the check on the volume and the energy change over the iteration period. On the molecular level, we essentially only deal with information flows. We thus carefully mark the individual signals.

The graphical model contains all the information except than the specification details of the unit cell. While it is possible to provide the graphical model for the unit cell, it is not very instructive. One better resorts to a modified version in which the number of molecules is provided as well as the type of interacting forces and the associated assumptions and those forces that are neglected.

3.4. Corrosion – a moving boundary process

An interesting class of models deal with moving boundaries. As an example we show a corrosion process in which the iron bars in concrete are attacked by an active component forming a rust layer. The reaction takes place in the boundary volume on top of the iron. It is modelled as a infinite small volume. The interfaces control the species transfer. So W, A, R are not transported into the iron, for example. So on the top we show the presence of the species while on the lower row, we indicate, which species are not transferred. Thus we define semipermeable walls representing the interfaces. The

circles again represent lumped systems, here iron. The bars are interfaces with the indication of what group of species is not transferred. The dot system is an infinite small capacity in which the reaction takes place.

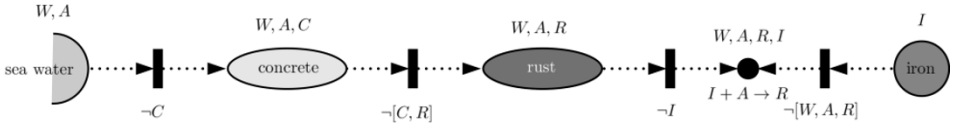


Figure 5: A corrosion process. The active component A and water W is diffusing through the concrete and the corrosion product layer (rust) to react with the substrate

1.2 Conclusions

We present a graphical method for representing process models on different scales. On the macroscopic level we have capacities on the three time scales: constant, dynamic and event-dynamic. The dynamic and the event-dynamic capacities are further detailed with the distribution information (1D-3D) and interfaces are abstracted to semi-permeable walls. An important aspect is the separation of the control of the plant, thus a physical object, and the control of the model. Latter implements some logic operations into the model, like a state limit is reached and the model changes structure like in the melting process. The result is always a directed graph showing the physical part of the process, the model control structure and the physical control structure. The directions provide the reference coordinate for each flow, which in the physical process are always extensive quantities.

Acknowledgments:
ProMo research was funded in parts by: (i) Bio4Fuels Research Council of Norway (RCN) project 257622 (ii) MARKETPLACE H2020-NMBP-25-2017 Grant Agreement No 760173 (iii) VIPCOAT Horizon2020 Grant Agreement No 952903 (iv) MODENA FP7-Framework Grant Agreement ID: 604271

References

Aspen, 2021, https://www.aspentech.com/enproducts/engineering/aspen-plus (last visited 27-10-2022)
Cape-Open, 2021, http://colan.org (last visited 27-10-2022)
ConceptDraw, 2021, https://www.conceptdraw.com/ (last visited 27-10-2022)
Elve, A. T. & Preisig, H. A., 2018, A framework for multi-network modelling, *Computer Aided Chemical Engineering, 44*, 2215ff
Functional Mock-up Interfaces, http://fmi-standard.org (last visited 27-10-2022)
gproms, 2021, https://www.psenterprise/.com/products/gproms (last visited 27-10-2022)
Preisig, H. A., 2021, Ontology-Based Process Modelling-with Examples of Physical Topologies *Processes, 9*
Preisig, H. A., 2020, ProMo -- A Multi-disciplinary Process Modelling Suite, *Computer-Aided Chemical Engineering,* 571-576
MatLab / Simulink, 2021, https://www.mathworks.com/products/simulink.html (last visitied 27-10-2022)
SWIG, 2021, http://www.swig.org (last visited 27-10-2022)

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50171-8

Simulation and CO_2 emission analysis for co-processing of bio-oil and vacuum gas oil

Jingyu Zhang, Shuai Zhang, Le Wu*, Yuqi Wang, Lan Zheng

School of Chemical Engineering, Northwest University, Xi'an 710069, China
lewu@nwu.edu.cn

Abstract

To reduce the CO_2 emission of a refinery and lower the production cost of bio-fuels, the co-processing of vacuum gas oil (VGO) and bio-oil has been proposed to add the renewable carbon into the co-processing products. However, the accurate CO_2 emission of the co-processing scheme is still unclear. In this work, a simulation of the whole co-processing scheme is built in Aspen Plus according to the date from actual industrial operations and reports. Based on the simulation results, the total CO_2 emission consisting of the direct emission from fuel combustions and indirect emission from utility consumptions is calculated to give a fully understand of co-pressing scheme. This work not only provides a whole simulation model, but also gives a basic understanding of the CO_2 emission and optimization for the co-processing scheme.

Keywords: Co-processing; Bio-oil; Vacuum gas oil; Aspen; CO_2 emission

1. Introduction

Environmental problems such as fossil energy shortage and global warming are becoming more and more serious. To ensure socially sustainable development, it is imperative to develop renewable fuels with low CO_2 emission. As the only carbon-containing renewable energy, biomass can be converted into gas, liquid and solid products, and bio-fuels derived from biomass can partially replace fossil fuels. Furthermore, as bio-fuels can effectively solve the shortage of resource and environmental pollution, bio-fuels development has gradually attracted attention.

The prices of bio-gasoline and bio-diesel are generally higher than those of petroleum-derived gasoline and diesel, because the biomass is more expensive than crude oil and bio-refinery needs a great deal of capital investment (Nuno, C et al, 2019). In addition, bio-diesel and bio-gasoline contain only partial fractions of diesel and gasoline and need further blending processes. Therefore, how to reduce the production costs of bio-fuels satisfying national fuel standards has become a research hotspot (Vasalos I A et al, 2016; Kan T et al, 2016).

Generally, there are three advantages in co-processing of bio-oil and vacuum gas oil in an existing refinery. Firstly, the utilization of the existing refinery infrastructure can greatly reduce the investment cost of bio-refineries. Secondly, the transportation cost from bio-fuel blending can be saved by using existing blending processes. Finally, the refinery mature fuel distribution system can be effectively used for the distribution of co-processing products. Thus, co-processing of bi-oil and vacuum gas oil is one of the effective ways to reduce bio-fuel production cost (Wu L et al, 2019).

According to the previous studies, co- processing of bio-oil and vacuum gas oil in an FCC to produce diesel and gasoline containing a certain amount of bio-carbon showed a highly technical feasibility. As fast pyrolysis and catalytic pyrolysis both can be used to produce bio-oil, they are both investigated for bio-oil co-processing with VGO. For the co- processing of fast pyrolysis oil and VGO, Pinho et al. (2017) co-fed fast pyrolysis and VGO directly in FCC. In order to reduce coke yield, the bio-oil co-processing ratio should less than 20%. Huynh et al. (2016) studied 20% hydrodeoxygenation (HDO) bio-oil and 80% VGO in FCC and found similar yields of gasoline and diesel were obtained compared with the yields of pure VGO cracking. Due to the high oxygen content and low enthalpy value, fast pyrolysis bio-oil needs further hydrogenation before co-pressing with VGO to ensure the yields of co-processing gasoline and catalytic diesel.

Due to high enthalpy value as well as low oxygen and water content, catalytic pyrolysis bio-oil can be directly co-processed with VGO. Wang et al. (2018) drew a conclusion that the diesel and gasoline yield would not be affected when co-processing 10% catalytic pyrolysis oil and 90% VGO. According to ^{14}C analysis, there are more than 7% renewable carbon in gasoline products. Lindfors et al. (2015) analyzed the differences in the co-processing of fast pyrolysis oil, HDO oil and catalytic pyrolysis oil with VGO in the FCC. The results showed that if the adding proportion of bio-oil exceeded 20%, the coke yield increased. The liquid fuel yield of fast pyrolysis oil was the lowest while the yields of HDO oil and catalytic pyrolysis oil were similar to the yields of pure VGO cracking (Sauvanaud L et al, 2018).

The key advantage of co- processing technology is to reduce environmental pollution by introducing renewable energy into fossil fuel refineries, and the technical and economic analysis of bio-oil co-processing with VGO in FCC by Wu et al. (2019) shows that the price of gasoline produced by co- processing was only $2.63/gallon. Thus, the further promotion and application of co-processing technology has attracted great attention. However, the CO_2 emission during co-processing are still not clear, especially in the context of "carbon peak and carbon neutralization".

In this paper, the co- processing of 10% bio-oil and 90% VGO in an FCC is taken as an example. The co-processing process of fast pyrolysis, catalytic pyrolysis and pure VGO was simulated and sensitivity analyzed by using Aspen Plus, and the CO_2 emission were analyzed and compared.

2. Process description

As it is shown in Figure 1, the co-processing process can be divided into two parts, the bio-oil production process and the co-processing of VGO and bio-oil in the existing infrastructure of the refinery. Considering that both fast and catalytic bio-oil can co-processing with VGO in FCC. In this paper, we simulate the fast pyrolysis bio-oil and VGO co-processing scenario (fast pyrolysis scenario), catalytic pyrolysis bio-oil and VGO co-processing scenario (catalytic pyrolysis scenario), and pure VGO cracking scenario (pure VGO scenario) to further compare and analyze the CO_2 emission of different biomass sources and pure VGO cracking, the specific flow is shown in Figure 1.

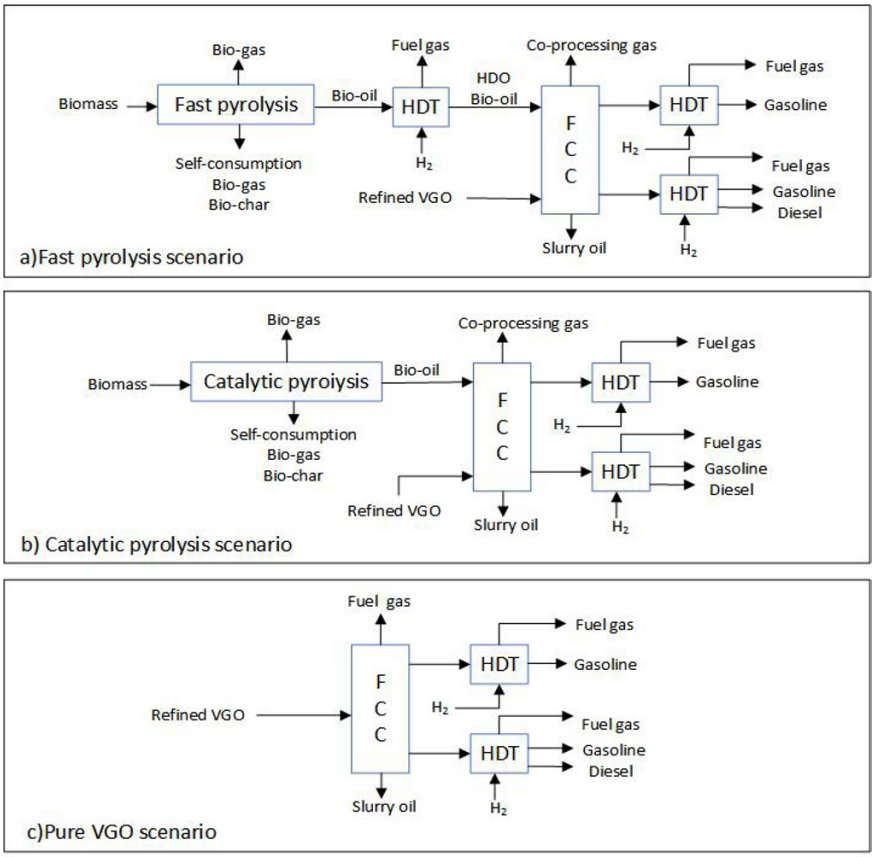


Figure 1 Diagram of co-processing of bio-oil and VGO. a) Fast pyrolysis, b) Catalytic pyrolysis, c) Pure VGO

2.1. Bio-oil production

In the fast pyrolysis scenario, the biomass raw material enters the grinder before entering the dryer to allow for a more efficient removal of additional water. Then mixture of dry biomass and hot sand enters the pyrolysis reactor to obtain bio-gas, bio-oil and biomass carbon. The reaction effluent is separated in a cyclone to obtain sand and bio-char from the bottom and bio-gas and bio-oil from the top. Bio-oil is separated with the bio-gas in a separator; the sand and bio-char are reentered into the pyrolysis reactor to reuse the sand. The biomass raw material and the recycle sand are heated by the combustion the bio-char and partial bio-gas. Due to the high content of oxygen, water and low enthalpy values, fast pyrolysis oil requires further hydrogenation treatment (HDT) to remove excess oxygen and water. Fast pyrolysis oil is pressurized by the pump and mixed with high pressure hydrogen into a hydrodeoxygenation (HDO) reactor. Then several separators are used to obtain reaction wastewater, fuel gas and HDO oil.

In the catalytic pyrolysis scenario, the catalytic pyrolysis process is similar to the fast pyrolysis process, and the catalyst regenerator is involved besides the above equipment. Since the high enthalpy values of the catalytic pyrolysis oil can be directly co-processed with the VGO, the bio-oil hydrogenation device is not included in this scenario.

Since the pure VGO scenario feeds only VGO, the scenario does not contain the bio-oil preparation process.

2.2. Bio-oil co-processing with VGO

According to Figure 1, HDO bio-oil or catalytic pyrolysis oil are co-fed with VGO into FCC reactor where the reaction effluent was separated into fuel gas, FCC gasoline, FCC diesel and slurry oil. Separated diesel and gasoline are refined in the corresponded HDT process to remove sulfur, nitrogen, and oxygen impurities from gasoline and diesel.

3. Aspen Simulation

Massachusetts Institute of Technology developed Aspen software for process simulation in the late 1970s. The software has been widely used in the chemical field, not only can conduct production device design, steady-state simulation optimization, but also can establish a process simulation system, so that the process can get more intuitive prediction results before operation. Effectively reduce casualty accidents caused by design errors. The software also establishes model simulation processes, sensitivity analysis, and physical properties and regression analysis.

3.1. Component provision and material method

The density of VGO is 0.85 to 0.9 g/cm^3, distillation range is 350 to 540 ℃; the density of bio-oil density is 1.1~1.2 g/cm^3, distillation range is 140 to 350 ℃. N_2, H_2O, CO_2, NO_2 and CH_4, C_2H_6, C_3H_8 are conventional fractions and type Conventional. The data in the oil virtual component is from the reference (Cruz P L et al, 2017). GRAYSON was selected as the physical property method.

3.2. Process module analysis

Firstly, select the Heater to pretreat the mixed oil; then, mixer module is selected to mix VGO and HDO oil, separation was performed according to the group shunt rate; select Plug and Yield reactor to simulate the co-processing process of the bio-oil and VGO; sep module shall be adopted and used to separate productions.

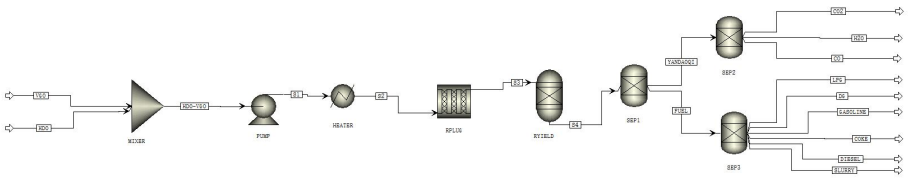


Figure 2 The process of the simulation by Aspen

3.3. Simulation results

The total CO_2 emission in co-processing process is 2.94×10^6 t·y^{-1} and the total CO_2 emission is 3.27×10^6 t·y^{-1} in pure vacuum gas oil. The amount of CO_2 emission from co-processing process was 10.36% less than in the case of pure vacuum gas oil. The result is similar to recent studies, which indicates that the simulation result is accurate.

4. Results and discussion

4.1. CO_2 emission

The total CO_2 emission consisting of the direct emission from fuel combustions and indirect emission from utility.

The CO_2 emission from fuels are calculated as formula (1).

$$E_{CO_{2,1}} = \frac{M_{CO_2}}{A_C} P\omega_C \tag{1}$$

where $E_{CO_{2,1}}$ is the CO_2 emission from raw material or product combustion, in t / a; M_{CO_2} and A_C is the molar mass of CO_2 and C, respectively, in g/mol; P is the raw material consumption or product generation, in t/a; ω_C is the carbon quality fraction in the raw material or product, in %.

CO_2 indirect emission from utility consumption are calculated as formula (2).

$$E_{CO_{2,2}} = u^{H_2} f_{CO_2}^{H_2} + u^{Stream} f_{CO_2}^{Stream} + u^{Water} f_{CO_2}^{Water} + u^{Elec} f_{CO_2}^{Elec} \tag{2}$$

where u^{Stream} and u^{Water} is consumption of steam and recycle water, in t/a; u^{H_2} is consumption of hydrogen, in m^3/a; u^{Elec} is consumption of electricity, in kW/a; $f_{CO_2}^{Stream}$ and $f_{CO_2}^{Water}$ is CO_2 emission coefficient of steam and recycle water respectively, in t-CO_2/t; $f_{CO_2}^{Water}$ is CO_2 emission coefficient of hydrogen, in t-CO_2/m^3; $f_{CO_2}^{Elec}$ is CO_2 emission coefficient of electricity, in t-CO_2/kW.

4.2. Results

By calculation, we obtained the results in Table 1.

Table 1 CO_2 emission

CO_2 emission t/y	Fast pyrolysis	Catalytic pyrolysis	Pure VGO
Utility	2.98×10^5	2.93×10^5	1.76×10^5
Product combustion	2.71×10^6	2.74×10^6	3.20×10^6
Total	3.01×10^6	3.03×10^6	3.38×10^6

The results in Table 2 show that the CO_2 emission of both co-processing methods, fast and catalytic pyrolysis, are less than in the case of pure vacuum gas oil. Therefore, the use of co-processing technology can effectively reduce the amount of CO_2. As an environmental protection technology, co-processing technology has great research prospects.

5. Conclusion

Using the co-processing process of 10% bio-oil and 90% vacuum gas oil in FCC, the co-processing process and the CO_2 emission were analysed. The results show that the total CO_2 emission at co-processing process was 2.94×10^6 $t\cdot y^{-1}$ and the total CO_2 emission was 3.27×10^6 $t\cdot y^{-1}$ for pure vacuum gas oil. To sum up, compared with the pure vacuum gas oil technology, co-processing technology is an environment-friendly

technology which can reduce carbon emissions to some extent. Therefore, when producing gasoline and diesel fuel which contain biomass, we should not only pay attention to the carbon emissions of the product, but also the CO_2 emissions generated in the process. While reducing the carbon emissions of the product, we should reduce the CO_2 emissions in the product production process.

References

Cruz P L, Montero E, Dufour J, 2017, Modelling of co-processing of HDO-oil with VGO in a FCC unit, Fuel, 196, 362-370.

Huynh T M, Armbruster U, Atia H, et al, 2016, Upgrading of bio-oil and subsequent co-processing under FCC conditions for fuel production. Reaction Chemistry & Engineering, 1, 239-251.

Kan T, Strezov V, Evans T J, 2016, Lignocellulosic biomass pyrolysis: A review of product properties and effects of pyrolysis parameter. Renewable & Sustainable Energy Reviews, 57, 28, 1126-1140.

Lindfors C, Paasikallio V, Kuoppala E, et al, 2015, Co-processing of Dry Bio-oil, Catalytic Pyrolysis Oil, and Hydrotreated Bio-oil in a Micro Activity Test Unit, Energy & Fuels, 29, 6, 3707-3714.

Nuno, C., Cruz, et al. 2019, Critical review of key variables affecting potential recycling applications of ash produced at large-scale biomass combustion plants, Resources Conservation & Recycling, 150, 104427.

Pinho A, Almeida M D, Mendes F L, et al, 2017, Fast pyrolysis oil from pinewood chips co-processing with vacuum gas oil in an FCC unit for second generation fuel production, Fuel, 188, 10, 462-473.

Sauvanaud L, Mathieu Y, Corma A, et al. 2018, Co-processing of lignocellulosic biocrude with petroleum gas oils, Applied Catalysis A: General, 551, 9, 139-145.

Vasalos I A, Lappas A, Kopalidou E P, et al, 2016, Biomass catalytic pyrolysis: process design and economic analysis, Wiley Interdisciplinary Reviews: Energy and Environment, 5, 3, 370-383.

Wang C, Venderbosch R, Fang Y, 2018, Co-processing of crude and hydrotreated pyrolysis liquids and VGO in a pilot scale FCC riser setup, Fuel Processing Technology, 181, 9, 157-165.

Wu L, Wang Y, Zheng L, et al, 2019, Design and optimization of bio-oil co-processing with vacuum gas oil in a refinery, Energy Conversion and Management, 195, 5, 620-629.

Wu L, Wang Y, Zheng L, et al, 2019, Techno-economic analysis of bio-oil co-processing with vacuum gas oil to transportation fuels in an existing fluid catalytic cracker, Energy Conversion & Management, 197.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50172-X

Dynamic Simulation and Optimization of a Subcritical Coal-Fired Power Plant During Load-Ramping Operations

Jinliang Ma[ac*], Miguel A. Zamarripa[ac], John C. Eslick[ac], Quang M. Le[d], Debangsu Bhattacharyya[d], Lorenz T. Biegler[e], Stephen E. Zitney[b], Anthony P. Burgard[a], David C. Miller[a]

[a]*National Energy Technology Laboratory, Pittsburgh, PA, 15236, USA*
[b]*National Energy Technology Laboratory, Morgantown, WV, 26507, USA*
[c]*NETL Support Contractor, Pittsburgh, PA, 15236, USA*
[d]*Dept. of Chem. Engineering, West Virginia University, Morgantown, WV, 26506, USA*
[e]*Dept. of Chem. Engineering, Carnegie Mellon University, Pittsburgh, PA, 15213, USA*
jinliang.ma@netl.doe.gov

Abstract

With increasing penetration of renewable energy sources, traditional power plants are required to operate more flexibly with higher ramp rates than originally planned for by design. A dynamic model for a subcritical coal-fired power plant was developed in the open-source, equation-oriented IDAES (Institute for the Design of Advanced Energy Systems) modeling framework. PID-based regulatory controllers and a supervisory-level coordinated control system (CCS) were also implemented. The plant-wide dynamic model and controls were used to investigate and optimize flexible operations. Plant-wide dynamic simulations were performed under load-ramping cycles from full load to 50 % load and back to full load at different ramp rates. In addition, dynamic optimizations were performed at ramp rates of 5 % full load per minute. The objective of the dynamic optimization was to maintain a stable main steam temperature (MST) at the boiler exit while maximizing plant efficiency during ramping. Dynamic simulation results show large excursions in MST, which are detrimental to equipment health. The CCS implementation leads to a drop in MST during the ramp down and an overshoot during the ramp up when attemperator water spray capacity is limited. Through dynamic optimization, the maximum MST deviation is reduced from 12 K, at a 5 %/min ramp rate from full load to 70 % load, to within 2 K by changing three manipulated variable profiles: the furnace stoichiometric ratio, the water spray valve opening, and the setpoint for sliding-pressure operation. This improves the average plant efficiency slightly, while the reduced MST excursion benefits the health of the steam turbine.

Keywords: Power Generation; Dynamic Simulation; Optimization; Flexible Operation.

1. Introduction

Due to the intermittency of renewable energy, conventional fossil-fueled power plants originally designed to operate at full load are increasingly load following. To conduct quantitative assessments of flexible operations, predictive nonlinear first-principles models are highly desired to capture equipment dynamics for off-design conditions. Moreover, distributed-parameter equipment models should be developed to resolve temperature and pressure distributions so that health-related performance variables such

as thermomechanical stresses can be calculated. The major contributions of this work are the development of an open-source, first-principles plant-wide dynamic model and its application to large-scale dynamic optimization for maximizing plant efficiency, while minimizing equipment degradation during transient load-ramping operation.

Dynamic models of coal-fueled power plants have previously been reported in the literature. Oko and Wang (2014) developed a gPROMS dynamic model for a 500 MW subcritical power plant and validated it at full load and three off-design conditions. However, the study did not consider optimization and equipment health under flexible operations. Chen et al. (2017) developed a 605 MW subcritical plant model in Modelica/Dymola and performed dynamic simulations with step changes in fuel flow rate. In an extended work, Chen and Bollas (2018) performed dynamic optimization to maximize the integral of plant efficiency over a predetermined time horizon of 24 hours based on a defined load profile proportional to a typical daily market demand. However, the ramp rate is very small and time step size is large (hours). Therefore, the dynamic model does not capture the effects of rapid (minutes) load ramping typical of a plant dispatch event. For a supercritical plant, Sarda et al. (2018) developed an Aspen Plus Dynamics model and compared three control configurations for load decreases from 100 % to 40 % at a ramp rate of 3 % maximum continuous rating per minute; however, no dynamic optimization was reported.

The work reported here is focused on dynamic simulation and optimization during load ramping. A generic (non-plant specific) 300 MW net subcritical coal-fired power plant is developed and used to simulate flexible operations at different ramp rates using a coordinated control system. To improve plant performance, dynamic optimization is conducted based on predefined load ramping profiles at 5 % full-load power output per minute. The optimization objectives are to limit deviation of MST from its setpoint to avoid high thermal shock to the superheaters and steam turbine while keeping plant efficiency as high as possible during ramping. The power plant dynamic flowsheet model is described in Section 2, followed by dynamic simulation and optimization results in Sections 3 and 4, respectively. Conclusions are drawn in Section 5.

2. Dynamic model description

Figure 1 shows the process flow diagram of the 300 MW subcritical power plant (boiler subsystem and steam cycle) developed using the open-source IDAES modeling framework (Lee et al., 2021). IDAES offers an extensive library of power plant models (see https://github.com/IDAES/idaes-pse) and associated documentation at https://idaes-pse.readthedocs.io/en/stable/technical_specs/model_libraries/index.html. Details of the first-principles unit operation models and their validation using operating data from an existing subcritical power plant are reported by Eslick et al. (2021). The dynamic models implement mass and energy balance equations with temporal or accumulation terms to account for fluid inventories in boiler waterwall and heat exchanger tubes and in large vessels including the boiler drum, deaerator, feedwater heaters (FWHs), and condenser hotwell. Since metal components such as tubes and vessel walls have high density and high heat capacity, their thermal holdups are also considered. Distributed-parameter models are used for convective heat exchangers and the boiler drum and waterwalls. Performance equations include first-principles relationships for heat transfer and pressure change. Due to its complexity, the boiler fire-side model is a surrogate model based on a rigorous 1D/3D hybrid model developed previously (Ma et al., 2016).

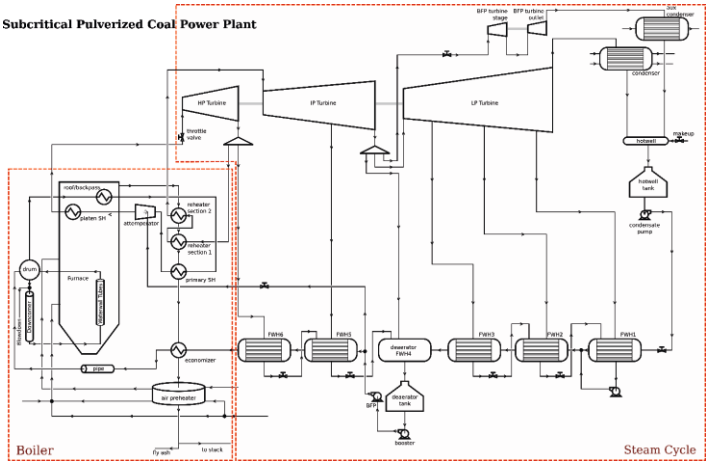


Figure 1. Process flow diagram of a 300 MW coal-fired subcritical power plant

Unit models with relatively small time constants, including the boiler fire-side, turbine stages, pumps, and valves, are treated as steady-state models in the dynamic flowsheet. Regulatory proportional-integral (PI) controllers are applied to control the water levels of the FWHs, condenser hotwell, and deaerator tank. The drum level is controlled by a three-element controller involving two cascading PI controllers. The MST is controlled by the attemperator through a proportional only controller by default. The power demand is met by a CCS with a turbine master controlling the power output by adjusting the steam flow through the throttle valve before the high-pressure (HP) turbine and a boiler master controlling the fuel and air flow rates to bring the main steam pressure to a desired sliding pressure, coordinated with the turbine master.

3. Dynamic simulations

Three dynamic simulations were performed for cycling from 100 % load to 50 % and back to 100 % using ramp rates of 5 %, 3 %, and 1 % of full load power output per minute, respectively. The simulations are based on the default settings of the control system with desired sliding pressure defined as a linear function of the load demand. Figure 2 shows the dynamic simulation results for the three load ramping cases. Figure 2(a) shows the predicted MST profiles (solid lines) along with the defined load demand profiles as inputs (dashed lines). Figure 2(b) shows the stoichiometric ratio (SR) and the attemperator control valve open fraction profiles calculated by the control system.

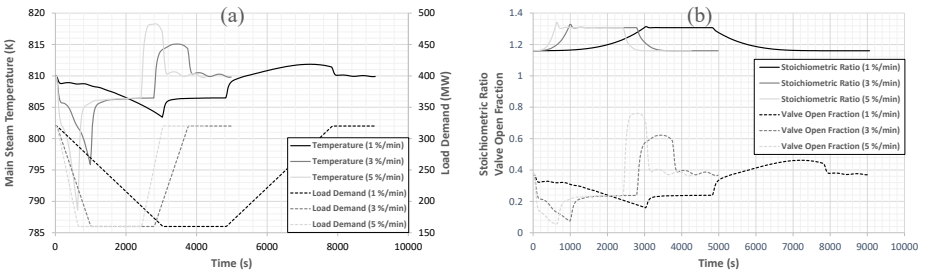


Figure 2. Dynamic simulation results of three load cycling cases: (a) Load demand and main steam temperature (MST), (b) SR and valve opening fraction

We observe that the MST drops below the setpoint of 810 K when ramping down. This occurs because the coal flow rate is reduced and is lower than the steady-state values due to release of stored energy in the system. The drop in MST cannot be corrected immediately by the attemperator. The 5 % ramping case has its maximum MST drop greater than 21 K during ramp down and the highest MST increase is greater than 8 K during ramp up. The deviation of the MST from its setpoint has detrimental effects on the health of superheaters and steam turbine. MST deviation can be partially mitigated by attemperator control, but it usually cannot eliminate the deviation, especially when ramping down at high rates and when the valve opening reaches its 5 % minimum set by the model. In the default controller setting, the furnace SR is set to be a polynomial function of coal flow rate with a higher SR used at a lower coal flow rate, such that MST maintains its setpoint of 810 K at steady-state. The large MST drop when ramping down is due to the release of stored energy in the system (mainly in drum and waterwalls), which reduces required coal flow rate below the corresponding steady-state value. To correct this, the SR can be increased beyond its required steady state value. This leads to higher air flow rate, slightly higher coal flow rate, and reduced flame temperatures inside the furnace, thus providing reduced heat absorption by the waterwalls, increased flue gas flow rate, and increased heat absorption by the superheaters. From Figure 2 we observe that higher ramping rates lead to higher MST deviations from the setpoint, since the effect of stored energy is stronger.

4. Dynamic optimization

While dynamic simulation provides insight into the impact of load ramping on system performance—especially MST deviation—dynamic optimization has the potential to reduce MST deviation. The 5 % ramp rate was chosen for dynamic optimization, since its impact is strongest among the three control cases considered here. To reduce CPU time for dynamic optimization, up and down ramps are modelled separately, and the load range was confined between 70-100 %. Only ramp down cases are reported here. Case A is without dynamic optimization, i.e., based on default controller settings. Case B has attemperator spray valve opening and furnace SR as decision variables for the optimization. Case C extends Case B with an additional decision variable, sliding-pressure deviation from the default value calculated as a linear function of load.

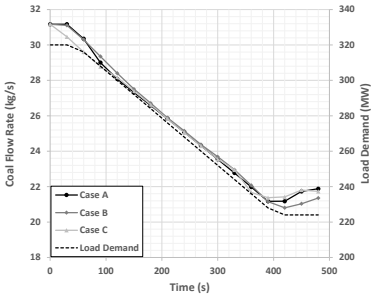
The objectives of the dynamic optimization are to minimize the MST deviation during the transient ramp while minimizing the coal flow rate or maximizing the plant efficiency. The objective function for the multi-objective optimization is defined as:

$$f_{obj} = 0.001 \sum_{i=0}^{N} \left(T_{steam,i} - T_{steam,setpoint}\right)^2 + \sum_{i=0}^{N} \frac{\dot{m}_{coal,i}}{\dot{m}_{coal,full\_load}} \quad (1)$$

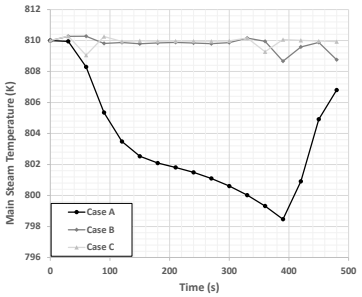
where $T_{steam,i}$ is MST at time point *i* and $T_{steam,setpoint}$ is the setpoint (810 K). $\dot{m}_{coal,i}$ and $\dot{m}_{coal,full\_load}$ are coal flow rate at time point *i* and at the initial steady-state full load, respectively. A weighting factor of 0.001 is adopted for the temperature deviation term. Ranges for the manipulated variables are between 1.15 and 1.3 for furnace SR, 5 % and 95 % for spray valve opening, and ±0.5 MPa for the sliding pressure adjustment.

The dynamic optimization cases were modelled on the IDAES platform leading to nonlinear programs with over 60,000 variables and were solved with IPOPT. The simulation problem (Case A) required 7 CPU minutes, while Cases B and C required the equivalent of up to 170 simulations. The dynamic optimization results are provided in Figure 3. Figure 3(a) shows the coal flow rates required to meet load demand.

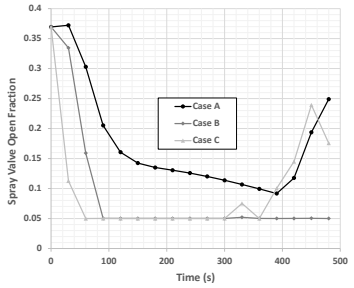
Compared to Case A, the average coal flow rates during the 8-minute period are reduced by 0.03 % for Case B and 0.29 % for Case C. Figure 3(b) shows the MST profiles. The MST drops as low as 12 K below the setpoint of 810 K near the end of the ramp in Case A. In contrast, the maximum MST deviation is reduced to about 1.5 K in Case B and to only 1 K in Case C. Figure 3(c) shows the profiles of the spray valve open fraction for the three cases. The valve is closed quickly to the minimum open fraction of 0.05 at the beginning of the ramp (before 100 s) in the two optimized cases (Cases B and C) to offset the MST drop. Figure 3(d) shows the furnace SR profiles. A higher SR is used for the two optimized cases, which is the main reason for preventing the MST from dropping too low, as discussed in Section 3. Figure 3(e) shows the main steam pressure profiles. Note that the profiles of Cases A and B follow the trend of load demand since the sliding pressure setpoint is proportional to load demand. In Case C, the main steam pressure is optimized to go even lower to improve plant efficiency. Figure 3(f) shows the steam temperature at the HP turbine inlet (after throttle valve). The temperature drop at the HP turbine inlet is also reduced for the two optimized cases, mitigating the negative impact on turbine health during load ramping.



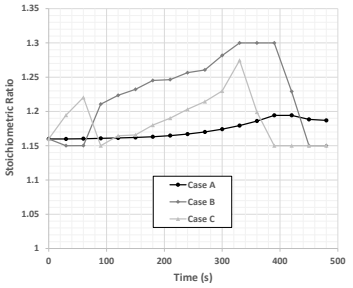
3(a) Coal flow rate



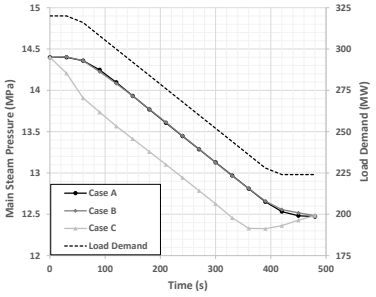
3(b) Main steam temperature (MST)



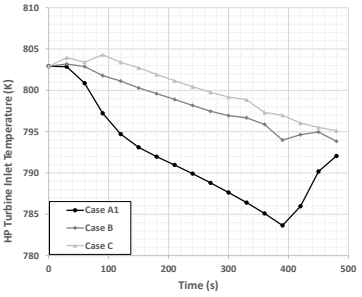
3(c) Spray valve open fraction



3(d) Furnace Stoichiometric ratio



3(e) Main steam pressure



3(f) HP turbine inlet temperature

Figure 3. Dynamic optimization results

4. Conclusions

An equation-oriented, plant-wide dynamic model of a subcritical coal-fired power plant was developed using the open-source IDAES model library and applied to perform dynamic simulations and optimization using state-of-the-art solvers. The nonlinear first-principles dynamic unit operation models, including distributed-parameter models, along with the coordinated control system, provide the ability to predict the excursion of performance variables such as the main steam temperature during fast load ramping. Through dynamic optimization of three manipulated variable profiles, main steam temperature deviation can be dramatically reduced while plant efficiency can be kept unchanged or slightly improved, thus mitigating the negative effect on equipment health.

Acknowledgement & Disclaimer: We graciously acknowledge support by the U.S. Department of Energy, Office of Fossil Energy and Carbon Management, Simulation-Based Engineering Program. This was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, nor the support contractor, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

References

C. Chen, Bollas, G. M., 2018, Dynamic optimization of a subcritical steam power plant under time-varying power load, Processes, 6(8), 114, https://doi.org/10.3390/pr6080114.

C. Chen, Z. Zhou, G. M. Bollas, 2017, Dynamic modeling, simulation and optimization of a subcritical steam power plant. Part I: Plant model and regulatory control. Energy Conversion and Management, 145(9), 324-334, https://doi.org/10.1016/j.enconman.2017.04.078

J.C. Eslick, M. Wang, J. Ma, M.A. Zamarripa, D. Bhattacharyya, S.E. Zitney, A.P. Burgard, D.C. Miller, 2021, Predictive Modeling of an Existing Subcritical Pulverized-Coal Power Plant for Optimization: Data Reconciliation, Parameter Estimation, and Validation, submitted for publication.

A. Lee, J.H. Ghouse, J.C. Eslick, C.D. Laird, J.D. Siirola, M.A. Zamarripa, D. Gunter, J.H. Shinn, A.W. Dowling, D. Bhattacharyya, L.T. Biegler, A.P. Burgard, D.C. Miller, 2021, The IDAES Process Modeling Framework and Model Library –Flexibility for Process Simulation and Optimization, Journal of Advanced Manufacturing and Processing, 3 (3).

J. Ma, J.P. Eason, A.W. Dowling, L.T. Biegler, D.C. Miller, 2016, Development of a first-principles hybrid boiler model for oxy-combustion power generation system, International Journal of Greenhouse Gas Control, 46, 136–157.

E. Oko, M. Wang, 2014, Dynamic modelling, validation and analysis of coal-fired subcritical power plant, Fuel, 135(1), 292-300.

P. Sarda, E. Hedrick, K. Reynolds, D. Bhattacharyya, S.E. Zitney, B. Omell, 2018, Development of a dynamic model and control system for load-following studies of supercritical pulverized coal power plants, Processes, 6(11), 226, https://doi.org/10.3390/pr6110226.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50173-1

Solvent Screening Methodology considering Techno-Economic and Environmental Sustainability criteria for Algae Lipid Extraction

Santiago Zapata-Boada[a*], Maria Gonzalez-Miquel[b], Megan Jobson[a], and Rosa Cuellar-Franca[a]

[a] *Department of Chemical Engineering, The University of Manchester, Manchester, M13 9PL, United Kingdom*
[b] *Department of Chemical and Environmental Engineering, Universidad Politécnica de Madrid, Madrid, 28006, Spain*
santiago.zapataboada@manchester.ac.uk

Abstract

Solvent extraction plays a crucial role in biorefinery processes, such as the extraction of lipids from algae biomass to produce biodiesel. Volatile organic solvents, such as hexane, are typically used to extract algae lipids, but this involves high energy consumption for solvent recovery and negative environmental impacts due to its hazardous, volatile, and flammable nature. This study proposes a systematic methodology combining molecular simulation techniques, data classification methods, and process simulation to screen solvents for lipid extraction from wet algae biomass. The novelty relies on the integration of techno-economic, environmental and safety criteria, such that all three sustainability dimensions are considered in solvent selection. First, thermodynamic equilibrium data of solvent-lipid systems are predicted with the COSMO-RS method. Then, a clustering method (K-means algorithm) is used to identify and select the solvents with the highest partition coefficient and selectivity towards lipids. After this, the selected solvents are further screened considering their physicochemical properties (solubility in water, density, viscosity, boiling point) and health, safety, and environmental performance. Finally, the lipid extraction process with the shortlisted solvents is simulated in Aspen Plus to obtain techno-economic and environmental sustainability performance indicators. Out of 88 initial solvent candidates, limonene and ethyl tert-butyl ether are identified as promising alternatives to the benchmark solvent (hexane), as they are non-hazardous and have higher selectivity towards lipids. However, their use involves higher energy requirements, operating costs, and greenhouse gas emissions in the lipid extraction process. This methodology can be applied to other extraction processes to find sustainable alternatives for the conventional solvents and identify trade-offs from a techno-economic and environmental perspective at the early stages of process design.

Keywords: COSMO-RS, process simulation, biobased solvents.

1. Introduction

Solvent screening methods are used to identify promising solvents depending on application-specific technical performance; health, safety, and environmental (HSE) criteria; and process-specific economic considerations (Clarke et al., 2018). Solvent technical performance (extraction efficiency and selectivity) can be evaluated with experimental equilibrium data or molecular simulation methods, such as COSMO-RS

(Eckert and Klamt, 2002). The capability of the COSMO-RS method to predict partition ratios as a key parameter for solvent selection has been successfully validated in previous studies (López-Porfiri et al., 2020). HSE criteria of solvents can be found in solvent selection guides, such as CHEM21 (Prat et al., 2016), which give a preliminary ranking of solvent "greenness" based on their physical properties and hazard statements. Process-specific considerations can be captured with process simulation, which is used to calculate performance indicators for measuring energy consumption, economic viability and environmental sustainability and for exploring design improvements (Sheldon, 2018).

Key applications of solvent extraction include the downstream processing of biorefinery applications, which are challenging due to the broad range of impurities, heat-sensitive components, and high dilution of the feed streams, and can account for 15-80% of the total costs (Ghosh, 2007). The extraction of lipids from algae biomass to produce biodiesel is an important biorefinery application (Chen et al., 2018). Conventional processes for lipid extraction use volatile organic solvents derived from fossil resources, such as hexane, but its toxicity and high volatility make it environmentally unfavourable (Clarke et al., 2018). Alternative solvents, such as terpenes, have been identified in experimental screening studies (Dejoye Tanzi et al., 2012) and proposed as ideal alternatives to hexane due to their higher extraction efficiency and selectivity. However, the impact of these alternative solvents on the economic viability and environmental sustainability at a process level has not been considered. This study presents a novel approach that combines molecular and process simulation techniques to integrate techno-economic, environmental and safety indicators to screen solvents for algae lipid extraction. The following section describes the methodology, followed by the results and the conclusions.

2. Methodology

The proposed methodology for solvent screening is illustrated in Figure 1. It consists of five steps: liquid-liquid equilibrium (LLE) data calculation using COSMO-RS method, LLE data analysis and classification, evaluation of physicochemical properties of solvents, evaluation of HSE performance of solvents, and process simulation.

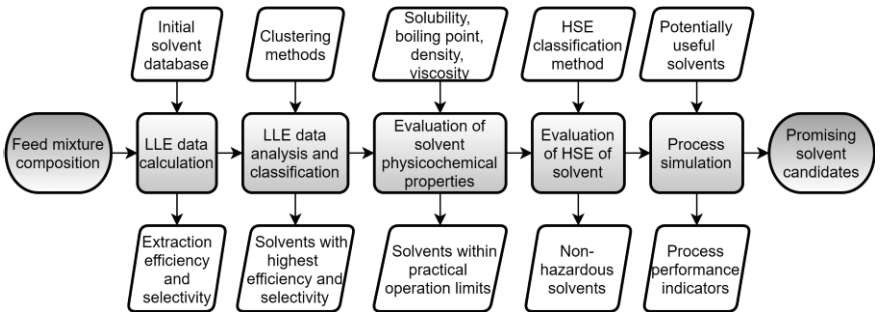


Figure 1. Methodology proposed to evaluate solvents at the process development stage

2.1. Liquid-liquid equilibrium data calculation

The feed (wet algae biomass) is a dilute aqueous stream with a water content of 80% wt., lipid content of 6% wt., and protein and carbohydrates content of 14% wt. It is assumed that the lipids are represented by the molecule triolein, which is one of the most common triglycerides found in algae lipids (Yao et al., 2015). Glutamine, a polar molecule found

in algae biomass, is assumed to represent other components. The extraction conditions are set to 25°C and 101.325 kPa, which are typical values for algae lipid extraction. An initial solvent database (Sels et al., 2020) is used to propose a 'long list' of candidate solvents. Then, solute (triolein and glutamine) activity coefficients at infinite dilution in the aqueous and organic phases are calculated using the COSMOthermX software, which implements the COSMO-RS method (Eckert and Klamt, 2002).

The extraction efficiency is expressed as the partition coefficient of the solute between the aqueous and organic phases (K_i). At low solute concentrations (dilute systems), the partition coefficient can be predicted by modelling the solute activity coefficient at infinite dilution ($\gamma^{\alpha,\infty}$) according to Eq.(1), where x is the mole fraction of the solute i, s corresponds to the solvent phase and w to the aqueous phase. The selectivity of the solvent between components i and j ($\beta_{i,j}$) is defined as the ratio of the partition coefficients of i (triolein) and j (glutamine).

$$K_i = \frac{x_i^s}{x_i^w} = \frac{\gamma_i^w}{\gamma_i^s} \approx \frac{\gamma_i^{w,\infty}}{\gamma_i^{s,\infty}} \tag{1}$$

2.2. LLE data analysis and classification

A clustering method, namely the k-means method (Bock, 2007), is used to classify the solvents according to the partition coefficient of triolein in the aqueous and organic phases and the selectivity of solvents to triolein. The number of clusters (seven) is selected based on an analysis of the cluster inertia (sum of squared distances of samples to the nearest cluster center), aiming for a small inertia value, which is a sign of good and meaningful clustering. The solvents belonging to the cluster with the highest partition coefficients and selectivity to triolein are selected for further evaluation.

2.3. Evaluation of physicochemical properties of solvents

Further screening of the selected solvents resulting from the previous step is conducted based on practical considerations: (a) avoiding loss of solvent to the raffinate phase; (b) avoiding thermal degradation of the solute; (c) ease of recovery; (d) ease of handling. The following limits are defined for the physicochemical properties of solvents: solubility in water < 1% wt. (to form a two-phase system and minimize solvent loss to the aqueous phase); boiling point < 200 °C (to prevent thermal degradation of triolein), density < 1.5 kg/L, and viscosity < 2 mPa.s (ease of handling and storage). Solvents that do not meet these criteria are discarded.

2.4. Evaluation of health, safety, and environmental performance of solvents

The HSE performance of the selected solvents from the previous step is evaluated using the CHEM21 methodology (Prat et al., 2016), which considers health, safety, and environmental scores, and indicates whether a solvent is 'hazardous', 'problematic' or 'recommended'. Hazardous solvents are discarded, as their substitution during process development is a priority of this work. Finally, a shortlist of potentially useful solvents for the application is obtained.

2.5. Process simulation

A conceptual process design is proposed for continuous extraction of lipids from 20 t h-1 of wet algae biomass, using the shortlisted solvents from the previous step. Aspen Plus V8.8 is used to simulate the process (see Figure 2). The NRTL activity coefficient model

is used to simulate phase equilibrium for the lipids-solvent-water system, and missing interaction parameters are estimated using the UNIFAC method. Solvents' performance is compared by assuming that a counter-current extraction column with five stages is needed to extract lipids from algae biomass (Stephenson et al., 2010). The solvent-to-dry feed mass ratio is assumed to be 5 to 1 (Davis et al., 2014).

Mass and energy flows obtained from the process simulation are used to calculate technical, economic, and environmental performance indicators of the extraction process. Two technical performance indicators are proposed: lipid recovery (LR), the ratio of extracted lipids to the total lipids in the feed; and solvent intensity (SI), the ratio of the mass of solvent entering the extractor to the extracted lipids. One economic performance indicator is proposed: total annualized cost (TAC), including capital and operating costs. These are calculated using Aspen Process Economic Analyzer. Two environmental indicators are used: energy intensity (EI), the ratio of the energy input to the process (heat used in the stripping column for solvent recovery) to the mass flowrate of extracted lipids; and global warming potential (GWP), estimated from the greenhouse gas emissions resulting from the combustion of fuel (natural gas) used to supply the energy to the process.

3. Results

Out of 88 initial solvent candidates, 37 solvents are identified in the cluster that exhibits the highest partition coefficients and selectivity to triolein. These solvents are selected for further evaluation. All these solvents are non-polar and include hydrocarbons derived from fossil resources (e.g., heptane); bio-based hydrocarbons (e.g., limonene); ethers (e.g., ethyl tert-butyl ether); and esters (e.g., methyl laurate). After evaluating the selected solvents' physicochemical properties and HSE performance, 26 solvents are discarded due to their solubility in water > 1% wt. (e.g., tetrahydrofuran), boiling point > 200 °C (e.g., methyl laurate), density > 1.5 kg/L (e.g., carbon tetrachloride), and hazardous nature (e.g., chloroform). As a result, only 11 solvents are shortlisted for further investigation via process simulation. Table 1 presents process performance indicators of the shortlisted solvents, calculated from the simulation results. The indicators consider the impact per unit of product (kg lipid).

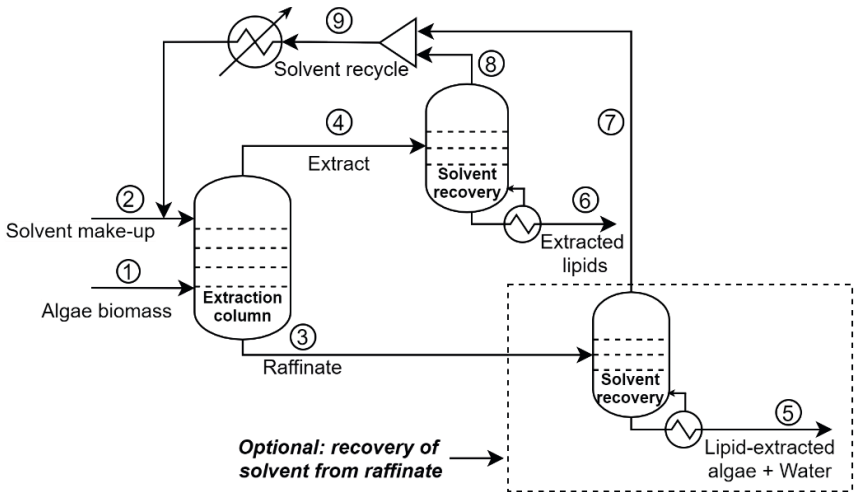


Figure 2. Process flow diagram of the lipid extraction process

Table 1. Process performance indicators of the screened solvents

Solvent	LR (%)	SI (kg solvent/ kg lipid)	TAC (USD/kg lipid)	EI (MJ/kg lipid)	GWP (kg CO_2 eq./kg lipid)
Hexane*	99.9	0.09	2.10	9.48	0.49
Cyclohexane	99.8	0.09	2.13	10.13	0.52
Methylcyclohexane	99.7	0.09	2.15	10.46	0.54
Heptane	99.9	0.09	2.17	10.73	0.55
Ethyl cyclopentane	99.7	0.09	2.14	10.71	0.55
Toluene	98.1	0.10	2.14	11.60	0.60
O-xylene	99.2	0.09	2.22	12.98	0.67
Limonene	100	0.08	2.67	13.26	0.68
Cymene	99.9	0.08	2.80	12.73	0.66
Pinene	96.6	0.09	2.54	14.98	0.77
ETBE	91.6	0.18	2.29	9.74	0.50
Dibutyl ether	96.0	0.10	2.57	12.65	0.65

LR: Lipid recovery; SI: Solvent intensity; TAC: Total annualized costs; EI: Energy intensity; GWP: Global warming potential. *Hexane is evaluated for benchmarking purposes

LR is over 98% for all the fossil-based hydrocarbons (cyclohexane, methylcyclohexane, heptane, ethyl cyclopentane, toluene, o-xylene), as they all have a high partition coefficient and high selectivity towards triolein. SI of all conventional solvents is comparable due to the similar losses of solvent in the extract and raffinate phases. Cyclohexane presents the lowest TAC, EI, and GWP among the fossil-based hydrocarbons due to the low reboiler duty required for evaporation. For example, TAC of cyclohexane is comparable to TAC of hexane (1.5% higher). On the other hand, TAC, EI, and GWP of o-xylene are the highest for all fossil-based hydrocarbons due to its high boiling point and enthalpy of vaporization.

LR of terpenes (limonene, cymene and pinene) are over 96%. As Table 1 shows, limonene has the highest LR among the solvents due to its high partition coefficient and selectivity to triolein. Also, terpenes perform better than hexane in terms of SI due to their lower volatility and evaporation losses. For example, SI of limonene is 9% lower. However, terpenes have higher TAC, EI, and GWP than hexane: for pinene, TAC, EI, and GWP are 21%, 58%, and 57% higher. LR of ethers (ETBE and dibutyl ether) are lower than LR of terpenes and fossil-based hydrocarbons, which can be attributed to the more polar nature of ethers. SI of ETBE is nearly double that of hexane due to the higher losses of ETBE in the lipid-extracted algae. However, ETBE is the only alternative solvent that performs similarly to hexane in terms of EI and GWP (within 3% of those of hexane).

Overall, limonene and ETBE are identified as promising alternative solvents to hexane. Both are non-conventional solvents that can be obtained from biomass-based resources. Limonene performs well in terms of LR and SI. ETBE has the lowest EI and GWP of all solvents other than hexane. Even though there is not a clear 'winner' that outperforms hexane in every dimension, as evidenced by the trade-offs in the process performance indicators, the proposed methodology provides a clear indication of the strengths and weaknesses of the solvent candidates, providing a shortlist of potentially attractive solvents for a more detailed evaluation.

4. Conclusions

This work proposes a systematic approach combining molecular simulation techniques, data classification methods, and process simulation to comprehensively screen solvents. The integration of techno-economic, environmental and safety criteria ensure that all three sustainability dimensions are considered in solvent selection by using this novel methodology. This methodology is demonstrated for the extraction of lipids from wet algae biomass to find alternative solvents to the benchmark solvent (hexane), whose use involves negative environmental impacts due to its hazardous, volatile, and flammable nature. A comprehensive database including 88 alternative solvents is used as a starting point, and after evaluation of techno-economic, safety, and environmental performance, limonene and ethyl tert-butyl ether are identified as promising alternatives to hexane. These solvents are non-hazardous, can be obtained from biomass-based resources, and show competitive performance indicators. Overall, the proposed methodology is helpful to identify a shortlist of potentially attractive solvents for a more detailed evaluation, providing a clear indication of the strengths and weaknesses of the solvent candidates. The methodology developed herein can be applied to other solvent extraction processes to promote the development and design of sustainable chemical processes using alternative solvents.

References

Bock, H.-H. 2007. Clustering Methods: A History of k-Means Algorithms. Berlin, Heidelberg: Springer Berlin Heidelberg.

Chen, J., Li, J., Dong, W., Zhang, X., Tyagi, R. D., Drogui, P. & Surampalli, R. Y. 2018. The potential of microalgae in biodiesel production. *Renewable and Sustainable Energy Reviews,* 90**,** 336-346.

Clarke, C. J., Tu, W.-C., Levers, O., Bröhl, A. & Hallett, J. P. 2018. Green and Sustainable Solvents in Chemical Processes. *Chemical Reviews,* 118**,** 747-800.

Davis, R., Kinchin, C., Markham, J., Tan, E., Laurens, L., Sexton, D., Knorr, D., Schoen, P. & Lukas, J. 2014. Process Design and Economics for the Conversion of Algal Biomass to Biofuels: Algal Biomass Fractionation to Lipid- and Carbohydrate-Derived Fuel Products. National Renewable Energy Lab. (NREL), Golden, CO (United States).

Dejoye Tanzi, C., Abert Vian, M., Ginies, C., Elmaataoui, M. & Chemat, F. 2012. Terpenes as green solvents for extraction of oil from microalgae. *Molecules,* 17**,** 8196-8205.

Eckert, F. & Klamt, A. 2002. Fast solvent screening via quantum chemistry: COSMO-RS approach. *AIChE Journal,* 48**,** 369-385.

Ghosh, R. 2007. *Principles of bioseparations engineering,* Singapore, World Scientific.

López-Porfiri, P., Gorgojo, P. & Gonzalez-Miquel, M. 2020. Green Solvent Selection Guide for Biobased Organic Acid Recovery. *ACS Sustainable Chemistry & Engineering,* 8**,** 8958-8969.

Prat, D., Wells, A., Hayler, J., Sneddon, H., McElroy, C. R., Abou-Shehada, S. & Dunn, P. J. 2016. CHEM21 selection guide of classical- and less classical-solvents. *Green Chemistry,* 18**,** 288-296.

Sels, H., De Smet, H. & Geuens, J. 2020. SUSSOL—Using Artificial Intelligence for Greener Solvent Selection and Substitution. *Molecules,* 25**,** 3037.

Sheldon, R. A. 2018. Metrics of Green Chemistry and Sustainability: Past, Present, and Future. *ACS Sustainable Chemistry & Engineering,* 6**,** 32-48.

Stephenson, A. L., Kazamia, E., Dennis, J. S., Howe, C. J., Scott, S. A. & Smith, A. G. 2010. Life-Cycle Assessment of Potential Algal Biodiesel Production in the United Kingdom: A Comparison of Raceways and Air-Lift Tubular Bioreactors. *Energy & Fuels,* 24**,** 4062-4077.

Yao, L., Gerde, J. A., Lee, S.-L., Wang, T. & Harrata, K. A. 2015. Microalgae Lipid Characterization. *Journal of Agricultural and Food Chemistry,* 63**,** 1773-1787.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50174-3

Modelling and simulation of the production of n-butyl lactate in a reactive distillation column at pilot plant scale

García, César [a], Yusti, Daniela [a] , Velandia, Jessica [a,b], Ochoa, Silvia [c] and Gil, Iván D. [a]*

[a]*Grupo de Procesos Químicos y Bioquímicos, Departamento de Ingeniería Química y Ambiental, Universidad Nacional de Colombia Sede Bogotá, Cra 30 # 45-03, Ed. 453, Of. 318, Bogotá – Colombia.*
[b]*Investigación y Desarrollo, Sucroal S.A., Recta Cali Palmira Km 18 desvío a Candelaria - Colombia.*
[c]*SIDCOP Grupo de investigación, Universidad de Antioquia, Colombia*
**idgilc@unal.edu.co, +57 1 3165000 ext: 14100*

Abstract

Petrochemical-based solvents are one of the main sources of volatile organic compounds (VOCs) in the atmosphere, which are reported to be carcinogenic under certain conditions. Different alternatives are being studied in order to reduce this problem. Special interest has been placed in lactate esters obtained from long chain alcohols like propyl, n-butyl and amyl alcohols because they are considered as green solvents due to their toxicological and environmental characteristics. These allow their use in many applications, especially as chemical feedstock, dairy-related flavouring, and drugs administration. Although most of the reported methods for lactate esters production include intensification technologies such as reactive distillation. There are not enough reports related to the evaluation of lactate esters production in a reactive distillation column at pilot plant scale, for obtaining a high productivity and a good quality control of the end-product specifications. In this work, a principles-based model, including NRTL-Raoult thermodynamic model and a pseudo-homogeneous kinetic model, was developed for a reactive distillation column at pilot plant scale located at Universidad Nacional de Colombia. The system was simulated in Matlab/Simulink R2020b considering the column design. The obtained results were validated with simulations in Aspen Plus V11, with a composition of butyl lactate >97% mole fraction.

Keywords: 1-Butyl lactate, Reactive distillation column, Phenomenological based model, Pilot plant scale, Optimization.

1. Introduction

According to the Environmental Protection Agency (EPA), Volatile Organic Compounds (VOCs) have proved to have a negative impact on people's health (EPA, 2021). In this area, chemical solvents constitute around 80% of compounds used in many important chemical processes. A percentage of 0.7% of VOC emissions come from this type of

industries, while 21% of these compounds come from the general uses of these solvents (Häckl & Kunz, 2018).

New solvents called "green solvents" are compounds with potential to replace VOCs. These solvents have advantages over conventional VOCs, such as low toxicity, biodegradability, and renewability (Li et al., 2016). The global market for green solvents was valued at $ 5.4 billion in 2015, in products such as bio-alcohols, bio-glycols, and lactate esters, among others (Grand View Research, 2016).

Nowadays, there is a growing interest in lactate esters and their uses in cosmetic, food and pharmaceutical applications. The conventional process to produce lactic acid esters involves multiple reactions and purification stages. Butyl lactate has great interest due to its wide range of industrial applications. Kumar & Mahajani (2007) studied the esterification kinetic between lactic acid and 1-butanol in the presence of cation exchange resins as a catalyst and assessed the applicability of reactive distillation for this process. Peña-Tejedor et al. (2005) and Garcia et al. (2021) studied the equilibrium of the butyl lactate – 1-butanol binary system at 1, 5 and 101.325 kPa. Velandia et al. (2021) studied the thermodynamic model of the esterification system for butyl lactate production, the reactive and nonreactive residue curves maps, and carried out a conceptual design for reactive distillation.

Due to the limited information available for this system, this work evaluates the feasibility of producing butyl lactate in a pilot-scale reactive distillation column. A phenomenological-based model was developed and implemented in Matlab/Simulink 2020b. This model was validated with experimental data acquired by Kumar & Mahajani (2007) and compared with Aspen Plus V11 simulation.

2. Process Modelling

2.1 Model description

The reactive distillation column at pilot plant scale has a reactive core, a structured packaging module with the Amberlyst 70 exchange resin and other modules containing Nutter Ring No. 07 packaging. Is recirculated the organic phase formed in the decanter. The water produced is removed in the aqueous phase.

Figure 1(a) is a representation of the reactive distillation column at the pilot plant located in the Laboratory of Chemical Engineering at Universidad Nacional de Colombia. The bottom flow rate was fixed at 2.02 kg/h to obtain the higher conversion, while hydraulic restrictions are fulfilled. The lactic acid flow rate was fixed at 4.042 kg/h with a mass fraction of 0.30 of lactic acid and inlet temperature of 353.15 K. The butanol flow rate was fixed at 1.656 kg/h with a mass fraction of 0.99 for butanol and inlet temperature of 363.15 K. Figure 1(b) is a representation of the column for modelling purposes, where stage 1 is the decanter, stage 2 is the condenser and stage 14 is the reboiler.

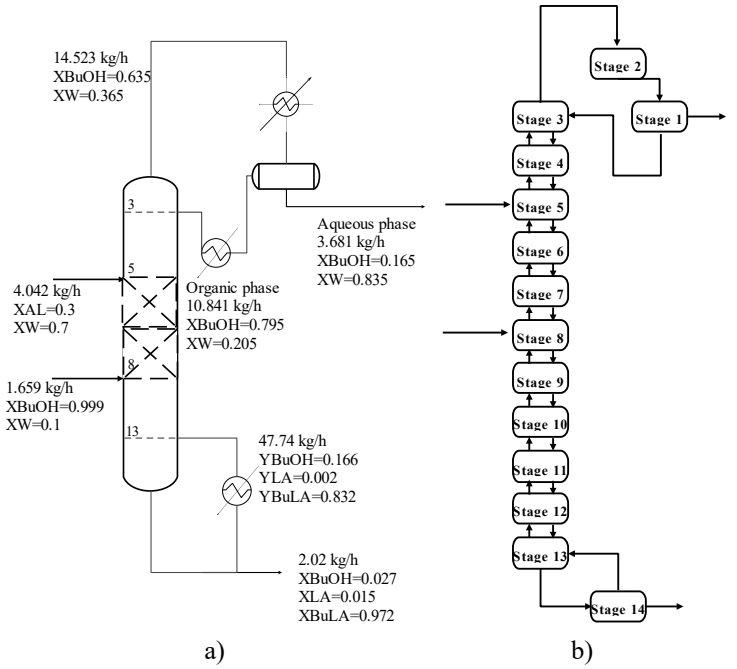


Figure 1. Continuous reactive distillation column: a) at pilot plant scale to produce butyl lactate. b) Distributed Reactive distillation column.

Figure 2 shows a representation of the equilibrium stage model for a general reactive stage. The stages for rectifying, reactive and stripping zones are determined based on the evaluated case (Figure 1 a).

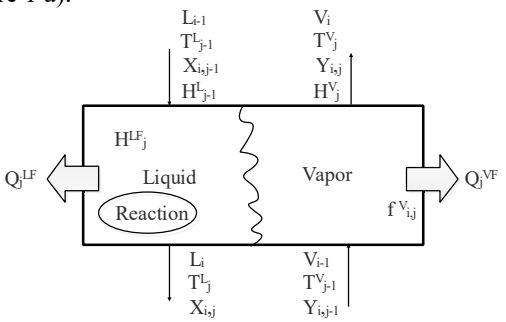


Figure 2. Equilibrium stage model

In this case, the reactive distillation column is represented by an equilibrium model kinetically controlled, which considers that there is no resistance to internal and external diffusive transport over the catalyst and that there is no resistance to transport over the fluid phases. For this reason, the phase equilibrium is achieved homogeneously.

2.2 Model assumptions

For modelling the liquid phase, NRTL model is used and the vapor phase is assumed ideal, as well as a pseudo-homogeneous kinetic model. In the decanter, NRTL model is used for predicting the liquid-liquid equilibrium and the k-value method is used to

determine compositions in extracted and refined flows. The presence of oligomers is neglected as well as the kinetic and potential energy. The reboiler and condenser are modelled as trays and the equilibrium controlled by kinetics is considered. Physical properties are constant and there is no heat loss from the column. The column pressure drop is neglected, and the reaction takes place in liquid phase.

2.3 Balance Equations

In this section, the generalized mass and energy balances for reactive and non-reactive stages (Eq. 1 to 3) are presented, based on the equilibrium stage model (Figure 2). For the three additional stages (decanter, condenser and reboiler), the set of mass and energy balances are assumed as equilibrium stages.

Total Material balance:

$$\frac{dZ_i}{dt} = L_{i-1} + V_{i+1} - L_i - V_i + F_i \tag{1}$$

Molar balance per component:

$$\frac{dx_{i,n}}{dt} = \frac{\left(x_{i-1,n}L_{i-1}+y_{i+1,n}V_{i+1}-x_{i,n}L_i-y_{i,n}V_i+\left(x_{F,n}\right)_i q_{F,i}F_i + R_{n,i}-x_{i,n}\frac{dZ_i}{dt}\right)}{Z_i} \tag{2}$$

Energy balance:

$$\frac{dT_i}{dt} = \frac{h_{i-1}L_{i-1}+H_{i+1}V_{i+1}-h_iL_i-H_iV_i+h_{F,i}q_{F,i}F_i+Q_{r,i}-h_i\frac{dZ_i}{dt}}{Z_i * C_{pmix,i}} \tag{3}$$

Where Fi represents the feed flowrate (Lactic acid +Butanol), $Rn_{,i}$ is the reaction rate, $X_{F,i}$ is the molar fraction of each compound (Lactic /butanol) in the feed and $Q_{r,i}$ the heat of reaction.

3. Model simulation

The column design specifications of the reactive distillation column at pilot plant scale located at Universidad Nacional de Colombia - Bogota are presented in Table 1.

Table 1. Reactive Distillation Column - Internal specifications (Sánchez et al., 2020).

Parameter	Value
Diameter [m]	0.0762
Number of reactive stages (5-8)	4
Number of non-reactive stages (1-4, 9-14)	10
Stripping zone length per stage [m]	1.1
Rectifying zone length per stage [m]	1
Catalyst (Katapak SP-11 of Sulzer)	Amberlyst 70
Operating pressure [kPa]	74.66

The thermodynamic information is presented by Garcia et al. (2021) and the NRTL binary parameters and kinetic model parameters are presented by Velandia et al. (2021). The model evaluation for the system of lactic acid (LA), butanol (BuOH), butyl lactate (BuLac) and water (W) was performed using Matlab/Simulink R2020b. The model was compared against Aspen Plus V11.

Figure 3 a) shows the composition profiles along the column. Results obtained with the developed Matlab/Simulink R2020b model are compared against Aspen Plus V11 simulator. In both cases, the results are in agreement and present the same tendency. With the presented operation conditions, BuLac with a purity of 97% can be obtained, while the BuOH is maintained in excess along the column. Figure 3 b) shows the temperature profile, where the model simulated in Matlab/Simulink shows good agreement with the results obtained in Aspen Plus.

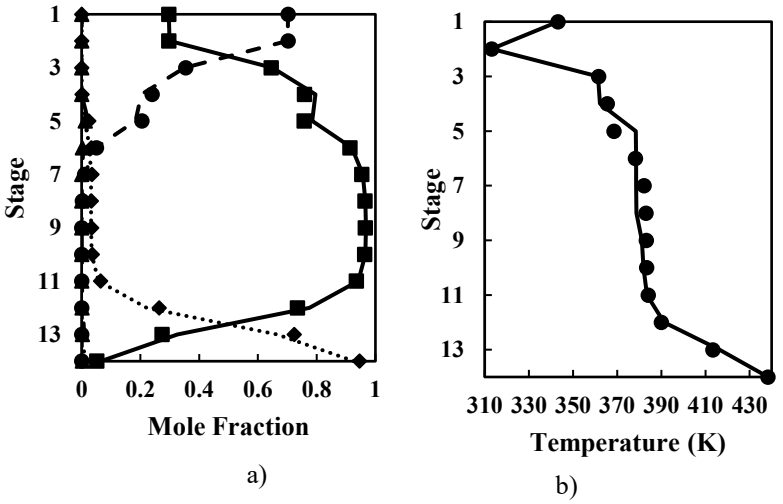


Figure 3. a) Mole fraction profiles: Aspen Plus V11 (●) W, (■) ButOH, (▲) LA, (♦) BuLac and Matlab/Simulink R2020b (**---**) W, (**-**) ButOH, (**-·-**) LA, (···) BuLac. b) Temperature profile: Aspen Plus V11 (●) and Matlab/Simulink R2020b (**-**).

4. Conclusions

Butyl lactate can be obtained in the reactive distillation column at pilot plant scale of Universidad Nacional de Colombia. In both simulations, the obtained feeding flow rates are in accordance with the technical specifications, which means that these are the nominal values of operation. The obtained BuLac, with a purity higher than 97%, follows product specifications for commercial purposes. It is important to compare these simulations against experimental information and this will be performed in a future work.

Reference

EPA, U. (2021). United States Environmental Protection Agency, Volatile Organic Compounds' Impact on Indoor Air Quality. https://www.epa.gov/indoor-air-quality-iaq/volatile-organic-compounds-impact-indoor-air-quality

Garcia, C., Velandia, J., Céspedes, M., Rodríguez, G., & Gil, I. D. (2021). Isobaric Vapor−Liquid Equilibrium for the Binary Mixture of 1-Butanol + Butyl L-Lactate at 1 and 5 kPa. Journal of Chemical and Engineering Data, 1–6.

Grand View Research. (2016). Green & Bio-based Solvents Market Size, Share & Trends Analysis Report By Product. (Lactate Esters, Methyl Soyate), By Application (Paints & Coatings, Adhesives).

Häckl, K., & Kunz, W. (2018). Some aspects of green solvents. Comptes Rendus Chimie, 21(6), 572–580.

Kumar, R., & Mahajani, S. M. (2007). Esterification of lactic acid with n-butanol by reactive distillation. Industrial and Engineering Chemistry Research, 46(21), 6873–6882.

Li, Z., Smith, K. H., & Stevens, G. W. (2016). The use of environmentally sustainable bio-derived solvents in solvent extraction applications - A review. Chinese Journal of Chemical Engineering, 24(2).

Peña-Tejedor, S., Murga, R., Sanz, M. T., & Beltrán, S. (2005). Vapor-liquid equilibria and excess volumes of the binary systems ethanol + ethyl lactate, isopropanol + isopropyl lactate and n-butanol + n-butyl lactate at 101.325 kPa. Fluid Phase Equilibria.

Sánchez, C. A., Gil, I. D., & Rodríguez, G. (2020). Fluid phase equilibria for the isoamyl acetate production by reactive distillation. Fluid Phase Equilibria, 2–30.

Velandia, J. J., García, C. A., Céspedes, M. A., Rodríguez, G., & Gil, I. D. (2021). Reactive and non-reactive residue curve maps analysis to produce Butyl Lactate by catalytic distillation. Chemical Engineering and Processing - Process Intensification, 168.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50175-5

Optimal Design of Offshore Wind Power Farm Considering Wind Uncertainty

Sunwoo Kim[a], Seongwhan Kang[b], Jay H. Lee[a,*]

[a]*Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology (KAIST), 291, Daehak-ro, Yuseong-gu, Daejeon, 34141, Republic of Korea*
[b]*R&D Campus Daejeon, LG Energy Solution (LGES), 188, Munji-ro, Yuseong-gu, Daejeon, 34122, Republic of Korea*
jayhlee@kaist.ac.kr

Abstract

This study presents an optimization-based framework for the design of offshore wind power farms (OWFs). First, the main characteristics of the OWF facilities are decided including the size and type of the turbines and the capacity of the inter-array cables. Second, a feasible location for installing the facility is identified by considering various geographical characteristics such as average wind speed and water depth, with the aid of a geographic information system (GIS). Finally, we construct a mixed-integer fractional programming (MIFP) optimization model to decide on the exact locations of the wind turbines and the offshore substations as well as the routes of the inter-array cables, by minimizing the unit cost which is the total installation cost divided by total power generated. A case study is conducted based on the 6th regional energy plan of Jeju island, South Korea, which plans to construct two 100 MW and one 125 MW OWFs. The optimized unit cost of the OWFs turns out to be 2.3 MM$/MW, which is at par with existing OWFs. Furthermore, we introduce a stochastic programming model to consider various potential wind scenarios directly in the design. The robust design obtained via stochastic programming approach showed an improved average unit cost by ~5% compared to the nominal design which assumed a specific scenario of wind speed and direction.

Keywords: Offshore Wind Power Farm (OWF), Geographical Information System (GIS), Stochastic Programming

1. Introduction

Offshore wind power farms (OWFs) are becoming popular, with GW-scale projects appearing in many parts of the world, owing to its high potential in terms of scale and efficiency. The OWF market is expected to grow significantly, i.e., from 23 GW in 2018 to 228 GW in 2030 which is about a ten-fold increase. However, the high install and operating cost of OWF is still a barrier to more widespread adoption. Since the overall cost of OWF is greatly affected by its geographical characteristics (e.g., wind speed, water depth) as well as its design (e.g., spatial arrangement of the turbines, routes of the inter-array cables), the location selection and design must be carried out carefully, both for economic feasibility evaluation and actual installation. Also, the intermittency of wind speed depending on season and weather must be precisely considered for the optimal design of OWF.

There have been relatively few studies that addressed the design of OWFs using GIS under considering wind uncertainty. Recently, a mixed-integer linear programming (MILP) optimization model that decides routes of the inter-array cables and locations of the substations was suggested and applied to the Navitus Bay Windpark in the U.K by Pillai et al. (2015). However, the locations of the wind turbines, which affect the installation cost and power generation greatly, were not considered in this study. Spyridonidou et al. (2020) suggested a design framework of OWFs which uses GIS where the case study is performed in Greece and, sixteen candidate areas for constructing OWF were selected considering geographical criteria. However, the task of determining the optimal layout of the OWF was not addressed in this study. Also, Banzo et al. (2011) suggested a stochastic optimization model to decide on the location of the substation and the cross-sections of cable connection considering wind uncertainty and applied the model to a case set in Hrons Rev OWF in Denmark. However, this study did not address the micro-siting problem of locating the turbines at the site.

To overcome the limitations of previous research, this paper suggests a stochastic programming based design framework of OWF using GIS and considering the wind variability/uncertainty. We screen for candidate locations of OWF utilizing GIS. Then an optimization model is formulated for each candidate location where the main decision variables are locations of the turbines and the offshore substations, and routes of the inter-array cable and the objective are to minimize the average unit cost over possible scenarios of the wind speed/direction.

This paper is organized as follows. The design framework for OWF is explained in section 2. Then the detailed formulation of the optimization is presented in section 3. The results of the case study involving the design of OWFs in Jeju Island, Korea are presented in section 4. Some conclusions are given in section 5.

2. Problem definition

Our research aims to develop an optimization based framework for designing OWFs while considering relevant geographical information and wind uncertainty. The consideration of geographical characteristics is important as they significantly affect the bottom-line economics of OWFs. Previously, Kang et al. (2020) presented a three-stage framework for designing a biorefinery, which consists of the steps of facility design, GIS analyses, and optimization model development. Firstly, key characteristics of the facility such as the sizes of the main processing units and main yield parameters are specified. Then, important geographical information such as the land use and climates are considered to select candidate locations. Finally, the specified facility and geographical information are used to construct a mathematical optimization problem that carries out specific designs for economic evaluations.

In our research, the above three-stage framework is adapted for the design of OWFs. In the first step, we select 5 MW wind turbines found in the reference by NREL. The power curve of the turbine and the installation costs of the various items of the facility are specified. In the second step, candidate locations for installing the facilities are identified with the aid of GIS analyses. Key geographical information such as water depth, wind speed, and distance data are collected to be used as input data for the optimization model. Also, we quantify the wake effect, the reduction in the wind speed as the wind passes through other turbines at each candidate location. Lastly, the objective function of the optimization model is defined. In our case, the unit cost, which is the total installation cost divided by the total power generation ($/kWh), is to be minimized. Finally, the

optimization was solved to obtain the optimal OWF design under 200 wind scenarios generated from the stochastic model. To analyze the economics, another 200 wind scenarios were generated to compare the average unit costs among the different designs.

3. Problem formulation

The power curve of the wind turbine is approximated as a piecewise linear function of wind speed as in Eq.(1) where a is the slope, b is the y-intercept, $WS_{i,a}$ is the wind speed and $P_{i,a}$ is the power generated at location i under scenario a. The installation costs of the turbine and the substation are estimated using Eq.(2) and Eq.(3), respectively, which were proposed by Gonzalez et al. (2017), where P^{total} is the total installed power capacity of the OWF (MW), N is the number of turbines installed in the OWF, WD is the water depth, P_i is the capacity of each turbine (MW), and *A, B, C, D* are constants. The capacity and installation cost of the inter-array cables are summarized in Table 1.

$$P_{i,a} = a \times WS_{i,a} + b \tag{1}$$

$$C^{turbine}(k\$) = A \times \frac{(P^{total})^{0.87}}{N} + (B \times WD^2 - C \times WD + D) \times P_i \tag{2}$$

$$C^{sub}(k\$) = 3040.8 + 106.4 \times P^{total} \tag{3}$$

Table 1. Installation cost and capacity of electric cables

Cable type	Capacity (Maximum number of turbines)	Capital cost ($/meter)
Inter-array cable	5	135
	7	250
	9	370
Export cable (Offshore)	25	885
Export cable (Onshore)	-	762

Secondly, the distance data of the wind turbines are obtained from GIS analysis and reduced wind speed by wake effect is calculated by Eq.(4), where $dc_{i,j,a}$ represents how much the wind speed of turbine at location i get reduced due to the adjacent turbine j and X_i is the binary variable representing whether the turbine is installed or not.

$$WS^d_{i,a} = WS_{i,a} \times X_i - \sum_j dc_{i,j,a} \times X_j \tag{4}$$

For stochastic optimization, wind speed scenarios are generated as following procedures. First, at each observatory near the candidate location, the deviation in the annual average wind speed compared to the average value of the data for the 12 years was calculated as a percentage value. Second, the calculated value of deviation is discretized into 10 steps based on the maximum and minimum value of the wind speed deviation. Then, the number of deviation data corresponding to each deviation step is counted to get the possibility distribution $Step_a$, Finally, the wind speed of scenario a is generated by randomly selecting one of the deviation steps from $Step_a$ and random variable $\xi_{i,a}$ is

used to fill the gap of possible wind speed range between each step randomly as presented in Eq.(5).

$$WS_{i,a} = \overline{WS_i} * (1 + Step_a + \xi_{i,a}) \quad (5)$$

Finally, a MIFP optimization problem is constructed to decide the locations of the substation, wind turbines, and the routes of the inter-array cables. The objective is to minimize the unit cost which represents described in Eq.(6), where $C^{facility}$ is the installation cost of each facility type and $P_{i,a}$ is the power generated by the turbine on location i under scenario a. Unit cost represents how much power is generated from the initial investment capital.

$$\min \frac{C^{total}}{Ptotal} = \frac{C^{turbine} + C^{sub} + C^{cable}}{\sum_{i,t} P_{i,t}} \quad (6)$$

The major constraints of our problem are the power flow balance and those arising from the spatial limitation. Energy flow $F_{i,j}$ is defined as the number of turbines connected through the inter-array cable. Notice that the offshore substation has only inflow energy, and the turbines have one more inflow of energy than outflows, which is expressed as Eq.(7). For each location, at most one facility can be constructed, which is expressed as Eq.(8), where $S_i^{100\,MW}$ and $S_i^{125\,MW}$ are the binary variables representing whether or not a substation of the particular scale is installed.

$$\sum_{i,j \neq i} F_{i,j} - F_{j,i} = X_i - 20 \times S_i^{100\,MW} - 25 \times S_i^{125\,MW} \quad (7)$$

$$X_i + S_i^{100\,MW} + S_i^{125\,MW} \leq 1 \quad (8)$$

4. Case study

According to the 6th regional energy plan of Jeju Island, two 100 MW and one 125 MW OWFs are being planned for construction. We use the three-stage framework to design the OWFs in Jeju Island according to this plan. In our case study, the main decision variables are the locations of the substation and wind turbines, and the routes of the inter-array cables. The main assumptions are summarized as below:

- The wind speed is represented by its annual average value.
- There is no transmission loss through an electric cable.
- Seven times the blade diameter is the minimum allowed distance between turbines, (882 m)
- Length of the onshore export cable is 20% longer than the Euclidean distance

To consider wind uncertainty, 200 scenarios of wind speed are generated and used to design the optimal OWF (multi-scenario design) and compared with the design where only the average wind speed is used (nominal-scenario design). Another 200 scenarios of wind speed are generated for economic analyses of each OWF design.

In Figure 1, the difference between the multi-scenario design and the nominal scenario design is illustrated. Although the average value of wind speed is the same, as the specifics of the wind speed differ for each scenario, the average amount of power generated is different. This is because the power curve shows varying slopes for different

wind speed ranges. The different OWF layouts lead to different economic results. In all zones, the multi-scenario design showed improvements over the nominal scenario design. The detailed economic results are presented in Table 2, where the multi-scenario design shows ~5% lower average unit cost compared to the nominal scenario design, and the percentage of the scenarios showing improvements is about 92%. In all zones, the use of 100 MW units resulted in a lower unit cost. Since the number of candidate locations is ~40, 100 MW units presented more options than 125 MW units which required only 21 installations. The comparison of the unit cost is given in Figure 2. As a result, we can conclude constructing 100 MW OWF in zone 1 and zone 2, and constructing 125 MW in zone 3 is the best solution for meeting the plan.

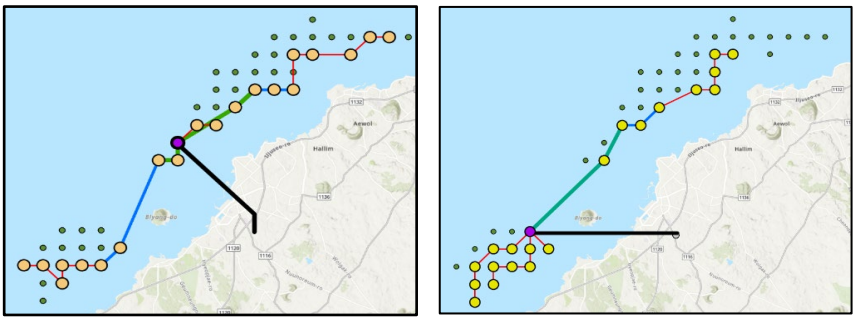


Figure 1. Optimized layout of the OWF in zone 1 (a) nominal scenario design (b) multi-scenario design

Table 2. Detailed economic comparison between the nominal scenario design and the multi-scenario design

Capacity	Zone	Scenario type	Average C^{total} (MM$)	Average P^{total} (GWh)	Average C^{total}/P^{total} ($/MWh)	Percentage of improved scenarios (%)
100MW	1	Nominal	223.1	230.4	968.4	7.5
		Multiple	231.4	242.4	954.5	92.5
	2	Nominal	228.2	245.8	928.4	7.5
		Multiple	231.6	257.3	900.0	92.5
	3	Nominal	223.6	289.7	771.8	8.5
		Multiple	205.3	281.4	729.6	91.5
125MW	1	Nominal	288.0	290.4	991.7	7.5
		Multiple	283.5	303.6	933.8	92.5
	2	Nominal	283.7	294.8	962.5	7.5
		Multiple	294.3	315.7	932.1	92.5
	3	Nominal	286.0	360.8	792.7	8.5
		Multiple	261.1	351.2	743.6	91.5

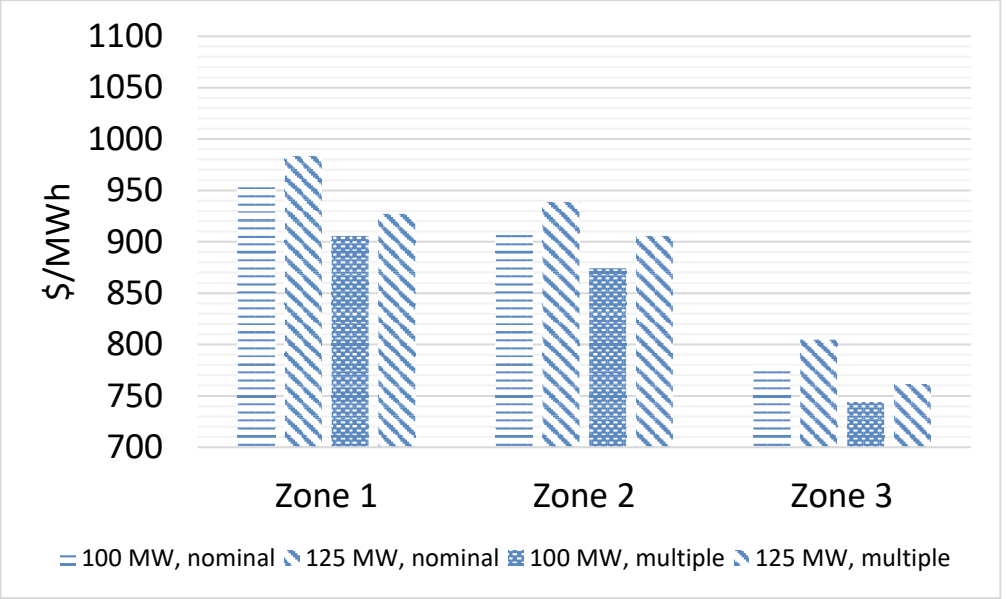


Figure 2. Comparison of the unit cost between the nominal scenario OWF and the multi-scenario OWF

5. Conclusions

In this paper, we adapt the three-stage framework originally developed for biorefinery design to the design of optimal OWFs considering wind uncertainty. Candidate locations are selected by using GIS and important geographical data from GIS are used in the optimization model. MIFP is constructed to design OWFs according to the 6th regional energy plan in Jeju Island as a case study. The economics of two different designs, a nominal scenario design and a multi-scenario design using stochastic programming were analyzed. In all candidate zones, the multi-scenario design showed improved economics with the average unit cost lower by ~5%, and improvements were shown for about 92% of the scenarios tried. For future research, the operation of designed OWFs will be addressed where various dispatch and storage decisions are made to manage energy supply and demand while addressing the intermittency and uncertainty of wind speed.

References

S. Kang et al. Three-stage design of high-resolution microalgae-based biofuel supply chain using geographic information system. *Applied Energy* 265 (2020): 114773.

M. Blanco et al., The economics of wind energy. *Renewable and sustainable energy reviews* 13.6-7 (2009): 1372-1382.Reviews, 13, Science Direct, Amsterdam, Netherlands 2009, 1372-1382.

A. Pillai et al., Offshore wind farm electrical cable layout optimization. *Engineering Optimization*. 2015;47:1689-708.

S. Spyridonidou et al., Strategic Planning of Offshore Wind Farms in Greece. *Sustainability*. 2020;12.

M. Banzo et al.,. Stochastic optimization model for electric power system planning of offshore wind farms. *IEEE Transactions on Power System*s Aug. 2011; 26(3): 1338–1348.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50176-7

Economic and environmental impact of fouling in produced water re-injection

Otavio Fonseca Ivo[a*], Lars Struen Imsland[a]

[a] *Department of Engineering Cybernetics, Norwegian University of Science and Technology, O. S. Bragstads plass 2, 7034 Trondheim, Norway*

fonseca.i.otavio@ntnu.no

Abstract

Produced water re-injection (PWRI) is an enhanced oil recovery technique which aims to increase the recovery factor of a reservoir while reducing produced water (PW) discharges to the ocean. Re-injection of PW is challenging as fouling in pipelines can lead to inefficient operation of the PWRI facility. In this work, we perform sensitivity analysis at a PWRI facility. As inputs, we have selected the riser roughness coefficient and processed PW flowrate. Furthermore, we consider two different re-injection strategies and assess the impact of varying those inputs over key economic and environmental performance indicators. Results show the negative effect that fouling has over several performance indicators. Furthermore, they suggest that one should constantly re-evaluate the fouling level at the pipelines of a PWRI facility before performing decision-making for production optimization.

Keywords: Sensitivity analysis, Waterflooding, Produced water re-injection, Production optimization, Fouling

1. Introduction

Produced water (PW) is by far the largest waste stream in the offshore oil and gas industry (Neff *et al.*, 2011). PW can be defined as formation water and/or water that has previously being injected into the formation. Due to its nature, PW is a complex mixture that contains dispersed oil, dissolved organic compounds, solid particles, and bacteria (Nasiri *et al.*, 2017). Moreover, it is considered a continuous source of contaminants to the continental shelf ecosystems (Beyer *et al.*, 2020). To increase operational environmental safety level, it is advised that operators follow the best practices for management of PW. These guidelines are grounded in four pillars: avoid, prevent, reduce, and treat (Miljødirektoratet, 2015).

To reduce marine pollution, several Norwegian operators started using waterflooding as a primary PW management solution (Steinar *et al.*, 2016). Waterflooding is an enhanced oil recovery (EOR) technique in which water is injected at a reservoir formation to displace residual oil towards production wells, increasing oil recovery factors (Bautista and Dahi Teleghani, 2017). However, continuous re-injection of PW can lead to fouling of scale-prone or tar-like (*i.e.* "schmoo") substances in the produced water re-injection (PWRI) facility (Bader, 2007), leading to inefficient operation (Rossini *et al.*, 2020). Inspired by

operational issues encountered at a real offshore PWRI facility, we investigate the economic and environmental impact of fouling in waterflooding operation.

2. Motivating example

The studied PWRI facility is represented in Figure 1. During normal operation, PW enters the system through the separation vessel. From the vessel, PW can be sent to either the ocean discharge and/or the pumping station (PS). In the PS, there are recycle systems located downstream of each pump to prevent inadequate pumping operation. When PW enters the pumping system, it is split between two fixed-speed pumps (FSPs). Downstream of the FSPs, the non-recycled PW is mixed and is boosted by the variable-speed pump (VSP). Downstream of the VSP, throttling valves regulate the passage of PW for re-injection in the re-injection templates (RITs) α and β. The connections between the PS and the RITs are performed by risers. Each RIT has choke valves which regulate the shared re-injection ratio between each re-injection well.

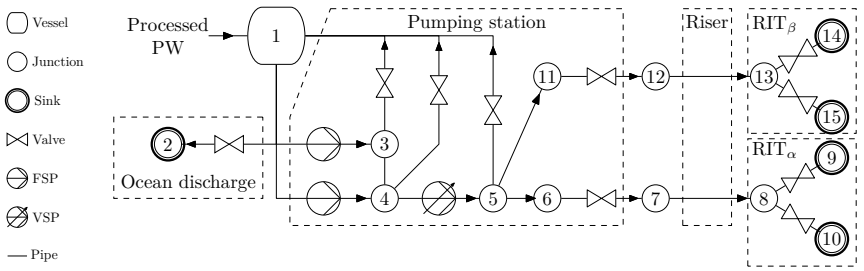


Figure 1: Produced water re-injection facility.

The PWRI model considered in this work is based on Ivo and Imsland (2021). It is desired to maximize the revenue (R) obtained during operation of the PWRI facility, which is given by:

$$\begin{aligned} R = &\$_{oil} \left(\lambda_\alpha \sum_{i=9}^{10} d_i + \lambda_\beta \sum_{i=14}^{15} d_i \right) \\ &- (\$_{fuel} + \$_{CO_2} E_d) \left(W^s_{(1,2)} + W^s_{(1,3)} + W^s_{(4,5)} \right), \end{aligned} \tag{1}$$

with constants $\$_{oil}$ as the market oil price; $\$_{fuel}$ as the fuel price; $\$_{CO_2}$ as the carbon tax; λ_α and λ_β as respectively the re-injection effectiveness of RIT$_\alpha$ and RIT$_\beta$; and E_d as the fuel energy content. Furthermore, $W^s_{(1,2)}$, $W^s_{(1,3)}$ and $W^s_{(4,5)}$ are the pumping shaft-power; and d_i is the PW demand at node i.

In addition, we consider key environmental performance indicators which are divided into two segments. The first is related to PW management, in which re-injection rate accounts for percentage of PW re-injected; and waste-to-value accounts for the conversion of re-injected PW to oil produced. These performance indicators are represented below:

$$RR = \left(\sum_{i=9}^{10} d_i + \sum_{i=14}^{15} d_i \right) / |d_1| \tag{2a}$$

$$WtV = \left(\lambda_\alpha \sum_{i=9}^{10} d_i + \lambda_\beta \sum_{i=14}^{15} d_i \right) / |d_1| \tag{2b}$$

where RR is the re-injection rate; and WtV is the waste-to-value conversion. The second environmental segment is related to CO_2 emissions, in which the total emission of CO_2 accounts for total emission; the CO_2 production intensity considers the emission rate relative to oil production; and CO_2 re-injection intensity contemplate the emission rate relative to PW re-injection. These CO_2 emission indicators are shown below:

$$T_{CO_2} = E_f \, E_d \left(W^s_{(1,2)} + W^s_{(1,3)} + W_{(4,5)} \right) \tag{3a}$$

$$I_{oil} = T_{CO_2} / \left(\lambda_\alpha \sum_{i=9}^{10} d_i + \lambda_\beta \sum_{i=14}^{15} d_i \right) \tag{3b}$$

$$I_{PW} = T_{CO_2} / \left(\sum_{i=9}^{10} d_i + \sum_{i=14}^{15} d_i \right) \tag{3c}$$

where T_{CO_2} is the total CO_2 emission; I_{oil} is the CO_2 production intensity; I_{PW} is the CO_2 re-injection intensity; and E_f is the CO_2 emission factor.

To perform sensitivity analysis in the PWRI facility, a parametric optimization problem was formulated. The problem is represented by the following nonlinear program (NLP):

$$\min_{\boldsymbol{\theta}} \quad \Phi(\boldsymbol{\theta}, \mathbf{p}), \tag{4a}$$

$$s.t. \quad \mathbf{g}(\boldsymbol{\theta}, \mathbf{p}) = 0, \tag{4b}$$

$$\mathbf{h}(\boldsymbol{\theta}, \mathbf{p}) \leq 0, \tag{4c}$$

where $\Phi : \mathbb{R}^{n_\theta} \times \mathbb{R}^{n_p} \to \mathbb{R}$ is the objective function; $\mathbf{g} : \mathbb{R}^{n_\theta} \times \mathbb{R}^{n_p} \to \mathbb{R}^{n_g}$ is the vector of equality constraints; and $\mathbf{h} : \mathbb{R}^{n_\theta} \times \mathbb{R}^{n_p} \to \mathbb{R}^{n_h}$ is the vector of inequality constraints. Moreover, $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta}$ is the vector of decision variables; and $\mathbf{p} \in \mathbb{R}^{n_p}$ is the vector of parameters. The compact form shown in Eq.(4) enables one to perform sensitivity analysis to study how optimal decision variables $\boldsymbol{\theta}^*(\mathbf{p})$ and optimal objective function $\Phi^*(\mathbf{p})$ behaves due to changes performed in parameters.

For the considered application, the sensitivity analysis was done by varying the PW entering the PWRI facility with $d_1 \in [-480, -1818]$ m^3/h. Moreover, we vary the roughness coefficient in the Hazen-Williams equation of riser$_{(7,8)}$ to represent an increase in flow resistance due to fouling, $C_{(7,8)} \in [140, 80]$. Notice that lower values of $C_{(7,8)}$ translates into more friction. The interval of interest for these parameters were based in operational data gathered from an offshore producer. Due to that, we analyze two strategies: the exclusive re-injection strategy (ERI), and the shared re-injection strategy (SRI). In ERI, re-injection should be performed exclusively at RIT$_\alpha$, while in SRI re-injection can be performed at both RITs.

3. Results and Discussion

3.1. Operational Behavior

Understanding the operational behavior of the PWRI facility is crucial for assessing the system performance. We showcase in Figure 2a the VSP operation; and in Figures 2b

and 2c the PW re-injected at the RITs. Each curve is associated with a roughness coefficient value and a strategy.

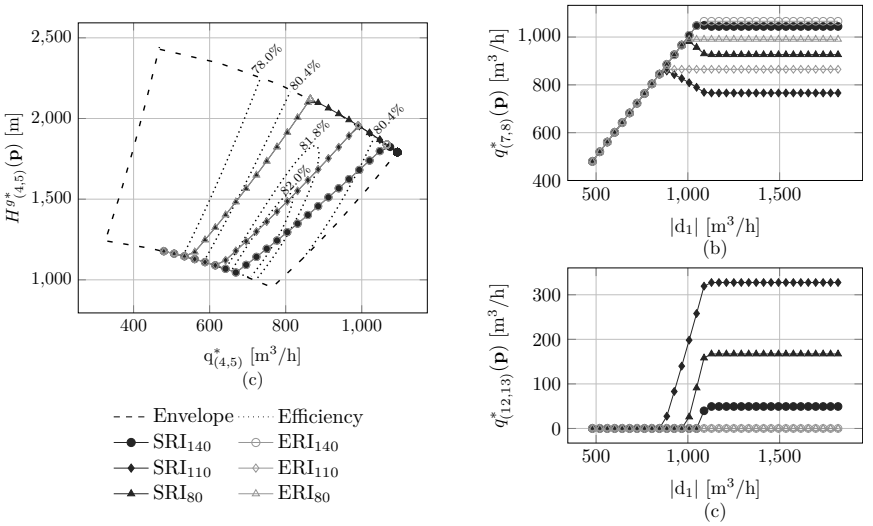


Figure 2: Overview of the PWRI facility operation for different roughness coefficients of the Hazen-William equation, in which: (a) Head gain of the VSP with operational envelope and efficiency islands; (b) re-injection flowrate of PW to the RIT$_\alpha$; and (c) re-injection flowrate of PW to the RIT$_\beta$.

The operational point of a VSP is given by the intersection of the system resistance and pumping curves (Gülich, 2008). For the lowest value of $|d_1|$, 480 m^3/h, the operation point of the VSP is at the minimum rotation and $q^*_{(4,5)}$ is also at 480 m^3/h. Once $|d_1|$ increases, valves$_{(6,7),(8,9),(8,10)}$ widen their opening to re-inject more PW at RIT$_\alpha$. This dislocates the system resistance curve to the right until these valves reach their maximum opening. At this point, the resistance curve in the ERI strategy cannot be further dislocated to the right. As for the SRI strategy, valves$_{(11,12),(13,14),(13,15)}$ can be opened to decrease the system resistance. However, due to the low effectiveness of RIT$_\beta$, it is instead optimal to re-injection more PW to RIT$_\alpha$ by increasing the VSP rotation. This tendency last for both strategies, until the VSP reaches its maximum operational rotation. For the ERI strategy an increase in re-injection cannot be achieved as the system lacks degrees of freedom, which represents a process bottleneck. For the SRI strategy, the valves associated with RIT$_\beta$ start to open until $q^*_{(4,5)}$ reaches the VSP envelope limit. This point represents a process bottleneck for the SRI strategy as the VSP reaches its operational limits.

3.2. Economic and environmental performance

The economic and environmental performance of each strategy with distinct roughness coefficient are shown in Figure 3. Overall, fouling has impacted negatively most of the considered indicators for both strategies at some interval of $|d_1|$. The only exception is the

re-injection ratio of the SRI strategy, as shown by the overlapping curves $\text{SRI}_{140,110,80}$ in Figure 3b.

While the VSP operates at minimum rotation, it is not observed any significant impact of fouling in the studied indicators. Nevertheless, when the VSP operates strictly inside its operational envelope, it is seen that fouling impacts negatively the total CO_2 emission in Figure 3d, CO_2 production intensity in Figure 3e, and CO_2 re-injection intensity in Figure 3f. This behavior is seen as fouling increases hydraulic losses in the $\text{riser}_{(7,8)}$, requiring additional usage of fuel to re-injection the same quantity of PW in RIT_α, which increase CO_2 emissions.

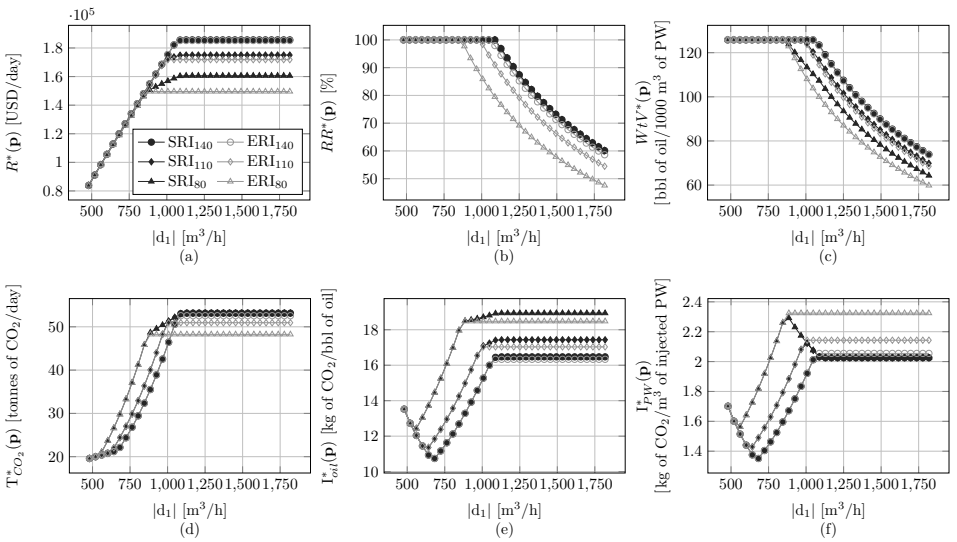


Figure 3: Economic and environmental performance for distinct strategies with different roughness coefficients of the Hazen-William equation. (a) Total revenue; (b) re-injection ratio; (c) waste-to-value; (d) total CO_2 emission; (e) CO_2 production intensity; and (f) CO_2 re-injection intensity.

As the VSP operates at maximum rotation, it is possible to see that fouling has worsened mostly indicators for the ERI strategy, except for total CO_2 emission in Figure 3d, which improves as less PW is re-injected at RIT_α. For the SRI strategy, fouling causes worsening of several indicators, except for re-injection ratio in Figure 3b, total CO_2 emission in Figure 3d, and CO_2 re-injection intensity in Figure 3f, in which differences due to fouling cannot be observed.

4. Conclusion

In this work, we have shown that fouling in PW re-injection can impact significantly the economic and environmental performance of a PWRI facility. Furthermore, it was shown that a RIT with lower effectiveness has the potential to mitigate several issues caused by fouling if included in a decision-making framework. However, it is important to notice

that scaling can also occur at other risers during operation. Thus, monitoring of scaling and revaluation of decisions should be regularly performed for improving short and long term decision-making over waterflooding operation.

5. Acknowledgement

This research is a part of BRU21 – NTNU Research and Innovation Program on Digital and Automation Solutions for the Oil and Gas Industry (www.ntnu.edu/bru21) and supported by OKEA ASA. In addition, we would like to acknowledge license partners Neptune Energy Norge AS, and Petoro AS for allowing the usage of field data. The authors would like to thank employees at OKEA Per Magne Bjellvåg, Alf Sebastian Lackner, and Raymond Hellerud for their valuable input.

References

J. Beyer, A. Goksøyr, D.Ø. Hjermann, and J. Klungsøyr, 2020, Environmental effects of offshore produced water discharges: A review focused on the Norwegian continental shelf, Marine environmental research.

J. Neff, K. Lee, and E. DeBlois, 2011, Produced Water: Overview of Composition, Fates, and Effects, Produced Water: Environmental Risks and Advances in Mitigation Technologies, New York, NY, USA.

J.F. Bautista, and A. Dahi Taleghani, 2017, The state of the art and challenges in geomechanical modeling of injector wells: a review paper, Journal of Energy Resources Technology, 139, 1.

J.F. Gülich, 2008, Centrifugal pumps, Berlin, Germany.

Miljødirektoratet, 2015, Utredning av beste tilgjengelige teknikker (BAT) for rensing av produsert vann som slippes ut fra petroleumsvirksomheten til havs, Norway.

M.S.H. Bader, 2007, Seawater versus produced water in oil-fields water injection operations, Desalination, 208, 1-3, 159-168.

M. Nasiri, I. Jafari, and B. Parniankhoy, 2017, Oil and gas produced water management: a review of treatment technologies, challenges, and opportunities, Chemical engineering communications, 204, 8, 990-1005.

N. Steinar, G. Eimund, and D. Egil, 2016, Produced Water Management Under the Norwegian "Zero Harmful Discharge Regime" – Benefits With the Risk Based Approach, Society of Petroleum Engineers - SPE International Conference and Exhibition on Health, Safety, Security, Environment, and Social Responsibility, 1-9.

O.F. Ivo, L.S. Imsland, 2021, Analysis of Optimal Control Strategies for Efficient Operation of a Produced Water Reinjection Facility for Mature Fields, IFAC-PapersOnLine, 54, 3, 115-121

S. Rossini, G. Roppoli, P. Mariotti, S. Renna, M. Manotti, A. Viareggio, and L. Biassoni, 2020, Produced water quality impact on injection performance: Predicting injectivity decline for waterflood design, International Petroleum Technology Conference.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50177-9

Mathematical Modelling of Reactive Inks for Additive Manufacturing of Charged Membranes

Xinhong Liu[a], Riju De[b], Alexander Pérez[c], John R. Hoffman[a], William A. Phillip[a], Alexander W. Dowling[a*]

[a]*Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN 46556, USA*
[b]*Department of Chemical Engineering Goa, BITS Pilani, Goa 403726, India*
[c]*Department of Mechanical Engineering, University of Puerto Rico-Mayagüez, Mayagüez, PR 00681, USA*
adowling@nd.edu

Abstract

Patterned charged membranes with engendered useful characteristics can offer selective transport of electrolytes. Chemical patterning across the membrane surface via a physical inkjet deposition process requires precise control of the reactive-ink formulation, which enables the introduction of charged functionality to the membrane. This study develops a new dynamic mathematical model for the primary step of the batch reactive-ink formulation considering an ink mixture of copper sulphate and ascorbic acid. Nonlinear least squares parameter estimation is performed to infer three kinetic model parameters by analysing data from nine dynamic experiments simultaneously. Global sensitivity and Fisher information matrix (FIM) analyses reveal only one kinetic parameter is identifiable from time-series pH measurements. The fitted model can capture the overall nonlinear dynamics of the batch reaction and works best for initial Cu^{2+} concentrations between 30 and 50 mM. Time-series Cu^{2+} or Cu^{+} concentration measurements are recommended in future experiments to elucidate the kinetics of reactive-ink formulation.

Keywords: Sensitivity analysis, Parameter estimation, Additive manufacturing, Fisher information matrix, Nanostructured membrane, Data science

1. Introduction

A critical need for more selective membranes has emerged as their applications in modern separations and sensing processes continue to expand. To this end, membranes with chemically patterned surfaces such as charge-patterned mosaics and Janus membranes are an emerging approach for accessing higher selectivity (Qu et al., 2017). The state-of-the-art fabrication process of such membranes involves printing reactive inks on the nanostructured substrates, where the charged functionality is induced through the copper-catalysed azide-alkyne cycloaddition (CuAAC) reaction mechanism. Hoffman et al. (2019) demonstrated that this approach requires precise control of the reactive-ink formulation to ensure the membrane manufacturing process remains in the transport-limited regime. The reactive-ink formulation comprises of a sequence of reactions occurring between reactants, i.e., copper sulphate ($CuSO_4$), ascorbic acid (DH_2), and an alkyne, to facilitate the formation of the dinuclear-copper alkyne complex (DNCAC). The concentration of the DNCAC is essential for the CuAAC reaction to proceed and therefore should be optimized to achieve a smooth chemical patterning while ensuring a controlled charged functionalization rate of the membrane at the same time (Worrell et al., 2013).

Focusing on the primary step which involves the reaction between $CuSO_4$ and DH_2, there are only a few models available in the literature that explains the mechanism of the DH_2-$CuSO_4$ reaction (Shtamm et al., 1979; Xu and Jordan, 1990). However, these models are limited to either the final Cu^{2+} reduction or the Cu^{+}/hydronium ions (H^{+}) formation. None have discussed a comprehensive and detailed rate law-based model that captures all the species involved in the reaction to the best of our knowledge.

In this paper, we develop a first-principles-based kinetic model for reactive ink formulation. Considering the anaerobic reaction scheme for the ascorbic acid dissociation as reported in the literature, rate constants that are unavailable in the literature are posed as model parameters and related to known equilibrium constants from Shtamm et al. (1979). Parameter estimation is performed to identify kinetic rate constants from time-series pH data. Local and global sensitivity analyses are performed to determine which model parameters are identifiable and design future experiments.

2. Dynamic mathematical model for reactive-ink formulation

The simplest reactive-ink solution is an aqueous mixture of ascorbic acid and copper sulphate. The reaction scheme involves a sequential dissociation of ascorbic acid via two reversible reactions in equilibrium as given by Eqs. (I), (II) and a final forward reaction as in Eq. (III), which describes the conversion of Cu^{2+} to Cu^{+}. Shtamm et al. (1979) proposed the following reaction scheme for the reduction of Cu^{2+} to Cu^{+}:

$$DH_2 \underset{k_{b1}}{\overset{k_{f1}}{\Leftrightarrow}} DH^- + H^+ \tag{I}$$

$$Cu^{2+} + DH^- \underset{k_{b2}}{\overset{k_{f2}}{\Leftrightarrow}} CuDH^+ \tag{II}$$

$$Cu^{2+} + CuDH^+ \overset{k_{i1}}{\rightarrow} 2Cu^+ + D + H^+ \tag{III}$$

The equilibrium constants for reaction (I) and reaction (II) reported by Shtamm et al. (1979) are $K_{e1} = {k_{f1}}/{k_{b1}} = 5 \times 10^{-5}$ M and $K_{e2} = {k_{f2}}/{k_{b2}} = 200$ M^{-1}, respectively. Here, k_{f1} and k_{f2} denote the forward rate constant, whereas k_{b1} and k_{b2} symbolize the backward rate constants for reactions I and II, respectively. The rate constant for the final reaction (III) reported by Shtamm et al. (1979) is $k_{i1} = 6000$ $mol^{-1}{\cdot}L{\cdot}min^{-1}$. We consider a fully dynamic kinetic model for reactions (I) to (III) with five differential variables and five ordinary differential equations:

$$\frac{d[DH_2]}{dt} = -k_{b1}K_{e1}[DH_2] + k_{b1}[DH^-][H^+] \tag{1}$$

$$\frac{d[DH^-]}{dt} = k_{b1}K_{e1}[DH_2] - k_{b1}[DH^-][H^+] - k_{b2}K_{e2}[DH^-][Cu^{2+}] + k_{b2}[CuDH^+] \tag{2}$$

$$\frac{d[H^+]}{dt} = k_{b1}K_{e1}[DH_2] - k_{b1}[DH^-][H^+] + k_{i1}[CuDH^+][Cu^{2+}] \quad (3)$$

$$\frac{d[Cu^{2+}]}{dt} = -k_{b2}K_{e2}[DH^-][Cu^{2+}] + k_{b2}[CuDH^+] - k_{i1}[CuDH^+][Cu^{2+}] \quad (4)$$

$$\frac{d[CuDH^+]}{dt} = k_{b2}K_{e2}[DH^-][Cu^{2+}] - k_{b2}[CuDH^+] - k_{i1}[CuDH^+][Cu^{2+}] \quad (5)$$

Here, $[DH_2]$, $[DH^-]$, $[H^+]$, $[Cu^{2+}]$, and $[CuDH^+]$ denote the concentrations of ascorbic acid, anion, hydrogen ion, cupric ion, and the complex, respectively, in unit of M. The model includes three unknown kinetic parameters, k_{b1} [mol^{-1}·L·min^{-1}], k_{b2} [min^{-1}] and k_{i1} [mol^{-1}·L·min^{-1}], that characterize the reaction progress.

3. Materials and methods

3.1. Materials

Initially, 200 mM ascorbic acid (DH_2) and copper sulphate ($CuSO_4$) solutions were prepared separately in 5 mL sample vials. Subsequently, the solutions were mixed to form 10 mL solutions. The pH of the mixtures was measured at every 1 min time interval for 5 min. The procedure was repeated in triplicate with $CuSO_4$ solutions at 200, 150, 100, 66.7, 50.0, 40.0, 33.3, 28.6, 25.0 mM for a total of (3 x 9 =) 27 experiments.

3.2. Parameter estimation with multi-start

To estimate the three unknown kinetic model parameters, $\boldsymbol{\theta} = \{\boldsymbol{k_{b1}}, \boldsymbol{k_{b2}}, \boldsymbol{k_{i1}}\}$, we solve the least-square nonlinear regression problem, shown in Eq. (6), which minimizes the square of the difference between the model predicted and the experimentally measured concentration of H^+ ions (residuals) for the 9 experimental conditions.

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n_{exp}} \sum_{j=1}^{n_{tri}} \left([H^+]_{pred,i,j} - [H^+]_{exp,i,j}\right)^2 \quad (6)$$

Eq. (6) is solved numerically using scipy.optimize in Python with a customized multi-start strategy to help find the (near) global optima.

3.3. Local and global sensitivity analysis

The Fisher information matrix (FIM), which is defined as the inverse of the parameter covariance matrix, is computed at the best fit parameter values $\widehat{\boldsymbol{\theta}}$ based on local sensitivities of the model predictions to each parameter. The eigendecomposition of the FIM reveals which parameters are identifiable (Rothenberg and Thomas, 1971).

Similarly, a global sensitivity analysis is conducted via grid search by evaluating the objective function over a wide range of model parameters: k_{b1} = [400, 900] mol^{-1}·L·min^{-1}, k_{b2} = [1E-05, 1] min^{-1} and k_{i1} = [1, 6000] mol^{-1}·L·min^{-1}.

4. Results and discussion

4.1. Parameter estimation with multi-start

Table 1 shows the parameters estimated from regressing time-series pH measurements from the 27 experiments simultaneously. When each dataset (experimental condition) is regressed independently, parameters k_{b1} varies from 260 to 1,100,00 $mol^{-1} \cdot L \cdot min^{-1}$ and k_{i1} varies from 12 to 6,000 $mol^{-1} \cdot L \cdot min^{-1}$. These non-unique parameters motivate both simultaneous regression (Table 1) as well as formal identifiability analysis.

Table 1. Parameter estimation and local sensitivity from analysing 9 experiments simultaneously.

Estimated parameters			Residuals squared	Eigenvalues of FIM	Eigenvectors of FIM		
k_{b1}	k_{b2}	k_{i1}			k_{b1}	k_{b2}	k_{i1}
($mol^{-1} \cdot L \cdot min^{-1}$)	(min^{-1})	($mol^{-1} \cdot L \cdot min^{-1}$)	(M^2)				
				2.18E-12	-1.00	1.55E-09	4.40E-04
17601.30	0.0074	105.69	68.62	2.74E-06	-4.40E-04	8.35E-07	-1.00
				2.84E+06	-1.92E-09	-1.00	-8.35E-07

4.2. Local and global sensitivity analysis

The eigenvalues and eigenvectors of the Fisher information matrix (FIM) are also reported in Table 1. Two eigenvalues are near zero, 2.18E-12 and 2.74E-06, which implies the FIM is near singular and model is partial non-identifiable. The corresponding eigenvectors are predominantly in the direction of k_{b1} and k_{i1}, respectively, which indicate these parameters cannot be reliably estimated from these data. Conversely, the eigenvector of the largest eigenvalue, 2.84E+06, is in the direction of k_{b2}. This difference of more than 12 orders of magnitude in eigenvalues implies only k_{b2} is identifiable based on the pH measurements from 9 experimental conditions considered in triplicate.

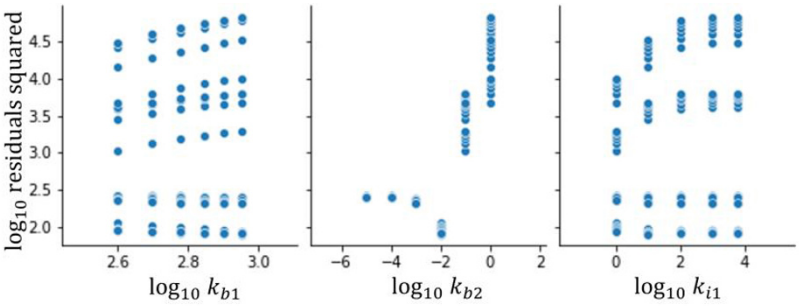


Fig. 1. Global sensitivity of the $\log_{10}$-transformed sum of residuals squared for three model parameters.

Fig. 1 shows the results from a global sensitivity analysis which confirm that k_{b2} is the most sensitive model parameter. Parameters k_{b1}, k_{b2} and k_{i1} are varied with a grid search, and the $\log_{10}$-transformed sum of residuals squared for the 27 experiments are computed. The plots with respect to k_{b1} and k_{i1} show the sum of residuals squared is only slightly impacted by the value of k_{b1} and k_{i1}. In other words, for a constant value of k_{b1} or k_{i1}, the sum of residuals squared changes by two orders of magnitude when varying the other two parameters. In a contrast, varying k_{b2} causes one to four orders of

magnitude changing in the sum of residuals squared. Moreover, the sum of residuals squared is minimized around $k_{b2} = 10^{-2}$ min^{-1}. For $k_{b2} \leq 10^{-2}$ min^{-1}, varying k_{b1} or k_{i1} has almost no impact on the sum of residuals squared, but the model becomes sensitive to $k_{b1}$and k_{i1} when $k_{b2} \geq 10^{-2}$ min^{-1}. A possible physical explanation for this result is that reaction (II) is rate limiting when $k_{b2} \leq 10^{-2}$ min^{-1}, hence the model predictions are not sensitive to the other reaction kinetic constants (provided they are sufficiently large to not be rate limiting).

4.3. Partial parameter estimation

The local and global sensitivity analyses above both show only k_{b2} can be reliably estimated from the available timeseries pH measurements from the 27 experiments. Since k_{b1} characterize the rate for the ascorbic acid dissociation, it only affects the beginning of Cu^{+} formation and has little influence on the subsequent reaction progress. Therefore, we select $k_{b1} = 600$ mol^{-1}·L·min^{-1} for reasonable H^{+} profiles and k_{i1}= 6000 mol^{-1}·L·min^{-1} based on literature (Shtamm et al., 1979).

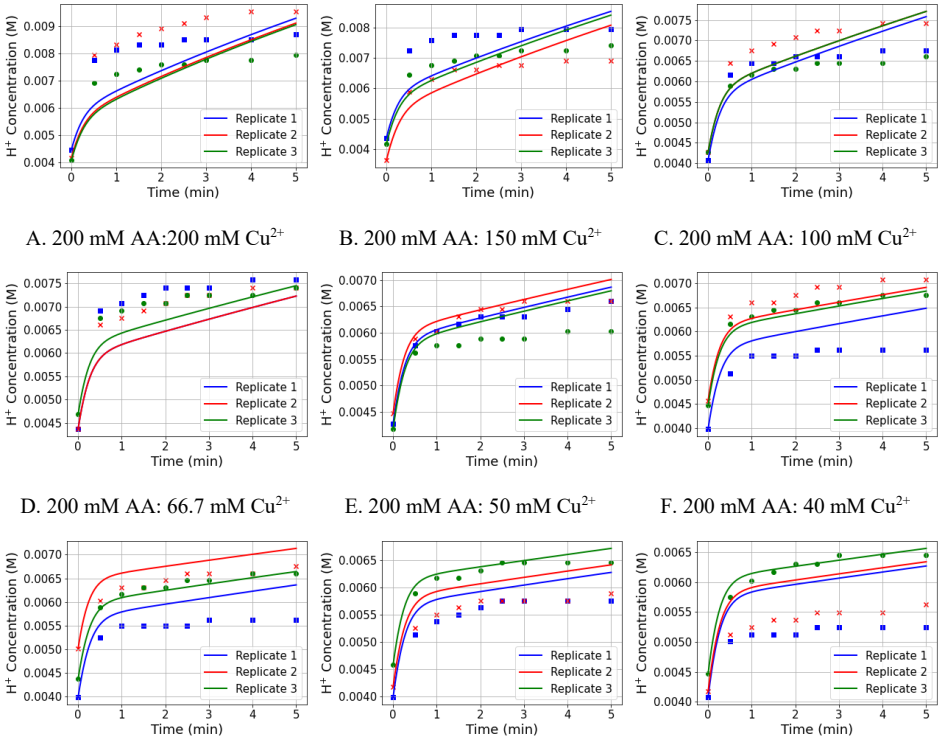


Fig. 2. Comparison for model predictions and experimental pH measurements for the 27 experiments (9 initial concentrations A-I repeated in triplicate).

Fig. 2 shows the results from partial parameter estimation with k_{b1} and k_{i1} fixed. The best fit value for k_{b2} is 0.0082 ± 0.0002 min^{-1}. As seen in Fig. 2, the continuous H^{+} concentrations predicted from the fitted model match the overall the reaction progress as measured by the timeseries pH data. Experiments with initial Cu^{2+} concentrations between

28.6 and 50 mM (Fig. 2E-H) are best predicted with their relative sum of residuals squared less than 4%. However, for the experiments conducted with higher $[Cu^{2+}]_0$, the model overestimates the H^+ concentration after 5 minutes (Fig. 2A-D). Also, the relative sum of residuals squared increases monotonically from 7 % to 44 % as $[Cu^{2+}]_0$ increases from 66.7 to 200 mM, respectively. Similarly, for $[Cu^{2+}]_0 = 25.0$ mM, the model overestimates the H^+ concentration and the relative sum of residuals squared equals 11%. Together, these results suggest that copper complexes, which are not modeled, are important.

5. Conclusions

We developed a mathematical model for the Cu-DH_2 reaction system. Local and global sensitivity analysis reveals that only reaction rate parameter, k_{b2}, can be identified from timeseries pH measurements. Partial parameter estimation is performed to estimate k_{b2}, which is, to our knowledge, one of the first reported values for k_{b2} in literature.

While the fitted model captures the overall nonlinear dynamics of the batch reaction, several extensions should be considered as future work. We hypothesize measuring timeseries Cu^{2+} or Cu^+ concentrations will enable some of the remaining model parameters to be identified. Likewise, adding intermediate chemical species such as complexes to the model may improve the quality of fit. Ultimately, a predicted dynamic model, such as the one presented in this paper, may be used to optimize the ink formulations for additive manufacturing of chemically patterned membranes.

Acknowledgments

We gratefully acknowledge support from the United States National Science Foundation (NSF) awards CMMI-1932206 and CBET-1941596. A.P. and J.R.H. thank Notre Dame for support from the NDnano Undergraduate Research Fellowship and the Patrick and Jana Eilers Graduate Student Fellowship for Energy Related Research, respectively.

References

J.R. Hoffman, A.D. Mikes, F. Gao and W.A. Phillip, 2019, Controlled postassembly functionalization of mesoporous copolymer membranes informed by fourier transform infrared spectroscopy, ACS Appl. Polym. Mater, 1, 8, 2120-2130, 10.1021/acsapm.9b00419

S. Qu, Y. Shi, S. Benavides, A. Hunter, H. Gao and W.A. Phillip, 2017, Copolymer nanofilters with charge-patterned domains for enhanced electrolyte transport, Chem. Mater, 29, 2, 762-772, 10.1021/acs.chemmater.6b04660

T.J. Rothenberg, J. Thomas, 1971, Identification in parametric models. Econometrica, 39, 577–592, 10.2307/1913267

E.V. Shtamm, A.P. Purmal and Y.I. Skurlatov, 1979, Mechanism of catalytic ascorbic acid oxidation system Cu2+-ascorbic acid-O2, Int. J. Chem. Kinet, 11, 5, 461-494, 10.1002/kin.550110503

B.T. Worrell, J.A. Malik and V.V. Fokin, 2013, Direct evidence of a dinuclear copper intermediate in Cu(I)-catalyzed azide-alkyne cycloadditions, Science, 340, 6131, 457-460, 10.1126/science.1229506

J. Xu and R.B. Jordan, 1990, Kinetics and mechanism of the reaction of aqueous copper(II) with ascorbic acid, Inorg. Chem, 29, 16, 2933-2936, 10.1021/ic00341a015

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50178-0

Economic Analysis of a Hydrogen Liquefaction Process Based on Techno-Economic and Energy Optimization

Heechang Son[a], Bjørn Austbø[b], Truls Gundersen[b], Jihyun Hwang[c, *], Youngsub Lim[a, d *]

[a]*Department of Naval Architecture and Ocean Engineering, Seoul National University, Seoul, 08826, Republic of Korea*
[b]*Department of Energy and Process Engineering, Norwegian University of Science and Technology (NTNU), NO-7491, Trondheim, Norway*
[c]*Scool of energy technology, Korea Institute of Energy Technology, Naju, 58217, Republic of Korea*
[d]*Research Institute of Marine Systems Engineering, Seoul National University, Seoul 08826, Republic of Korea*
jihyun.hwang@kentech.ac.kr
s98thesb@snu.ac.kr

Abstract

Efforts to be eco-friendly are becoming essential, and hydrogen is attracting attention as an eco-friendly fuel. As the hydrogen demand increases, the interest in liquid hydrogen is increasing because it is safer and more efficient for large-scale hydrogen transport and storage than compressed gaseous hydrogen. Since hydrogen must be cooled to about −253 ℃ to be liquefied, this represents a high proportion of the cost in a hydrogen liquefaction plant. Reducing the unit price of hydrogen is the key to achieve a hydrogen economy, and it is also important to reduce the cost of the hydrogen liquefaction process for price competitiveness of liquid hydrogen. The purpose of this study, therefore, is to perform an economic analysis of a hydrogen liquefaction process based on both techno-economic and energy-based optimization. The objective functions for techno-economic optimization and energy optimization are total annualized cost and total power consumption, respectively. The results show that the influence of operating expenditure is greater than that of capital expenditures, and the cost of the compressors is dominant in the capital expenditures. These results provide technical background data that can be used to improve the economic viability of the hydrogen liquefaction process as an important element in the hydrogen economy.

Keywords: Hydrogen liquefaction process; Techno-economic optimization; Energy optimization; Economic analysis.

1. Introduction

Eco-friendly efforts to solve the climate crisis are becoming an essential factor, leading to an era of great energy transition. Accordingly, hydrogen is in the spotlight as a promising fuel in the future because of its advantages: 1) abundant amount, 2) higher gravimetric energy density than conventional fossil fuels, and 3) eco-friendly fuel with no carbon emission (Yin and Ju, 2020).

The hydrogen value chain can be divided into three areas (production, storage/transportation, and utilization), and various research and development efforts are underway to lower the unit price of hydrogen in each area. In the hydrogen storage/transportation stage, three methods are being discussed as promising means to replace the conventional compressed hydrogen gas in order to increase the storage and transportation efficiency; i.e. Liquid organic hydrogen carrier (LOHC), Ammonia (NH_3) and Liquid hydrogen (LH_2). LOHC is a method of storing hydrogen in a liquid compound, and methylcyclohexane (MCH) is often considered as a candidate for this. Similarly, NH_3 synthesis is widely considered as a method for storing and transporting hydrogen. In LH_2, the hydrogen is stored and transported in liquid form, similar to the method of liquefied natural gas (LNG).

Since LH_2 has the high volumetric energy density compared to the gaseous hydrogen and the high gravimetric energy density compared to the LOHC and NH_3, it is expected to be the most suitable hydrogen storage method in large-capacity scenarios. However, liquefying hydrogen is technically challenging.

First, hydrogen is liquefied at −253 ℃, and therefore a harsher cryogenic environment than LNG is required. Accordingly, a hydrogen liquefaction process requires about 30 times higher energy consumption compared to a natural gas liquefaction process (e.g. typical specific energy consumption (SEC) of natural gas liquefaction is about 0.3 kWh/kg LNG, while a typical SEC of hydrogen liquefaction is about 10 kWh/kg LH_2). In addition, hydrogen molecules occur in two different spin isomers, orthohydrogen (ortho-H_2) and parahydrogen (para-H_2). Normal hydrogen (normal-H_2) that can be observed at room temperature is composed of 75 % ortho-H_2 and 25 % para-H_2. As hydrogen is cooled from room temperature, the ratio of spin isomers forming equilibrium hydrogen (equilibrium-H_2) changes. As the temperature is reduced, ortho-H_2 is converted into para-H_2, which is an exothermic process. Since the natural ortho-/para-H_2 conversion (OPC) process occurs relatively slowly, liquefied non-equilibrium-H_2 can be vaporized by the heat of conversion unless the OPC processes are rapidly achieved through catalytic reactions while cooling hydrogen. This is because the heat of conversion (about 1063 J/mol) is higher than the heat of vaporization of liquid hydrogen (about 954 J/mol) (Zhuzhgov et al., 2018). Therefore, in an actual hydrogen liquefaction plant, a catalyst-filled heat exchanger is used to achieve a catalytic reaction for rapid OPC, such that the conversion heat can be removed in the cooling process.

Currently, much effort is made to lower the unit price of hydrogen to achieve a hydrogen economy, and thereby it is important to reduce the cost of the hydrogen liquefaction process as much as possible in order to have price competitiveness for liquid hydrogen. This study, therefore, aims to present technical background data for an economical hydrogen liquefaction process design through economic analysis based on techno-economic and energy optimization.

2. Process description

The Claude cycle presented by Berstad et al. (2021) is selected as a base model for a hydrogen liquefaction process in this study. Figure 1 shows the process flow diagram of the hydrogen liquefaction process in consideration. First, hydrogen feed gas is introduced in the precooling cycle at a rate of 125 ton/day at 20 bar and 298.15 K. The hydrogen feed gas is cooled to 114 K by passing PC-HX in the precooling cycle and then enter the main cryogenic cycle. Here, it is assumed that any impurities are removed

through an adsorber before entering the main cryogenic cycle. A mixed refrigerant (MR) is used for the precooling cycle. The MR is composed of nitrogen (N_2), methane (C_1), ethane (C_2), propane (C_3), *n*-butane (*n*-C_4), and *i*-pentane (*i*-C_5). The MR is compressed to about 35 bar by two compression stages with intercoolers. It is assumed that the adiabatic efficiencies of each compressor and the pump in the precooling cycle are 85 % and 75 %, respectively. The high-pressure MR is pre-cooled to 114 K by passing PC-HX and then expanded to low pressure by a Joule-Thomson (J-T) valve bringing the stream to a lower temperature. The resulting low-temperature MR stream passes through PC-HX, cooling the hydrogen, and returns to the compression system.

After passing through an ortho-/para-H_2 conversion unit, the hydrogen gas is cooled from 117.9 K to 106 K in MC-HX2. The resulting hydrogen stream undergoes OPC through another conversion unit. In the simulation, equivalent heat exchangers using conversion reactor units are applied to simulate catalyst-filled heat exchangers for the OPCs as shown in Figure 1. The hydrogen is cooled to 22 K through HX3-HX8 with the OPCs and then expanded to 1.5 bar by a J-T valve. The final LH_2 conditions are 1.5 bar, 21.7 K, and 97 % para-LH_2 concentration. In the main cryogenic cycle, the normal-H_2 refrigerant is compressed to about 30 bar by four compression stages with intercooling in between. It is assumed that the adiabatic efficiency of C-3 and C-4 and the adiabatic efficiency of C-5 and C-6 are 82 % and 85 %, respectively. Side streams from the resulting refrigerant are drawn and expanded to intermediate pressures by expanders. The isentropic efficiency of each expander is assumed to be 85 %. The main refrigerant stream from MC-HX5 is depressurized by a liquid expander and a J-T valve. The resulting refrigerant stream at about 21 K passes through all heat exchangers and is returned to the compression system to provide the cooling duty. The side streams are mixed with the returning main refrigerant stream at intermediate pressure. The process parameters applied in the simulation are shown in Table 1.

Process modeling was done using Aspen HYSYS® V10.0. A modified Benedict-Webb-Rubin equation of state (MBWR) and the Peng-Robinson equation of state (PR) were applied to the main cryogenic cycle and precooling cycle, respectively.

3. Process optimization

For energy optimization, the net power consumption was used as the objective function, as shown in Eq.(1). The net power consumption can be calculated by subtracting the power produced in all *j* expander stages from the power consumed in all *i* compressor stages.

For techno-economic optimization, the cost estimation methodology presented by Turton et al. (2008) was used to calculate equipment cost, module cost and capital expenditures (*CAPEX*) of the hydrogen liquefaction process. Table 2 shows the equipment types considered here and their corresponding cost parameters. In this study, different types of heat exchangers for the precooling cycle and the main cryogenic cycle were selected in consideration of the type of refrigerant used. The cost of the catalyst used was assumed to be equal to 15 % of the cost of the heat exchanger without catalyst, and this assumption was used to calculate the cost of equivalent heat exchangers to which the OPCs are applied. The calculated *CAPEX* can be converted to an estimated annual *CAPEX* as shown in Eq.(2). Here, it was assumed that the interest rate (*i*) and the lifetime of the plant (*L*) are 10 % and 20 years, respectively. In addition, operating expenditures (*OPEX*) were estimated considering the total power consumption and the

electricity cost (c_{elec} = 0.06 USD/kWh) as shown in Eq.(3). Consequently, the total annualized cost (*TAC*) of the hydrogen liquefaction process can be calculated by summing annual *CAPEX* and *OPEX*. This was used as the objective function for the techno-economic optimization, as shown in Eq.(4).

$$\min \sum \dot{W}_{\text{total}} = \left(\sum_i \dot{W}_i - \sum_j \dot{W}_j \right) \tag{1}$$

$$\text{Annual } CAPEX = CAPEX \cdot \frac{(i * (1+i)^L)}{(1+i)^L - 1} \tag{2}$$

$$OPEX = c_{\text{elec}} \sum \dot{W}_{\text{total}} \tag{3}$$

$$\min TAC = \text{Annual } CAPEX + OPEX \tag{4}$$

Here, a genetic algorithm (GA) was used for the optimization, and the process optimization was performed by connecting the GA loaded from MATLAB R2021a with Aspen HYSYS®. Optimization variables include the discharge pressure of each compressor and the expansion pressure of each expander, as well as the MR flow rate and normal-H_2 flow rate that are the flow rates of refrigerants.

Table 1. Process parameters used in the simulations

Parameter	Value	Unit
Hydrogen feed pressure	20	bar
Hydrogen feed temperature	298.15	K
Hydrogen feed mass flow	125	t/d
Inlet fraction of para-H_2	25	mol %
Liquid hydrogen pressure	1.5	bar
Liquid hydrogen temperature	21.6	K
Outlet fraction of para-H_2	≥ 95	mol %
Adiabatic efficiency of compressors	82–85	%
Isentropic efficiency of expanders	85	%
Minimum temperature approach of heat exchangers	0.5	K
Intercooler temperature	298.15	K

Table 2. Types and cost parameters for each equipment in the hydrogen liquefaction process

Equipment	Type	Cost parameters		
		K_1	K_2	K_3
Main cryogenic heat exchanger	Flat plate	4.6656	-0.1557	0.1547
Precooling heat exchanger	Spiral tube	3.9912	0.0668	0.243
Compressor	Centrifugal	2.2897	1.3604	-0.1027
Expander	Axial gas turbines	2.7051	1.4398	-0.1776

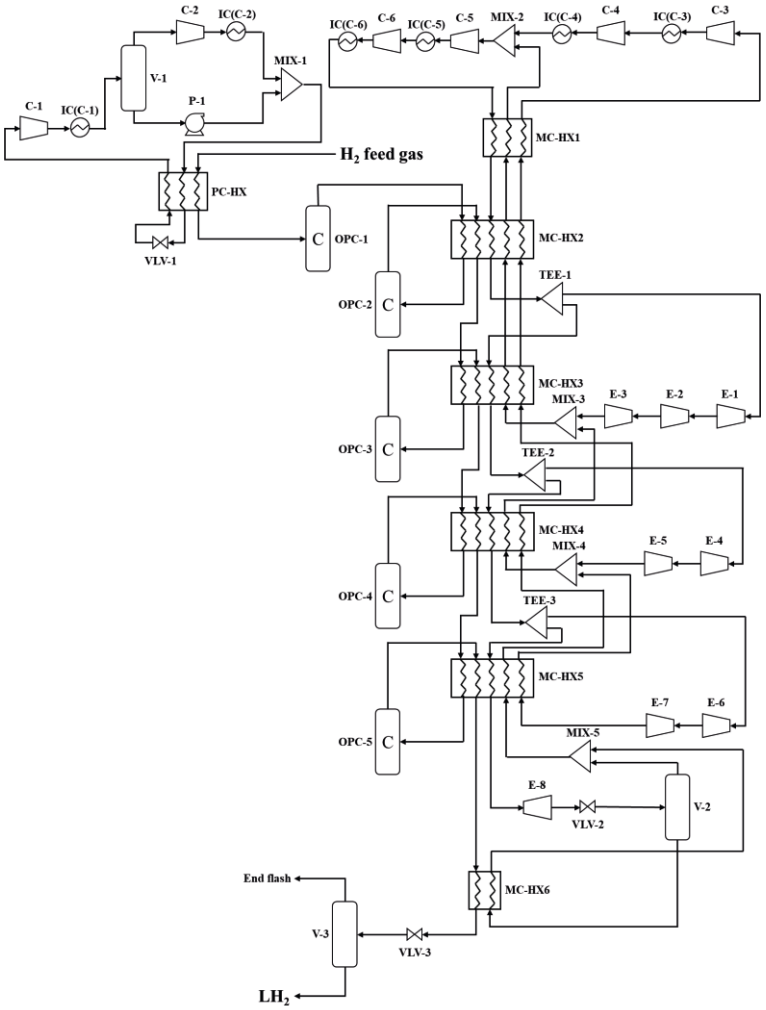


Figure 1. Process flow diagram of the hydrogen liquefaction process

4. Results

The economic analyses of the energy and techno-economic optimization results as well as the base model are shown in Figure 2. Since the hydrogen liquefaction process consumes a lot of energy, the influence of *OPEX* in *TAC* is greater than that of *CAPEX*, and therefore the compressor cost is dominant in the *CAPEX*. However, in the case of energy optimization, the *TAC* is larger compared to the base model and techno-economic optimization because the heat exchanger capacity is excessively increased.

As a result, for the hydrogen liquefaction process, it is advantageous in terms of *TAC* to increase process efficiency and save *OPEX*, but care must be taken not to excessively increase the capacity of the heat exchangers during energy optimization.

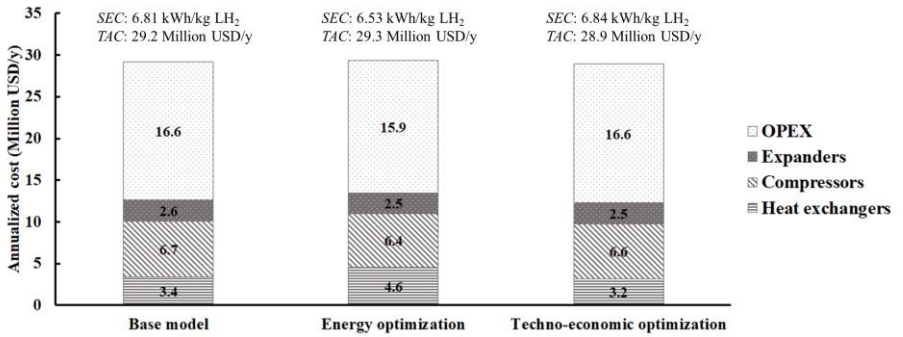


Figure 2. Economic analyses for base model, and energy and techno-economic optimization

5. Discussion

In this study, values for the overall heat transfer coefficient (U values) were assumed when estimating the heat exchanger cost in the techno-economic optimization; i.e. U value for the main cryogenic heat exchanger was set to 3000 W/m^2·°C, while the U value for the precooling heat exchanger was set to 5000 W/m^2·°C. Since these values are taken from natural gas liquefaction process data, future studies are required to apply more appropriate U values for the hydrogen liquefaction process. In addition, since the liquid expander (E-8) cost-wise was treated in the same way as the gas expanders, this also needs to be improved in future studies.

6. Conclusions

An economic analysis of a hydrogen liquefaction process was performed based on techno-economic and energy optimization. The total annualized costs were calculated by applying the cost estimation methodology selected for the techno-economic optimization. In conclusion, the influence of *OPEX* on the hydrogen liquefaction process cost is high, and thereby it is important to increase the process efficiency. In addition, if energy optimization is used for the hydrogen liquefaction process design, it should be noted that the heat exchanger capacity can be excessively increased leading to an increase in *TAC*.

References

D. Berstad, G. Skaugen, and Ø. Wilhelmsen, 2021, Dissecting the exergy balance of a hydrogen liquefier: Analysis of a scaled-up claude hydrogen liquefier with mixed refrigerant pre-cooling, International Journal of Hydrogen Energy, 46, 8014–8029.

R. Turton, R. C. Bailie, W. B. Whiting, and J. A. Shaeiwitz, 2008, Cost Equations and Curves for the CAPCOST Program in Analysis, synthesis and design of chemical processes, 3rd ed., London, UK: Pearson Education, 923–954.

L. Yin and Y. Ju, 2020, Process optimization and analysis of a novel hydrogen liquefaction cycle, International Journal of Refrigeration, 110, 219–230.

A. V. Zhuzhgov, O. P. Krivoruchko, L. A. Isupova, O. N. Mart'yanov, and V. N. Parmon, 2018, Low-Temperature Conversion of *ortho*-Hydrogen into Liquid *para*-Hydrogen: Process and Catalysts. Review, Catalysis in Industry, 10, 1, 9–19.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50179-2

How Digital Twins are Propelling Metals Industry to Next Generation Decision-Making: A Practitioner's View

Yale Zhang[a*], Mitren Sukhram[a] and Ian Cameron[a]

[a] *Hatch Ltd., 2800 Speakman Drive, Mississauga, Ontario Canada L5K 2R7*
[*] *yale.zhang@hatch.com*

Abstract

The digital twin is a technology to digitally transform asset lifecycle in the metals industry, from improving project delivery to empowering operational intelligence toward next-generation decision-making. In this paper, Hatch's digital twin framework is presented and demonstrated using a real-world blast furnace twin example, followed by development practice and lessons learned from our practice experience.

Keywords: Digital Twin, Analytics, Decision-making, Mining and Metals, Blast Furnace

1. Introduction

The metals industry is concerned with the processing of bulk ore resources into various basic materials, such as steel, aluminum, copper, etc. It is an asset-centric industry, where key assets are heavily invested and have their own lifecycle. With the growing maturity of the Industrial Internet of Things (IIoT), Machine Learning and Artificial Intelligence (ML/AI) technologies, the digital twin becomes a powerful tool to transform the industrial asset lifecycle from process design, construction to operation. It drives the metals industry toward better decision-making to improve project delivery, yield and resource efficiency, as well as plant operation and maintenance.

At Hatch, the digital twin is often considered under a bigger picture of industrial asset lifecycle and enterprise value chain. As shown in Figure 1, it first contextualizes the asset information during the process design, engineering, and construction phase. Such information like asset structure, Piping and Instrumentation Diagrams (P&IDs), design basis, equipment information, etc. are all connected to static and dynamic process models as well as discrete event simulation models to optimize project design, Capex/Opex, and schedule. The dynamic process models (with possible expansion to include thermo- and fluid dynamics) are further integrated with a control system emulator to build digital twins for operator training and accelerate the process commissioning phase. Once processes are in operation, the above models together with data-driven, ML/AI algorithms are used to develop digital twins to support operational intelligence, for example, "process twins" for anomaly detection, scenario analysis, and operational guidance; "asset twins" for predictive maintenance, and risk management; and "connected twins" considering integrated mass flow, energy flow, cash flow, and carbon flow for end-to-end value chain simulation and optimization.

This paper focuses on the digital twins of operating assets. It starts with a review of three selected digital twin frameworks and then presents a commercial digital twin use case for

ironmaking blast furnaces. Hatch's digital twin development approach and lessons learned are also shared from an industrial practitioner's perspective.

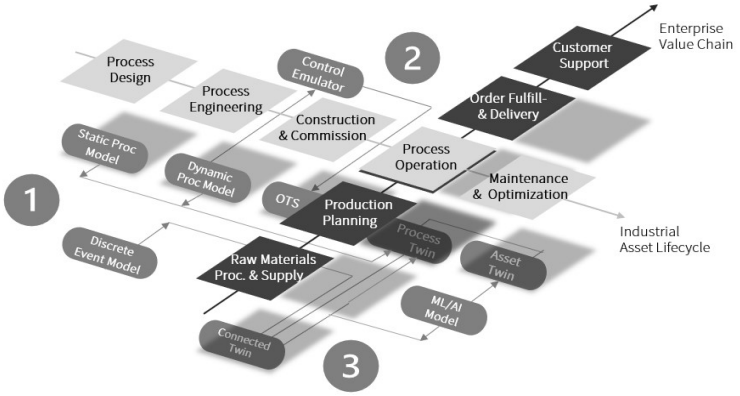


Figure 1. Digital twins across industrial asset lifecycle and enterprise value chain

2. Understand Digital Twin Technology

The digital twin concept has been around for many years, and it is always evolving with the latest development of Industry 4.0 technologies. Many different digital twin definitions and frameworks have been presented recently by technology vendors and industrial practitioners. Three representative digital twin frameworks are selected and discussed in this section. Each of them has its own foci, but they share some common components, which are crucial to digital twin design and development.

Deloitte is one of the early practitioners in applying digital twin technology to help their manufacturing clients create new values. In its definition, the digital twin framework is a near real-time comprehensive integration between the physical and digital worlds (Mussomeli, *et al.*, 2020). Such integration relies on four key components: sensors and actuators from the physical side, and data and analytics from the digital side. Sensors are distributed throughout the industry processes, continuously generating real-time data to represent the current situation of physical assets. The data are ingested into a digital twin platform and consumed by analytical algorithms and models to produce actionable insights. The insights can be implemented by actuators to apply direct changes to physical processes in an automatic or semi-automatic way. DNV is an independent assurance and risk management organization, who proposed a unique framework by decomposing a digital twin into multiple manageable functional elements (DNV-GL, 2020). Each element contains five key components: (1) data streams, (2) asset information model, (3) computation models, (4) dashboard and quality indicator, and (5) need and key decision. These components normally serve one specific purpose (i.e., use case) but may also be shared across multiple elements. In DNV's view, the digital twin is a collection of functional elements, which work together holistically to provide decision support over the entire life cycle of a physical asset. Hatch 3I's digital twin framework is presented in Figure 2 (Zhang, *et al*. 2021), which consists of Integration, Intelligence, and Interaction.

Integration is to establish a dynamic visualization environment, where all sources of engineering, production, maintenance, environment, health, and safety data together with high-fidelity 2D/3D asset models are combined to create a "Single Source of Truth". It

presents on-demand, contextualized information and ensures data transparency and availability for all designated users through a secured cloud infrastructure.

Intelligence is the key differentiator between the previous generation of BI dashboards and a digital twin. The latter provides a fully and meaningfully connected analytical platform to effectively combine first-principle models, big data, and ML/AI technology to generate actionable insights and make intelligent decisions related to the physical twin's safety, reliability, efficiency, and profitability.

Interaction focuses on value connectivity. A digital twin adds value to business by creating innovative and more meaningful points of interaction between humans and machines through rich visualization and services, for example, augmented operator training, remote expert support, and collaborative decision-making.

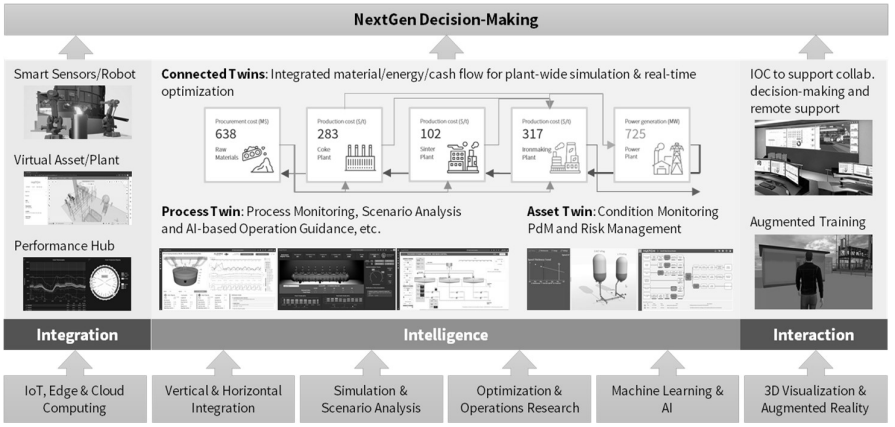


Figure 2. Hatch 3I's Framework of Digital Twin

Some common components of digital twin are evident based on the above frameworks and should be considered for every digital twin implementation. They are: (1) **Data** is a fundamental component of digital twins. Effective decision-making requires insights based on true and contextualized data from multiple sources. It becomes critically important that a well-designed data platform is utilized by digital twins for data connection, ingestion, manipulation, and storage. (2) **Model** is the core of digital twins. The models based on first principles, ML/AI algorithms, simulation, and/or optimization technology are used in different digital twin use cases to generate insights and make decisions for specific purposes. (3) **Visualization** is the aspect that describes the user interface of a digital twin. It represents actionable insights through various technologies, from widely used business intelligence dashboards, 3D visualization, to the latest virtual and augmented reality.

3. An Industrial Example: Blast Furnace Digital Twin

Several digital twins have been successfully implemented at Hatch clients' sites by a joint digital and process engineering team following Hatch 3I's framework, for example, blast furnace digital twins for the steel industry, autoclave digital twins for the refractory gold industry, and others. In this section, one use case of blast furnace digital twin is presented as an example to demonstrate digital twin's value through timely decision-making.

Casting is an important part of blast furnace stable operation. It is a critical task for cast-house floor operators to control hearth liquid heights and avoid large liquid accumulation. The presence of excessive liquids can exert back-pressure on the tuyere raceways in the furnace, distort the gas flow, and create an overall negative impact on furnace operations, such as slow burden movement, hanging and slipping, elevated heat load on the bosh and lower stack leading to increased coke consumption and carbon emissions. Many blast furnace operators follow casting rules that are enacted from experience but in most cases, there are no real-time measurements of hearth liquid heights to validate the rules.

As part of the blast furnace digital twin, a casting guidance use case was developed for a better understanding of the accumulation and drainage of iron and slag. It integrated real-time furnace operational data (e.g., hot metal smelting rate, hot blast pressure, slag and blow time, etc.) with the first-principle model developed by Cameron, *et al.* (2020) to establish a true representation of blast furnace hearth and provides an operational intelligence tool with the following three key functions.

Real-time monitoring and prediction of hearth liquid heights to improve process safety and stability. As illustrated in Figure 3, this function provides a 3D visualization of the blast furnace hearth together with real-time casting information, hot metal/slag heights, and their predictions. High liquid height alerts, operational recommendations, and non-adherence to casting Standard Operating Procedures (SOPs) will be sent to operators directly either through an on-screen notification center or mobile text messages.

Figure 3. Example of blast furnace digital twin

Scenario analysis of future cast operations to provide operational guidance towards optimal production. With this function, process engineers and/or cast-house planners can develop a custom response to the current material heights by adjusting some operational levers such as taphole sequence, cast start time, drill bit size, hot metal production rate, etc. Through various scenario analyses based on the current operating conditions, they can experiment with the ideal response to a challenging hearth condition, for example, high slag heights.

Model parameter auto-tuning based on real-time data to minimize plant-model mismatch. This function is performed at the end of each cast to automatically update key process model parameters such as taphole roughness and wear rate to make the digital

twin a true representation of its physical twin through consistent self-learning. Our study showed that the auto-tuning function contributed to a 17% improvement in model prediction accuracy.

The casting guidance use case has been rolled out to the client's Blast Furnace Command Center since late 2020, which provided a data-centric decision tool to furnace/shift coordinators and also enabled Hatch experts to provide remote technical support. Some tangible benefits have been observed, including more consistent hot metal and slag heights inside the blast furnace, a 45% decrease of non-casting time along with a notable reduction in furnace hearth temperature and energy consumption.

4. Development Approach and Lessons Learned

It has been a very challenging but rewarding journey over the past three years to establish digital twin as one of the service offerings at Hatch Digital. We have made some great successes to deploy digital twins across a plant site but also failed by mistakes. This session discusses some lessons learned as a way of sharing our experience from a practitioner's perspective.

Focusing on value delivery is the only way to meaningfully apply a digital twin. At the beginning of digital twin development, it is critically important to identify digital twin values specific to users and differentiate them from other technologies such as Business Intelligence (BI) dashboard or simulation applications. The values need to be tangible and measurable to justify potential investments. We have experienced a predictive maintenance use case, in which the initial costs of developing a failure prediction model are difficult to justify due to the low probability of the failure occurrence.

Digital twins are always evolving so as the development approaches – be agile. It should be completely fine to start with a Minimum Viable Product (MVP) just to address one or two key features, and then add new ones gradually based on users' feedback. Taking the casting guidance as an example, along the six-month development process, we added new features almost every month, from 3D unity model for enhanced visualization, real-time notification via mobile text messages, to model tuning and auto-tuning functions. Today we are developing a new AI technology to characterize the changes of taphole based on image analysis to further improve the model accuracy.

Accelerate your time-to-value by adopting a low-code digital twin platform. Developing a digital twin requires the connection of real-time data, orchestration of process models, and customization of user interfaces to support data visualization and user interactions. We found that some low-code digital twin platforms (such as XMPro, etc.) have developed great built-in data connection and visualization features, allowing us to concentrate more on intelligence and insight development and help our clients realize benefits sooner and faster, which is key to the success of digital twin projects.

Make your digital twin a trustworthy tool to operators. The most common failure in digital twin implementation is that users stopped using it for decision support after the initial commissioning stage because the digital twin can no longer represent its physical processes. This can be caused by several different reasons related to, for example, data quality, model/algorithm robustness, and/or digital system stability. To prevent this situation from happening, certain quality assurance measures need to be incorporated into digital twin design. Some of them may include real-time validation of input operational data, continuous monitoring of model health, built-in model self-learning capabilities to

adapt to possible process changes over t.me, and/or a consistent and meaningful way to interpret and present digital twin results to avoid any ambiguity and misunderstanding from users' point of view.

Last but not least, y*our commercial model is either a roadblock or a lubricant to the success of digital twin projects*. Although this is not a technical issue, we often found that in many cases, a promising digital twin project may quickly lose the support either from the developer- or client-side as the commercial model is not sustainable. In other words, the costs of creating and running a digital twin are not met by the values created. We have to respect that there is no one commercial model that can fit all. A flexible commercial model with a focus on long-term benefits will help solve the issue.

Keep these lessons learned in mind, we have established a proven, agile development approach to transforming a conceptual idea into a real-world digital twin solution, which consists of the following five steps:

(1) focusing our attention on understanding a wide range of user needs, and grouping them into multiple categories such as functional requirements, data requirements, user experience requirements, or security requirements, etc.

(2) conducting persona workshops to identify goals, key users, and pain points for each use case, and prioritizing them based on their potential benefit and implementation cost.

(3) proposing a solution with high-priority use cases, and using an interactive mockup as a communication tool to visualize different ideas and solicit stakeholders' feedback.

(4) commissioning an MVP through plant trials, targeting a quick delivery with added values, and continuously improving it during the following development.

(5) comparing with the original requirements to ensure all needs are fully satisfied by MVP or its further improved versions. It is worth mentioning that this step often serves as a new round of innovation process that may lead to further design improvements.

5. Conclusion

Our view on digital twins, development practice as well as lessons learned from the past project experience are shared from an industrial practitioner's perspective in this paper. We believe, with strong support from Industry 4.0 technologies, the digital twin will continue to be a powerful technology tool to solve different plant operation, maintenance, and value chain optimization challenges, and provide a good starting point of the digital journey of driving toward next-generation decision-making for the metals industry.

References

A. Mussomeli, A. Paarot, B. Umbenhauer and L. Warshaw, "Digital twins: Bridging the physical and digital," Deloitte, 15 January 2020.

DNV-GL, "DNVGL-RP-A204: Qualification and Assurance of Digital Twins," DNV, 2020.

Y. Zhang, M. Sukhram, I. Cameron, and A. Rozo, "Industrial Perspective of Digital Twin Development and Applications for Iron and Steel Processes," *Iron and Steel Technology,* no. 3, pp. 34-43, 2021

I. Cameron, M. Sukhram, K. Lefebvre, and W. Davenport, *Blast Furnace Ironmaking Analysis, Control, and Optimization.* Mississauga: Elsevier. 2020

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50180-9

The study on feasibility of HFO refrigerants in BOG re-liquefaction process

Taejong YU[a], Donghoi KIM[b], Truls GUNDERSEN[c*], Youngsub LIM[a,d*]

[a]*Department of Naval Architecture and Ocean Engineering, Seoul National University, 1 Gwanak-ro,*
Gwanak-gu, Seoul, 08826, Republic of Korea
[b]*SINTEF Energy Research, Sem Sælands vei 11, NO-7465 Trondheim, Norway.*
[c]*Department of Energy and Process Engineering, Norwegian University of Science and Technology*
(NTNU), Kolbjørn Hejes Vei 1B, NO-7491, Trondheim, Norway
[d]*Research Institute of Marine Systems Engineering, Seoul National University, 1 Gwanak-ro,*
Gwanak-gu, Seoul, 08826, Republic of Korea

Abstract

As regulation of ship emission established, using liquefied natural gas (LNG) as ship fuel is one of the method to reduce ship emission. For LNG carrier, boil-off gas (BOG) generated from the storage tank can be used as fuel and it is economical to adopt BOG re-liquefaction system on the vessel. In this study, we investigate the feasibility of hydro-fluoroolefins (HFO) refrigerants in the BOG re-liquefaction system. We simulate BOG re-liquefaction system and optimize it to minimize energy consumption. Then, the global warming potential(GWP) generation during lifecycle of the ship is estimated. As the results, HFO refrigerants has similar performance in BOG re-liquefaction system in terms of energy consumption compared to hydrocarbon refrigerants. In case using HFO refrigerants, the specific power consumption (SPC) of re-liquefaction process is 0.479 kWh/kg and in case using hydrocarbon refrigerants it is 0.471 kWh/kg. However, although HFO refrigerants has lower GWP than hydrocarbons, the total GWP generation is larger than when using hydrocarbons due to GWP generation caused by CO2 emission.

Keywords: BOG re-liquefaction process, hydro-fluoroolefin refrigerants, optimization, the global warming potential.

1. Introduction

For long distance, it is economical to transport liquid form of natural gas (NG) by LNG carriers. However, LNG is stored in the storage cargo tank at -160 °C under ambient pressure (Lim et al, 2013), so part of LNG is vaporized to boil-off gas (BOG) by heat ingress into the tank. For LNG carrier, it is economical to use BOG as fuel and re-liquefy the remaining BOG to recover it to the tank. There are two types of gas propulsion engines mounted on ships: a high-pressure gas propulsion engine (ME-GI) developed by MAN B&W and a low-pressure gas propulsion engine (X-DF) developed by WIN-GD. In the initial gas propulsion engine market, the market share of high-pressure gas propulsion engines was high, but as emission regulations for nitrogen oxides (NOx) and sulfur oxides (SOx) were tightened (IMO, 1999, 2016, 2017), interest in low-pressure gas propulsion

engines is increasing (IGU, 2019). For low pressure gas propulsion engine, it is advantageous in terms of energy consumption to liquefy and recover BOG using an external refrigerant cycle due to its low operating pressure.

Meanwhile, hydro-fluoroolefins (HFO) are consider as suitable alternative to conventional refrigerants established through the Montreal Protocol and other such treaties in heating, ventilation and air-conditioning (HAVC) and automobile industries. The conventional refrigerants in HAVC and automobile industries include hydrofluorocarbons (HFCs), chlorofluorocarbons (CFCs) and hydrochlorofluorocarbons (HCFCs) which have high global warming potential (GWP) or ozone depletion potential(ODP). The HFOs have double bonds in its molecular structure, so the rate of decomposition in the air is high. Thus they have little environmental impact, resulting in low GWP and ODP.

Currently, there is little research on the re-liquefaction of BOG using HFO refrigerants, so it is worth studying their feasibility in the BOG re-liquefaction system. In this study, BOG re-liquefaction process using HFO refrigerants is modeled and optimized to minimize the energy consumption. Then, GWP generation during lifecycle is estimated and compared to case of using conventional hydrocarbon refrigerants.

2. Process design

2.1 Design basis

The LNG and BOG composition is assumed as shown in Table 1. ASPEN HYSYS V10 is used as a software for process simulation, and this study uses the Peng–Robinson equation of state, which is suitable for simulating LNG, a light hydrocarbon, up to high pressure and cryogenic conditions.

Table 1. LNG (Aspelund et al, 2010) and BOG composition.

Type	Unit	Value
LNG composition		
Nitrogen	mol%	0.37
Methane	mol%	95.89
Ethane	mol%	2.96
Propane	mol%	0.72
Butane	mol%	0.06
BOG composition		
Nitrogen	mol%	0.48
Methane	mol%	99.49
Ethane	mol%	0.03

The amount of BOG generated from the tank can be estimated the Equation (1).

$$BOG_{tank} = V \cdot L_v \cdot \rho \cdot V_\gamma \quad (1)$$

where BOG_{tank} is the rate of BOG generated from the storage tank, V is the volume of the storage tank, L_v is the liquid level, ρ is the average density of the LNG, and V_γ is the

evaporation rate of the storage tank. Assuming that the tank volume is 170,000 m^3, the liquid level is 95%, the average density is 437.9 kg/m^3, and the evaporation rate is 0.1%/day, the amount of BOG generated per hour could be estimated to be about 2938 kg/h.

To determine the BOG consumption rate for the main propulsion engine and auxiliary power generation engine, the following Equation (2) could be used.

$$BOG_{fuel} = \frac{P_e \cdot SFOC \cdot l}{LHV} \quad (2)$$

where BOG_{fuel} is the fuel consumption in kg/hr, P_e is the engine power in kW, SFOC is the specific fuel oil consumption in kJ/kWh, LHV is the lower heating value of BOG in kJ/kg based on the BOG composition, and l is the engine load in %. Table 2 shows the corresponding values of each engine are assumed.

Table 2. Engine specification (Wartsila, 2019) (WinGD, 2018).

Type	Unit	Main Engine	Aux. Engine
Manufacturer		WinGD	Wartsila
Model		X62DF	16V34DF
Power	kW	11,925 x 2	8,000
SFOC	kJ/kWh	7,132	7,679
Load	%	40	50
LHV of BOG	kJ/kg	42,900	
Fuel consumption	kg/h	1,586	716

By subtracting the fuel consumption obtained from through the above calculation from the BOG generation, the amount of BOG supplied to the BOG re-liquefaction system can be estimated as 639 kg/h.

In this study, the basic process configuration set as the single mixed refrigerant (SMR) cycle. Figure 1 shows the BOG re-liquefaction process.

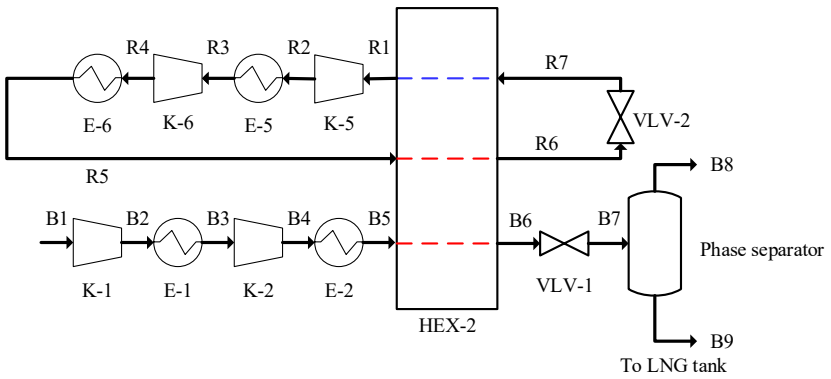


Figure 1. Process flow diagram of BOG re-liquefaction system.

2.2 Optimization

One of main purpose of this study is to investigate the cooling efficiency of HFO refrigerants. To do this, the specific power consumption (SPC), one of the key indicator of energy consumption in liquefaction process, is optimized. The SPC can be estimated as following Equation (3).

$$SPC = \frac{\sum \dot{W}}{\dot{m}_{LNG}} \tag{3}$$

where $\sum \dot{W}$ is the total power consumption of the liquefaction system, and $\dot{m}_{LNG}$ is the the total mass flow of LNG generated at the rear end of the phase separator. After selecting SPC as the objective function for comparative analysis, the optimum point to minimize SPC is found in each process. The process variables are the pressure ratio of the compressors (K-1 to K-4), expansion pressure at the rear end of the valve (R7), cooling temperature (R6), and the composition and mass flow rate of the mixed refrigerant. The constraint set the minimum temperature difference of the heat exchanger to 3 °C to prevent oversizing of heat exchanger and the vapour fraction of compression inlet stream to 1 to prevent malfunction of compressors.

2.3 Environmental assessment

In this study, total GWP generation during the lifecycle of an LNG carrier is estimated by two parts: (1) GWP caused by the disposal of refrigerants ($GWP_{disposal}$) after life cycle of the ship and (2) GWP caused by CO2 emission for power generation (GWP_{CO2}). The lifecycle of an LNG carrier is assumed to be 20 years. The $GWP_{disposal}$ is the GWP generation from a disposal of refrigerants after lifecycle of the ship. Equation (4) shows GWP value due to the disposal of the refrigerants.

$$GWP_{\text{disposal}} = GWP \cdot V_{re-liq} \cdot \rho_{\text{mr}} \cdot f_{\text{disposal}} \tag{4}$$

where $GWP_{disposal}$ is the GWP generation due to the disposal of refrigerants, GWP is the value of the global warming potential of the mixed refrigerants, V_{re-liq} is the inventory of the re-liquefaction system, and f_{disposal} is the percent of refrigerant disposed and assumed 30% for medium and large commercial refrigeration (U.S. EPA, 2014). To estimate the inventory of the refrigeration in the system (V_{re-liq}), Equation (5) can be used.

$$V_{re-liq} = V_{HEX} \cdot \rho_{\text{mr}} \cdot f \tag{5}$$

where V_{HEX} is the volume of the heat exchanger and estimated by using ASPEM EDR V10 as the heat exchanger simulation program. ρ_{mr} is the density of the mixed refrigerant. f is the factor to estimate total amount of refrigerant from inventory of the heat exchanger and is assumed to be 2.38.

The GWP_{CO2} is the GWP generation due to CO2 emission for power supply to the re-liquefaction system in per year, and it can be estimated by Equation (6) (Trozzi, 2010).

$$GWP_{\text{CO2}} = P_e \cdot \text{EF} \tag{6}$$

where P_e is the power consumed in the re-liquefaction process and EF is the carbon emission factor of the fuel. It refers to the amount of CO2 generated to produce a unit of energy depending on the type of fuel and is assumed as 61.0 for LNG fuel (Australia.

Department of Industry, Science, Energy and Resources, 2017). Availability of the re-liquefaction process is assumed to be 50% per year.

3. Results and discussion

In this section, cases are classified according to the combination of refrigerants. Case 1 uses the mixture of conventional hydrocarbons. It consists nitrogen (N2), methane (C1), ethane (C2), propane (C3), and normal butane (nC4). In case 2, C3 and nC4 are replaced with HFO refrigerants: HFO 1234yf, HFO 1234ze, HFO 1233zd. Table 3 shows the results of the SPC optimization.

Table 3. Results of SPC optimization.

Parameter		Unit	Case3		Case3	
Composition		mol%	N2	11.1	N2	15.9
		mol%	C1	28.2	C1	26.3
		mol%	C2	29.5	C2	29.8
		mol%	C3	0.0	HFO 1234yf	12.2
		mol%	nC4	31.2	HFO 1234ze	2.4
					HFO 1234zd	13.4
MR flow rate		kg/h	4074		4886	
BOG feed pressure		bar	90.33		95.02	
MR compression pressure	inlet	bar	3.7		4.3	
	outlet	bar	22.8		41.8	
BOG temperature after HEX			-156.8		-156.9	
Total power consumption		kW	268.0		272.6	
SPC		kWh/kg	0.471		0.479	

In Case 1, the total flow of the MR is 4074 kg/h. The BOG is compressed to 90.33 bar and then flow into the heat exchanger and cooled to -156.8 °C and expanded in the valve. The total power consumption is 268.0 kWh, and SPC is 0.471 kWh/kg. In Case 2, the total flow rate of the MR is 4886 kg/h, and after BOG is compressed 95.02 bar, it is cooled down to -156.9 °C. The MR requires compression 41.8 bar, the total power consumption is 272.6 kWh, and SPC is 0.479 kWh/kg. As shown in the results, the total power consumptions of each case similar. There is only 1.7% difference when replacing C3 and nC4 with HFO refrigerants. After the optimization of energy consumption, GWP generation is estimated. Table 4 shows the results of GWP estimation.

Table 4. GWP generation during the lifecycle.

Case	GWP_{disposal}	GWP_{CO2}	Total GWP (during life cycle)
Case1	2,973	217,207	4,347,121
Case2	1,762	220,936	4,420,473

Case 2 has the lower $GWP_{disposal}$ compared to Case 1 due to lows GWP values of HFO refrigerants. If an HFO refrigerants are used, GWP generated by disposal of refrigerants after the lifecycle can be reduced by 40% compared to the hydrocarbon mixture. Although $GWP_{disposal}$ of Case 2 is lower than that of Case 1, the total GWP generation during lifecycle is much larger due to GWP_{CO2} proportional to energy consumption. This means that the use of low GWP materials for refrigerants may not contribute to reduce the GWP generation, keeping low SPC is more effective to decrease the annual GWP.

4. Conclusion

In this study, we investigate the feasibility of HFO refrigerants in the BOG re-liquefaction system. We modelled BOG re-liquefaction system and optimized it to minimize the specific power consumption of the process. Then, we estimated GWP generation during the lifecycle of the ship. As the results, when HFO refrigerants replaces hydrocarbon refrigerants, there is no large difference in energy consumption. The conventional hydrocarbon mixture has the SPC of 0.471 kg/kWh and the mixture of HFO refrigerants and hydrocarbon has 0.479 kWh/kg. However, this small difference effects on GWP generation. Although HFO refrigerants generate lower GWP when disposed, it has larger GWP value totally due to GWP generation caused by CO2 emission. When using HFO refrigerants, $GWP_{disposal}$ is 1762 and when using hydrocarbons, it is 2973. GWP_{CO2} is 217,207 and 220,936 in each case. Consequently, total GWP generation is 4,437,121 in case of HFO refrigerants and 4,420,473 in case of hydrocarbon refrigerants. This means that energy consumption has a greater influence on GWP production than GWP of refrigerant.

References

Aspelund. A, Gundersen. T, Myklebust. J, Nowak. M, Tomasgard. A, An optimization-simulation model for a simple LNG process, Comput. Chem. Eng, 34, 1606–1617.

Austraila. Department of Industry, Science, Energy and Resources, 2017, National greenhouse accounts factors.

EPA, 2014, Direct Fugitive Emissions from Refrigeration, Air conditioning, Fire Supression, and Industrial Gases, Environmental Protection Agency, US.

IGU, 2019, World LNG Report 2019, Internatinal Gas Union.

IMO, 1999, International Convention for the Prevention of Pollution from Ships MARPOL 73/78: The Regulations for the Prevention of Air

Pollution from Ships (AnnexVI), Internatinal maritme organization, London, UK, 1999.

IMO, 2016, Marine Environment Protection Committee (MEPC), 70th Session, Internatinal maritme organization, London, UK.

IMO, 2017, Prevention of Air Pollution from Ships, Internatinal maritme organization, London, UK.

Lim. W, Choi. K, Moon.I, 2013, Current Status and Perspectives of Liquefied Natural Gas (LNG) Plant Design, Ind. Eng. Chem. Res, 52, 3065-3388.

Trozzi. C, 2010, Emission Estimate Methodology for Maritime Navigation, T. Consulting, Editor.

Wartsila, 2019, Wärtsilä 34DF Brochure; Wartsila, Helsinki, Finland.

WinGD, 2018, Low-Pressure X-DF Engines FAQ, WinGD, Winterthur, Switzerland.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50181-0

Crude Oil Blending Process Optimization with Precise Consideration of Fraction Properties

ZHENG Wanpeng[a], GAO Xiaoyong[a*], KUI Guofeng[a], ZUO Xin[a], ZHU Guiyao[a],XIE Yi[b]

[a] *Department of Automation, College of information Science and Engineering, China University of Petroleum, Beijing 102249, CHINA*
[b]*Beijing pipeline company of national pipeline network group, Beijing 100101,CHINA*
x.gao@cup.edu.cn

Abstract

Crude oil blending process is an integral part of the petroleum supply chain, including multiple industrial processes such as crude oil distribution, transportation, storage and blending in the production process of refinery enterprises. The optimization of crude oil blending process scheduling has high academic and industrial application value, and its related research work is currently a hot topic of academic interest. However, there are still urgent problems to be solved in the current research work. Crude oil blending process not only needs to consider a variety of delivery and distribution of crude oil, but also considers the constraint conditions that the blended product meets the production demand. Therefore, based on the continuous-time representation, a crude oil blending optimization model that precisely considers the properties of the fraction is proposed in this paper. Firstly, the important achievements in the research field of crude oil blending process optimization are briefly introduced, and the development trend and defects of the current research work are summarized. Subsequently, the MINLP model is described in detail. The model especially considers the properties demand and supply demand of mixed products in the secondary process. Finally, we verified the effectiveness of the proposed model in solving the actual blending formula optimization problem. The simulation results of a real case of a fuel refinery show that a product formulation is used to optimize the crude oil blending process, which can effectively improve the overall yield of petroleum fractions while meeting the demands of the secondary processing device.

Keywords: Optimization; Model; Simulation.

1. Introduction

Due to the variety of crude oils purchased by most refineries, properties of the material exist significantly different, while the refinery requires homogeneous and stable materials to ensure the quality and yield of the subsequent secondary processing process. Therefore, the refinery mixes crude oil according to different blending formulations in proportion during the crude oil blending process, to make the material properties meet the demand of the secondary processing device. In the 1980s, to adapt to the competitive global market, oil refining enterprises began to focus on the adjustment and optimization of industrial structure and production process. Crude oil blending can improve the quality grade of oil products, and then achieve higher economic benefits for refining enterprises. Therefore, the research on crude oil blending optimization has gradually become the focus of related fields.

In recent years, many achievements have been made in the research of crude oil blending optimization. The in-house center discretization method is proposed to deal with the long-horizon tank blending scheduling problem (Beach et al., 2020). The Lagrangean decomposition algorithm is used to solve the integration of crude oil scheduling and refinery planning (Yang et al., 2020). The operation optimization of crude oil blending and intermediate oil processing is solved to achieve the collaborative optimization of materials processing and product deployment (Li et al., 2020). With desalination as a separate task, the crude oil refinery operation of feeding a single desalination tank to multiple crude oil distillation units is studied (Bayu et al., 2020). The crude oil selection scheduling optimization is solved by two-stage stochastic programming (Li et al., 2021). Zhao et al. (2017) proposed a modelling method based on priority slots for crude oil scheduling with inconsistent component concentrations. Castillo et al. (2017) proposed a global optimization algorithm to solve a continuous-time MINLP blending scheduling model. The algorithm uses piecewise McCormick relaxation and normalized multiparametric disaggregation to calculate the global optimal estimate. Menezes et al. (2017) proposed a quantitative analysis method to optimize crude oil blending in oil refining enterprises to narrow the decision-making gap between crude oil procurement and production scheduling operations. Menezes et al. (2019) integrate scheduling operation details and time steps into different types of PSE solutions to get solutions to industrial problems faster. Franzoi et al. (2018) realized the effective optimization of complex process systems by means of parameter feedback after data coordination calculation. Franzoi et al. (2019) studied the

factors affecting scheduling production performance and proposed a detailed design of a hybrid scheduling and processing unit considering flexibility, responsiveness, and management capabilities.

The optimization of the crude oil blending process generally has two main points that need to be paid attention to: one is a typical scheduling problem, such as stock limitation and device processing capacity, and the other is the properties of the blended products. During the operation of the refinery, the blended product is separated by the atmospheric and vacuum distillation unit, and then become the material for the secondary processing device. Material properties that do not meet the demand of the processing devices will bring serious problems, affect the production of the factory. For example, for catalytic reforming device, when producing BTX, the required material is petroleum fraction between 60-145°C intervals. When producing high-octane gasoline, the required material is petroleum fraction between 80-200°C intervals. In addition, the catalytic reforming device has different limits on the sulfur content, nitrogen content, heavy metal content and molecular structure of the materials when producing different pro-ducts. Therefore, according to the demands of secondary processing devices, it is valuable to consider the properties of petroleum fractions precisely, but the current research work is too simple to consider the properties of petroleum fractions.

In this paper, we precisely consider the demand for secondary processing devices on the properties of materials. Inspired by Mendez's work in the field of gasoline blending optimization, based on continuous-time representation method, a crude oil blending optimization model is proposed considering the yield of blending products, sulfur content, nitrogen content, molecular structure and other properties, and minimizes the gap between the blended product and the desired product as the objective function (Mendez et al., 2006). According to an actual case, we obtained the results through program simulation, and verified that the model can be applied to different types of refining enterprises, provide reliable blending formula, and ensure that the properties of blending products can meet the demands of the secondary processing device.

2. Model description

In the production process of refining enterprises, different secondary processing devices always have different demands on the properties of petroleum fractions obtained by constant vacuum distillation of blended products. For example, catalytic cracking mainly uses heavy distillates and residuals as processing materials, while hydrocracking mainly uses VGO, CGO and residuals as processing materials. There are great differences in the properties of these petroleum fractions. In the atmospheric and vacuum distillation process, the main influencing factor that determines the distribution of petroleum distillates is to adjust the yield of crude oil in each fraction interval. In addition, high sulfur content will cause catalyst pollution and inactivation, affect the quality of petroleum products and cause device corrosion, high nitrogen content can cause catalyst poisoning and affect the stability of petroleum products, heavy metal content (Ni+V) will cause catalyst deactivation and bed blockage, the difference in molecular structure will affect the production process and the selection of catalysts. Therefore, according to the processing demands of the secondary processing device, the model not only needs to consider the yield of the blended crude oil in each fraction, but also needs to consider whether the sulfur content, nitrogen content，heavy metal content and molecular structure in the blended crude oil meet the demand of the device.

Before presenting the proposed mathematical models, the nomenclature in the model formula is explained as follows:

Indices

e – secondary processing device
i – crude oil component
p – CDU device processed product
t – time slot

Sets

I – set of the crude oil component
P – set of CDU device processed product
T_d – set of time slots postulated for the sub-interval ending at due date d

Parameters

d – due dates of product demand
h – time horizon
w – fraction temperature
$cor_{p,t}$ – correction factor of product p in time slot t
$cp_{i,p}^{\max}$, $cp_{i,p}^{\min}$ – maximum/minimum concentration of component i in product p
$D_{p,d}$ – demand of product p
$D_{p_l,d}$, $D_{p_c,d}$, $D_{p_h,d}$ – demand of light/middle/heavy fraction of product p at the end
$d_{p,d}$ – demand of product p to be satisfied at due date d
f_i – constant delivery rate of component i
$fr_p^{\max}$, $fr_p^{\min}$ – maximum/minimum flowrate of product p in time slot t
$gra_{p,t}$ – specific gravity of product p in time slot t
in_p, in_i – initial inventory of product p/component i
$met_p^{\max}$, $N_p^{\max}$, $S_p^{\max}$ –maximum heavy metal/nitrogen/sulfur content specified by device e
met_i, N_i, S_i – heavy metal/nitrogen/sulfur content of component i
$mst_{p,e}^{\max}$, $mst_{p,e}^{\min}$ – maximum/minimum molecular structure (hydrocarbon or gum composition) ratio specified by device e

mst_i – molecular structure ratio of component i

n_t – maximum number of blenders that can be working in parallel in time slot *t*

$V_p^{\max}$, $V_p^{\min}$ – maximum/minimum storage capacity of product *p*

$V_i^{\max}$, $V_i^{\min}$ – maximum/minimum storage capacity of component *i*

$vin_{p,w}$ – ideal yield of product *p* at temperature *w*

TEP_{p_l} , TEP_{p_c} , TEP_{p_h} – the initial boiling point of the light/ middle/heavy fraction of product *p*

$TIBP_p$ – the final boiling point of product *p*

$TIBP_{p_l}$, $TIBP_{p_c}$, $TIBP_{p_h}$ – the final boiling point of the light/middle/heavy fraction of product *p*

$tbp_{p,w}^{\max}$, $tbp_{p,w}^{\min}$ – maximum/minimum yield of product *p* at temperature *w*

$tbp_{i,w}$ – yield of component *i* at temperature *w*

$tbploss_i$ – yield loss of component *i*

ρ_i – density of component i

Variables

Ap,t – binary variable denoting that product *p* is blended in time slot *t*

e_t – ending time of time slot *t*

$F_{i,p,t}^I$ – the amount of component *i* being transferred to product *p* in time slot *t*

$F_{p,t}^P$ – the amount of product *p* being blended in time slot *t*

s_t – starting time of time slot *t*

$V_{p,t}^P$ – the amount of product *p* stored at the end of time slot *t*

$V_{i,t}^I$, $V_{i,t}^{I'}$ – the amount of component *i* stored at the end/ beginning of time slot *t*

$TBP_{p,w,t}$ – yield value of product *p* at time *t*, temperature *w*

$TBP_{p,i}$ – yield value of product *p*

TEP_p – the initial boiling point of product *p*

The MINLP model we proposed partitions the entire time horizon into a predefined number of sub-intervals, the length of each sub-interval will depend on the product due dates. In addition, the proposed model has the following features: 1)Divide the scheduling horizon into multiple sub-intervals, and set a set of time slots with unknown time and location for each sub-interval; 2) Assuming that the flow of components and product is constant throughout the scheduling horizon; 3) Final product properties are based on a volumetric average and a correction factor computed through the proposed iterative process; 4) A particular product demand can be satisfied by one or more periods. The constraints and variables of the model are described as follows:

2.1. Composition and material balance constraint

To ensure that the product *p* is mixed with the crude oil component *i* of a limited type, the component equation constraint is as shown in constraint (1); to meet product quality demands and crude oil supply constraints, upper and lower limits can be imposed on the component concentration of product *p*, as shown in Constraint (2); to ensure the material balance of product *p*, the constraint of material balance equation can be applied, as shown in constraints (3); to ensure the material balance of component *i*, the constraint of material balance equation can be applied, as shown in constraints (4) and (5).

$$\sum_i F_{i,p,t}^I = F_{p,t}^P, \forall p,t \tag{1}$$

$$cp_{i,p}^{\min} F_{p,t}^P \le F_{i,p,t}^I \le cp_{i,p}^{\max} F_{p,t}^P, \forall i,p,t \tag{2}$$

$$V_{p,t}^P = in_p + F_{p,t}^P - d_{p,d}, \forall p,t \tag{3}$$

$$V_{i,t}^I = in_i + f_i e_t - F_{i,p,t}^I, \forall i,t \tag{4}$$

$$V_{i,t}^{I'} = in_i + f_i s_t - F_{i,p,t}^I, \forall i,t \tag{5}$$

2.2. Device and inventory constraint

To meet the operating limits of the device, the upper and lower limit is imposed on the volume flow of product *p*, as shown in constraint (6); to meet the inventory limit, the upper and lower limit constraints are imposed on the quantity of product *p* and component *i*, as shown in constraint (7), (8) and (9); to ensure the balance of the productivity of the CDU device, equation constraints are proposed as shown in (10) and (11).

$$\begin{aligned} &fr_p^{\min}(e_t - s_t) - fr_p^{\min} h(1 - Ap,t) \\ &\le F_{p,t}^P \le fr_p^{\max}(e_t - s_t), \forall p,t \end{aligned} \tag{6}$$

$$V_p^{\min} < V_{p,t}^P < V_p^{\max}, \forall p,t \tag{7}$$

$$V_i^{\min} < V_{i,t}^I < V_i^{\max}, \forall i,t \tag{8}$$

$$V_i^{\min} < V_{i,t}^{I'} < V_i^{\max}, \forall i,t \tag{9}$$

$$\sum_p TBP_{p,i} + tbploss_i = 1 \tag{10}$$

$$TBP_{p,i} = \sum_{w=TIBP_p}^{TEP_p} tbp_{p,w} \tag{11}$$

2.3. Supply constraint

To meet the total supply-demand of subsequent secondary processing process, a lower limit constraint can be imposed on the quantity of product *p*, as shown in Constraint (12); in addition, the supply constraints of light fraction, middle fraction and heavy fraction processed by CDU device are shown in constraints (13), (14) and (15); the demands for different fractions must be less than or equal to the total demand, as shown in constraint (16).

$$ini_p + \sum_{t<d} F_{p,t}^P \ge D_{p,d} \tag{12}$$

$$\sum_{t<d} \sum_{w=TIBP_{p_l}}^{TEP_{p_l}} tbp_{p,w} F_{p,t}^P \ge D_{p_l,d} \tag{13}$$

$$\sum_{t<d} \sum_{w=TIBP_{p_c}}^{TEP_{p_c}} tbp_{p,w} F_{p,t}^P \ge D_{p_c,d} \tag{14}$$

$$\sum_{t<d}\sum_{w=TIBP_{p_h}}^{TEPp_h} tbp_{p,w}F_{p,t}^{P} \geq D_{p_h,d} \quad (15)$$

$$D_{p,d} \geq D_{p_h,d} + D_{p_c,d} + D_{p_l,d} \quad (16)$$

2.4. Yield and property constraint

To meet the processing demands, the upper and lower limits can be imposed on the product yield, as shown in constraint (17); to meet the demand of secondary processing device for the properties of materials, the upper limits on sulfur content, nitrogen content, and heavy metal content (Ni+V) can be imposed, as shown in constraint (18), (19) and (20).

$$\begin{aligned} & tbp_{p,w}^{\min}F_{p,t}^{P} \leq \sum_i tbp_{i,w}F_{i,p,t}^{I} \\ & \leq tbp_{p,w}^{\max}F_{p,t}^{P}, \quad \forall i,p,t \end{aligned} \quad (17)$$

$$\sum_i S_i F_{i,p,t}^{I} \leq S_{p,e}^{\max}F_{p,t}^{P}, \forall i,p,t \quad (18)$$

$$\sum_i N_i F_{i,p,t}^{I} \leq N_{p,e}^{\max}F_{p,t}^{P}, \forall i,p,t \quad (19)$$

$$\sum_i met_i F_{i,p,t}^{I} \leq met_{p,e}^{\max}F_{p,t}^{P}, \forall i,p,t \quad (20)$$

In particular, the distribution of sulfur, nitrogen and heavy metal elements is similar: with the increase of boiling point, their content will increase, mainly concentrated in residual oil. Therefore, sulfur content, nitrogen content and heavy metal content in different fractions can be estimated roughly according to the material balance equation and analytical test results.

2.5. Set of molecular structure constraints

To meet the production needs and device limitations of the secondary processing process, the upper and lower limits can be imposed on the molecular structure of the product, as shown in constraints (21) and (22).

$$\begin{aligned} & mst_{p,e}^{\min}F_{p,t}^{P} \leq \sum_i mst_i F_{i,p,t}^{I} + cor_{p,t}F_{i,p,t}^{I} \\ & \leq mst_{p,e}^{\max}F_{p,t}^{P}, \quad \forall i,p,t \end{aligned} \quad (21)$$

$$\sum_i mst_i F_{i,p,t}^{I} + cor_{p,t}F_{i,p,t}^{I} = \frac{\sum_i mst_i \rho_i F_{i,p,t}^{I}}{gra_{p,t}}, \forall i,p,t \quad (22)$$

In particular, for light fraction, middle fraction and heavy fraction, the molecular structure constraints are different. For light fraction, the main molecular structure constraints include the ratio of alkanes, cycloalkanes and aromatic hydrocarbons; for middle fraction, the constraints include the ratio of alkanes, cycloalkanes, aromatic hydrocarbons and alkenes; for heavy fraction, the constraints include the ratio of saturated hydrocarbons, aromatic hydrocarbons and gums. Since the crude oil distillation process is purely physical, the molecular structure of the product depends on the concentration of the crude oil in the blending process. For the secondary processing process, molecular structure constraints are necessary, which can effectively guarantee the stable operation of processing devices and improve efficiency and profitability.

The molecular structure constraint was first proposed in the relevant research on crude oil blending optimization. Although the current constraint still needs to be compensated and corrected, the modified result can still ensure that the product properties of fractions can meet the needs of secondary processing devices, which is of great significance.

2.6. Set of time slot timing constraints

To allow more flexible solutions and avoid overlapping time slots, correct order and sequence between postulated time slots must be established through the next set of constraints. Time slot duration constraint is shown in constraint (23); time slot sequencing constraint is shown in constraint (24); sub-interval bounds constraint is shown in constraint (25) and (26); time slot assignment constraint is shown in constraint (27).

$$e_t - s_t \leq h\sum_p A_{p,t}, \forall t \quad (23)$$

$$e_t \leq s_{(t+1)}, \forall t \quad (24)$$

$$s_t \geq d-1, \forall t \quad (25)$$

$$e_t \leq d, \forall t \quad (26)$$

$$\sum_p A_{p,(t+1)} \leq n_t \sum_p A_{p,t}, \forall (t,t+1) \in T_d \quad (27)$$

2.7. Objective function

While satisfying all the above constraints, the objective function of the optimization model is to minimize the gap between the blended crude oil yield of each fraction and the expected ideal crude oil yield, as shown in equation (28).

$$\min \sum_w \left| \frac{(tbp_{p,w} - vin_{p,w})}{vin_{p,w}} \right| \quad (28)$$

In particular, the expected ideal yield $vin_{p,w}$ is the product optimal value obtained under a certain operating mode assuming that the refinery has sufficient reserves for each type of crude oil, and the inventory capacity and processing capacity of the device can meet the demands. But in reality, due to inventory, devices and other factors, the yield of ideal products is impossible to achieve.

3. Simulation case

In order to verify the effectiveness and stability of the model in solving practical problems, we use a set of real fuel-based refinery devices data for simulation, the initial product is *p*. The production materials of the fuel-based refinery device come from the petroleum distillate separated from the CDU device: the material of catalytic reforming unit is naphtha and gasoline distillate in the range of 80-200℃; the material of catalytic cracking unit is heavy distillate and residual distillate in the range of 450-530 ℃; the material of hydrocracking unit is CGO distillate and VGO distillate in the range of 200-450℃; the delayed coking unit the vacuum residuum fraction above 530℃. The simulation device set includes ten storage tanks, three blending tanks, one crude oil distillation unit, one catalytic reforming unit, one catalytic cracking unit, one hydrocracking unit, one delayed coking unit. The simulation components and product data set include ten crude oil components and two products. An illustrative view of the set of simulation devices is shown in Figure 1.

The yield of each component and product is divided into 23 interval ranges according to actual production needs. In the simulation process, we consider two real production situations, the first situation is that the process plan is mainly used to produce gasoline, and the by-product is liquefied petroleum gas, petroleum coke, etc., the petroleum fraction product is $p1$ at this time. The second situation is the process plan mainly produces diesel, with by-products such as gasoline, petroleum coke, etc., and the petroleum distillate product currently is $p2$.

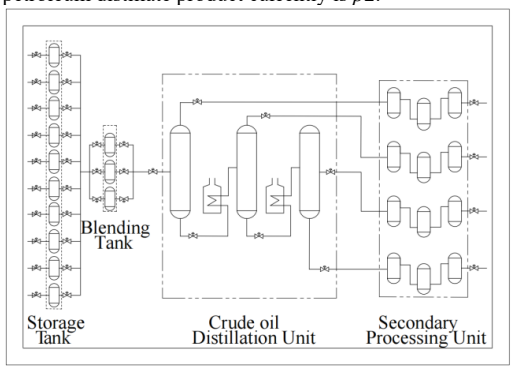


Fig.1. The illustrative view of the fuel-based refinery device

According to the optimization model proposed above, the two production situations can be expressed as minimizing the yield gap between the actual product and the ideal product in the objective function on the premise of satisfying the constraint (1-25). In order to better illustrate the results, we selected three optimization models proposed by research work related to the optimization of crude oil blending process between 2018 and 2021, and simulated and solved the above case. It should be noted that for the selected model, usually, only a few types of crude oil participate in the blending process. Therefore, increasing the types of crude oil participating in the blending process will have a huge negative impact on the solution results. In order to ensure the validity of the solution results, we performed relaxation and disaggregation on some constraints of the selection model. The solution results are compared with the model proposed in this paper. Table.1. details the product yield in a specific distillation interval according to the solution results.

Table.1. Product yield in specific distillation interval

Product name	*p*	*p**	*p(m)*	*p(m1)*	*p(m2)*	*p(m3)*
The first situation：mainly produce gasoline						
80-100℃ interval	2.97	4.95	4.41	4.12	3.62	4.00
100-130℃ interval	4.65	6.25	5.85	5.56	5.18	5.47
130-160℃ interval	4.73	6.00	5.76	5.42	5.16	5.51
160-180℃ interval	3.55	3.84	3.78	3.72	3.64	3.66
180-200℃ interval	2.97	3.09	3.09	3.04	3.01	3.04
80-200℃ interval	18.87	24.13	22.89	21.86	20.60	21.68
The second situation: mainly produce diesel						
350-365℃ interval	2.70	2.53	2.60	2.60	2.52	2.68
365-395℃ interval	4.35	4.65	4.46	4.52	4.46	4.47
395-425℃ interval	4.94	5.28	5.17	5.03	5.09	4.98
425-460℃ interval	4.28	4.89	4.72	4.43	4.57	4.38
460-475℃ interval	2.43	2.65	2.58	2.46	2.47	2.49
475-500℃ interval	3.49	3.62	3.62	3.59	3.54	3.57
350-500℃ interval	22.19	23.61	23.14	22.61	22.65	22.57

(*p* - Initial blended product; *p** - Blended product with ideal yield; *p(m)* - Blended product optimized by the model proposed in this paper; *p(m1), p(m2), p(m3)* - Blended products were optimized by the models proposed in the other three papers)

Solved by the simulation program, the gap of gasoline fraction yield between *p(m)* and *p** is 0.23, and the yield of the fraction used to produce gasoline is increased by 4.02% compared to the original product *p*; the gap of diesel fraction yield between *p(m)* and *p** is 0.62, and the yield of the fraction used to produce diesel is increased by 0.95% compared to the yield of the original product *p*. Compared to the model proposed in this paper, the blended products obtained by the other three optimization models have a larger gap with the ideal product, and the target fractional yield is less improved than the initial product: the gap of gasoline fraction yield between *p(m1), p(m2), p(m3)* and *p** is 2.27, 3.52, 2.45, and the yield of the fraction used to produce gasoline is increased by 2.99%, 1.73%, 2.82% compared to the original product *p*; the gap of diesel fraction yield between *p(m1), p(m2), p(m3)* and *p** is 1.00, 0.94, 1.04, and the yield of the fraction used to produce diesel is increased by 0.42%, 0.46%, 0.38% compared to the yield of the original product *p*. The comparison between the *p(m1), p(m2), p(m3)* and *p(m)* is shown in Figure 2 and Figure 3.

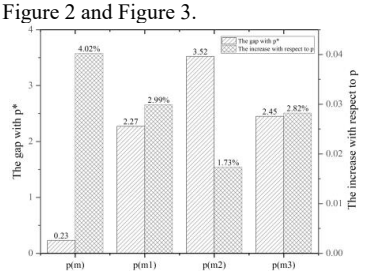


Figure.2. Simulation result comparison of the first situation

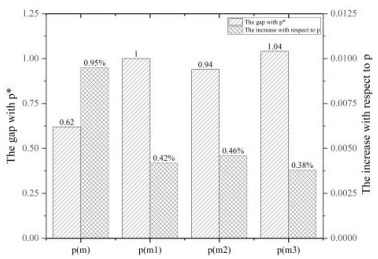


Figure.3. Simulation result comparison of the second situation

According to the comparison of simulation results, the optimization model proposed by us can effectively reduce the yield gap between the blended product and the expected ideal product, and ensure that the properties of the blended product are stable. And in different cases, our model improved the yield of the blended product in the ideal fraction, effectively optimized the product distribution of the blended product after being processed by the CDU unit, and then met the material requirements of the subsequent secondary processing unit. Combined with the previous simulation results, we verified the effectiveness of the proposed model in solving the actual blending formula optimization problem, and can reduce the fluctuation of distillate properties, improve the yield of distillate required by the secondary processing unit, greatly improve the production efficiency of oil refining enterprises, and ensure the stable operation of production and processing.

4. Conclusions

This paper first briefly summarizes the main research results and progress in the field of crude oil blending optimization in recent years. Subsequently, this paper considers the constraints of typical scheduling issues such as material balance, device constraints, and inventory constraints in the crude oil blending process. On this basis, it also considers the product yield, sulfur content, nitrogen content, heavy metal content and molecular structure constraints. These constraints are closely related to the secondary processing process. A MINLP model based on continuous-time representation is described in detail through constraints and variables. Finally, through simulation, this paper verifies that the model has certain advantages in maximizing the yield of the target fractionation interval. It can provide refiners with more accurate blending plans, optimize blended crude oil formulas, and meet the demands of subsequent secondary processing devices. However, current research work also has shortcomings. The model we proposed does not consider the economic benefits of the entire blending process, which leads to the higher cost of the crude oil blending formula. In the next stage, we will consider the multi-objective optimization of the economic benefits and properties of the crude oil blending process.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (No. 22178383, 21706282), and the Research Foundation of China University of Petroleum (Beijing) (No. 2462020BJRC004, 2462020YXZZ023).

Reference

B.Beach, R.Hildebrand, K.Ellis, B.Lebreton.
An Approximate Method for the Optimization of Long-Horizon Tank Blending and Scheduling Operations.
Computers & Chemical Engineering (2020)
DOI: 10.1016/j.compchemeng.2020.106839

H.Yang, D.E.Bernal, R.E.Franzoi, F.G.Engineer, K.Kwon, S.Lee, I.E.Grossmann.
Integration of Crude-oil Scheduling and Refinery Planning by Lagrangean Decomposition.
Computers & Chemical Engineering(2020)
DOI: 10.1016/j.compchemeng.2020.106812

M.Li.
Refinery Operations Optimization Integrated Production Process and Gasoline Blending.
Journal of Physics: Conference Series(2020)
DOI: 10.1088/1742-6596/1626/1/012111

F.Bayu, M.A.Shaik, M.Ramteke.
Scheduling of Crude oil Refinery Operation with Desalting as a Separate Task.
Asia-Pacific Journal of Chemical Engineering (2020)
DOI: 10.1002/apj.2539

F.Li, F.Qian, W.Du, M.Yang, J.Long, V.Mahalec.
Refinery Production Planning Optimization under Crude oil Quality Uncertainty.
Computers & Chemical Engineering(2021)
DOI: 10.1016/j.compchemeng.2021.107361

Y.Zhao, N.Wu, Z.Li.
A Novel Solution Approach to Scheduling of Crude-oil Operations Based on An Improved Continuous-time Formulation.
2017 IEEE 14th International Conference on Networking, Sensing and Control. May 16-18, 2017, Calabria, Italy.
DOI: 10.1109/ICNSC.2017.8000140

P.A.C.Castillo, P.M.Castro, V.Mahalec.
Global Optimization of Nonlinear Blend-scheduling Problems.
Engineering(2017)
DOI: 10.1016/J.ENG.2017.02.005

B.C.Menezes, I.E.Grossmann, J.D.Kelly..
Enterprise-wide Optimization for Operations of Crude-oil Refineries: Closing the Orocurement and Scheduling Gap.
Computer Aided Chemical Engineering(2017)
DOI: 10.1016/B978-0-444-63965-3.50210-5

B.C.Menezes, J.D.Kelly.
High-quality Blend Scheduling Solution for Sizing, Selecting, Sequencing, Slotting and Spotting in the Processing Industries.
Computer Aided Chemical Engineering(2019)
DOI: 10.1016/B978-0-12-818634-3.50303-9

R.E.Franzoi, B.C.Menezes, J.D.Kelly, J.W.Gut.
Effective Scheduling of Complex Process-shops Using Online Parameter Feedback in Crude-oil Refineries.
Computer Aided Chemical Engineering(2018)
DOI: 10.1016/B978-0-444-64241-7.50208-1

R.E.Franzoi, B.C.Menezes, J.D.Kelly, J.W.Gut.
Design for Online Process and Blend Scheduling Optimization.
Computer Aided Chemical Engineering(2019)
DOI: 10.1016/B978-0-12-818597-1.50030-8

C.A.Mendez, I.E.Grossmann, I.Harjunkoski, P.Kabore.
A Simultaneous Optimization Approach for Off-line Blending and Scheduling of Oil-refinery Operations.
Computers & Chemical Engineering (2006)
DOI: 10.1016/j.compchemeng.2005.11.004.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50182-2

Novel Design of Optimum Heat Exchanger Networks for Textile Dyeing Process to Maximize Wastewater Heat Recovery Efficiency

Yurim Kim[a§], Jonghun Lim[a,b§], Hyungtae Cho[a], Juwon Lee[a,b], Il Moon[b], Junghwan Kim[a*]

[a]*Green Materials and Process R&D Group, Korea Institute of Industrial Technology, 55, Jonga-ro, Jung-gu, Ulsan 44413, Republic of Korea*
[b]*Department of Chemical and Biomolecular Engineering, Yonsei University, 50, Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea*
[§]*Yurim Kim and Jonghun Lim contributed equally to this work as first authors.*
kjh31@kitech.re.kr

Abstract

The current dyeing process uses steam to heat the fresh water used for reactive dyeing. After the dyeing process, the wastewater is forcibly cooled and discharged for biological treatment. A large amount of energy is consumed for heating the fresh water and cooling the wastewater owing to the absence of a heat recovery process. This study suggests a novel design of optimum heat exchanger networks for textile dyeing process to maximize wastewater heat recovery efficiency. First, a process model was developed by integrating a heat exchanger with a heat pump. The model was designed to recover high-temperature and low-temperature wastewater through a heat exchanger and a heat pump, respectively. Second, the heat exchanger network of the developed process model was retrofitted based on pinch analysis to reduce operating costs. From the simulation results, the hot and cold utility consumption of the proposed model was quantitatively analyzed. Finally, a techno-economic analysis (TEA) was conducted to confirm the appropriateness of the proposed process model. The total annualized cost (TAC), based on the equivalent annual cost and the total product cost, was determined for both the current and potential future economies. As a result, the TAC was reduced by 43.2 %, and the economic efficiency of the proposed energy-saving measures demonstrates a payback period (PBP) of up to 0.65 y.

Keywords: Textile dyeing process; Wastewater heat recovery; Modeling, Analysis, and Simulation; Heat integration; Techno-economic analysis.

1. Introduction

The exhaust dyeing method is commonly used in the textile dyeing process for increased productivity. Here, a reactive dye is dissolved in a solution containing cellulose fibers, and is fixed to the surface of the fiber. The advantage of this method is that a large amount of fibers can be dyed over a short period. However, the fresh water must be heated for reactive dyeing, and the wastewater must be forcibly cooled for biological treatment after the dyeing process, resulting in high operating costs. To reduce the high operating costs, active research has been undertaken recently to recover the heat from the large amount of wastewater discharged and reuse it in the dyeing process. This is because the discharged wastewater has a significant amount of heat that can be recycled, so if it is used to preheat fresh water, operating costs for heating fresh water and cooling wastewater can be

reduced. Rakib et al. (2017) reduced the operating costs of 47,100 $/y by installing a counter-flow heat exchanger to utilize the wastewater heat during the dyeing process. They confirmed the feasibility of installing a heat exchanger through a short payback period (PBP) of approximately 11 months. Pulat et al. (2009) reduced the total annualized cost (TAC) by 423,837 $/y by designing a system that recovered the heat of the collected wastewater. They confirmed the validity of the wastewater heat recovery (WWHR) system through a PBP of approximately four months.

Despite the contributions of conventional WWHR systems to increase energy efficiency, several challenges remain. First, because the textile industry discharges wastewater at different temperatures over time, the energy efficiency of the WWHR system varies depending on how the wastewater heat is recovered. However, few studies have considered the various method of recovering heat from wastewater. Second, the conventionally developed WWHR system is not optimally configured, so the energy recovery efficiency cannot be maximized. Thus, heat recovery from wastewater in different ways and maximizing the energy recovery efficiency by optimally configuring a WWHR system is crucial.

To address these challenges, we propose a novel design of optimum heat exchanger networks for textile dyeing process to maximize WWHR efficiency. First, to improve the energy efficiency of the WWHR system, a process model was designed to recover high-temperature wastewater through a heat exchanger and low-temperature wastewater through a heat pump. Second, to maximize the energy efficiency of the wastewater, the heat exchanger network of the developed process model was modified through pinch analysis. Finally, techno-economic analysis (TEA) was conducted to confirm the appropriateness of the proposed process.

The novelty of this study is as follows. First, the WWHR system integrated a heat pump with a heat exchanger is proposed for the first time to improve heat recovery efficiency. Second, since the developed WWHR system was modified by pinch analysis to maximize heat recovery efficiency, this study can contribute to economic and environmental improvement in the textile dyeing industry.

2. Process description

Currently, the textile industry is largely classified into the pre-treatment, dyeing, and after-treatment processes to improve the quality of the final dyed products. First, in the pre-treatment process composed of bleaching, washing, and acidification, bleaching agents are used to remove impurities in the fibers that interfere with dyeing and to make the fibers white. Second, in the dyeing process, the fibers whose impurities have been removed are dyed. Finally, in the after-treatment process, consisting of cold rinsing, washing, hot rinsing, and finishing, reactive dyes and residues in the dyed fabric are removed. Table 1 shows the operation time, discharged wastewater temperature, and flow rate during each stage of the textile industry.

Table 1. The operation time, temperature and flow rate during each stage of the discharged wastewater

	Process	Process time [min]	Temperature [℃]	Mass flow rate [kg/h]
1	Bleaching	30	96	2,900
2	Washing	20	96	4,350
3	Acidification	10	50	8,700

4	Dyeing	60	96	1,450
5	Cold rinsing	10	30	8,700
6	Washing	20	90	4,350
7	Hot rinsing	10	70	8,700
8	Finishing	20	40	4,350

The dyeing process was performed in several steps. First, the fresh water was heated from 18 °C to 60 °C before the dyeing process, and then with reactive dyes, additives, and fabrics, it was heated to 96 °C to operate the dyeing process. Finally, because the wastewater discharged at each stage contains chemicals such as reactive dyes, discharging without additional treatment can have a fatal impact on the environment. Therefore, to treat the dyeing wastewater discharged from each stage, it was cooled to 40 °C.

3. Methodology

3.1. Development of the process model

In this study, Aspen Plus V 11.0 was used to model a process model of the WWHR system. In the process model, the Peng-Robinson equation was used as the thermodynamic model to consider the equilibrium among the various mixed substances in dyeing wastewater.

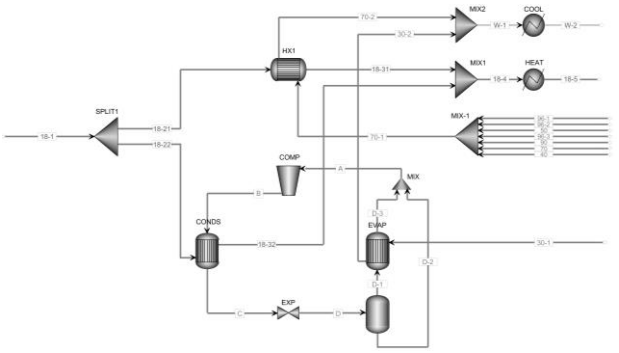


Figure 1. Process model of the WWHR system

Figure 1 presents the process model of the WWHR system. To recover the heat from the wastewater discharged at different temperatures and flow rates over time efficiently, a heat pump was used when the temperature of the wastewater was below 30 °C. Above this temperature, the heat was recovered through a heat exchanger (Kemp, 2007). Finally, to satisfy the target temperature of wastewater and fresh water, they were passed through a cooler and a heater, respectively.

3.2. Model modification by pinch analysis

The process model of the WWHR system was modified by pinch analysis to increase its energy efficiency. This method is used to modify the configuration of the heat exchanger network (HEN) to increase the efficiency of the heat recovery process. The more the energy recovery from the heat exchange between the streams, the lower the energy consumption, thus lowering the operating costs. However, the capital cost of the heat recovery process increases with increasing the heat exchange area to exchange more heat; thus, there is a trade-off relationship between the capital cost and the operating cost. Therefore, it is crucial to apply the supertargeting method that derives the targets of

designing a cost-optimal heat recovery process with improved energy efficiency to pinch analysis.
In this study, first, after deriving the cost-optimal pinch point through supertargeting, the energy target, heat transfer area target, the number of units target, and cost targets for the process modification were obtained. Then, the process was modified according to the targets by re-sequencing and re-piping existing heat exchangers, installing a new heat exchanger, and adding a stream split.

3.3. Techno-economic analysis

In this section, the TEA was conducted to confirm the feasibility of the proposed WWHR system through pinch analysis. The TAC and PBP of the proposed WWHR system were calculated by considering both the capital cost and the operating costs. The equations for the TAC and PBP are as follows (Eqs. 1-3) (Lim et al., 2021):

$$TAC = AF \times CC + OC \tag{1}$$

$$AF = \frac{\left(\frac{ROR}{100}\right) \times \left(1 + \frac{ROR}{100}\right)^{PL}}{\left(1 + \frac{ROR}{100}\right)^{PL} - 1} \tag{2}$$

$$PBP = \frac{CC}{OC - OC_0} \tag{3}$$

where CC and OC represent the capital cost and operating cost of WWHR system, respectively. AF, ROR, and PL represent the annualization factor, rate of return, and plant life, respectively. This study assumed a ROR of 10 % and a PL of 5 years. OC_0 represents the operating cost of the conventional dyeing process in the textile industry without a WWHR system.

4. Results and discussion

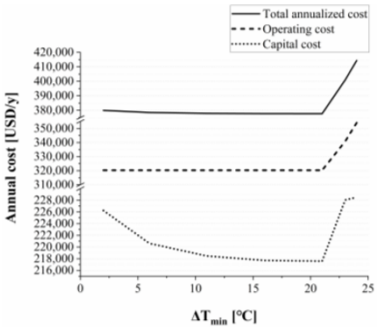


Figure 2. The cost targets of the HEN according to ΔT_{min}

Table 2. Model modification targets at cost-optimal ΔT_{min}

Targets	Units	Value
Heat exchanger area	m^2	25.9
The number of units	-	4
Hot utility	kJ/h	3,916,942
Cold utility	kJ/h	-
Capital cost	$/y	217,569
Operating cost	$/y	320,285
TAC	$/y	377,679

Figure 2 shows the cost targets of the HEN of the WWHR system process model according to ΔT_{min}. First, the operating cost when ΔT_{min} of the HEN falls below 21.46 °C

does not change by the threshold problem requiring only hot utilities because the heat exchange between the streams is sufficient. However, if ΔT_{min} is higher than 21.46 ℃, the operating cost increases because both hot and cold utilities are required. Therefore, when ΔT_{min} falls below 21.46 ℃, the operating cost does not change, but it increases when ΔT_{min} is higher than 21.46 ℃. Second, the capital cost of the HEN increases as ΔT_{min} decreases. If ΔT_{min} becomes higher than 21.46 ℃, the capital cost of the HEN also increases because both the heater and the cooler are required. Therefore, the cost-optimal ΔT_{min} of the HEN of the proposed process model was 21.46 ℃. Table 2 lists the model modification targets at the cost-optimal ΔT_{min}.

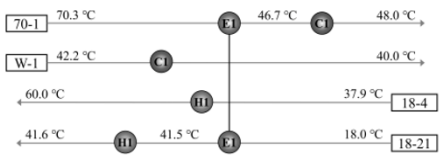


Figure 3. Grid diagram – HEN of the WWHR system before modification

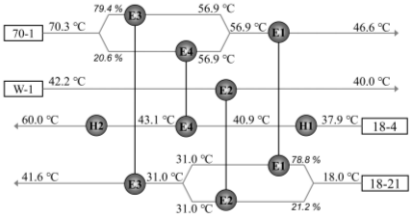


Figure 4. Grid diagram – HEN of the WWHR system after modification

Figure 3 shows the grid diagram – HEN of the WWHR system before modification, and Fig. 4 shows the grid diagram – HEN of the WWHR system after modification according to Table 2.

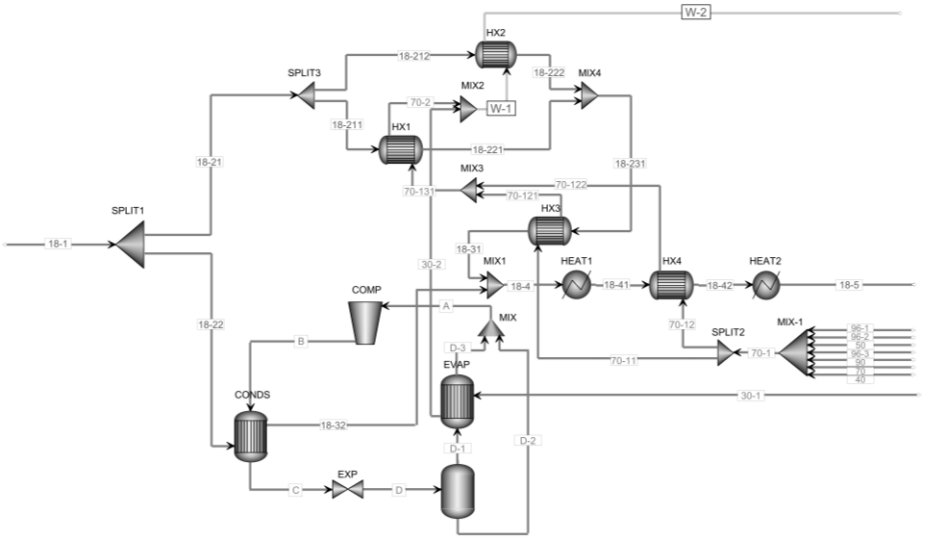


Figure 5. Modified process model of the WWHR system

Table 3. The specification of the modified process model of the WWHR system

	Units	Value
Heat exchanger area of HEN	m^2	27.3
The number of units of HEN	-	6
Hot utility of HEN	kJ/h	3,916,942

Cold utility of HEN	kJ/h	-
Total capital cost	$	217,569
Total operating cost	$/y	320,285
TAC	$/y	377,679

Figure 5 shows the modified process model of the WWHR system, and Table 3 shows the specifications of the HEN modification results. As a result, the TAC of the suggested WWHR system was $ 377,679, which reduced the TAC by 43.2 % compared to the conventional textile industry. Also, the PBP of the proposed WWHR system was 0.65 y, confirming the feasibility in applying the WWHR system to the textile industry.

5. Conclusions

In this study, we propose a novel design of an optimum WWHR system in the textile industry to maximize energy efficiency. This study makes two major contributions to the literature. First, this work is the first attempt to integrate a heat exchanger with a heat pump to effectively recover wastewater heat. Second, a pinch approach is suggested to maximize the energy efficiency of the WWHR system. Through this novel 2-step approach, the optimum WWHR system reduced the TAC by 43.2 % compared to the conventional textile industry without the WWHR system, and the PBP was 0.65 y confirming the validity of the proposed system. This study provides a guideline for future WWHR system designs.

Acknowledgements

This work was supported by the Korean Institute of Industrial Technology within the framework of the following projects: “Development of Global Optimization System for Energy Process [grant number EM-21-0022, IR-21-0029, IZ-21-0052]” and “Development of complex parameter smart analysis modules for color customering [grant number EH-21-0008]”

References

M.I. Rakib, R. Saidur, E.N. Mohamad, A.M. Afifi, Waste-heat utilization – The sustainable technologies to minimize energy consumption in Bangladesh textile sector, J. Clean. Prod. 142 (2017) 1867–1876. https://doi.org/10.1016/j.jclepro.2016.11.098.

E. Pulat, A.B. Etemoglu, M. Can, Waste-heat recovery potential in Turkish textile industry: Case study for city of Bursa, Renew. Sustain. Energy Rev. 13 (2009) 663–672. https://doi.org/10.1016/j.rser.2007.10.002.

I.C. Kemp, 3 - Data extraction and energy targeting, in: I.C. Kemp (Ed.), Pinch Anal. Process Integr. (Second Ed., Second Edi, Butterworth-Heinemann, Oxford, 2007: pp. 41–98. https://doi.org/https://doi.org/10.1016/B978-075068260-2.50008-0.

Lim, J., Jeong, S., Kim, J., Deep neural network-based optimal selection and blending ratio of waste seashells as an alternative to high-grade limestone depletion for SOX capture and utilization, Chem. Eng. J. 133244 (2021). https://doi.org/10.1016/j.cej.2021.133244

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50183-4

Study on the Kinetic Parameters of Crystallization Process Modelled by Partial Differential Equations

Jiali Ai[a], Jindong Dai[a], Jianmin Liu[a], Chi Zhai[b*] and Wei Sun[a*]

[a]*College of Chemical Engineering, Beijing University of Chemical Technology, North Third Ring Road 15, Chaoyang District, Beijing and 100029,China*
[b] *Kunming University of Science and Technology, 727 South Jingming Road, Chenggong District, Kunming and 650500, China*
zhaichi@kmust.edu.cn, sunwei@mail.buct.edu.cn

Abstract

The physical properties of the crystal are affected by the crystal morphology, and the morphology will be affected by the crystallization environment. To obtain the expected physical properties, it is necessary to study the relationship between crystal growth conditions and crystal morphology. As snowflake is rich in crystal morphology, it is very suitable for studying this relationship. Crystal growth kinetics can be used to explore the mechanism of how the crystal growth conditions influence on crystal morphology. However, due to the impact of growth conditions in both time and space coordinates, it is difficult to identify the kinetic parameters of the crystallization process. In this work, the dynamic equation of snowflake growth was established through data regression, and the relationship between snowflake morphology and snowflake growth conditions was studied. In general, cellular automata (CA) is used to simulate the growth process of snowflakes, and the data of whether different positions are in crystallization state and the change of water vapor density with time are extracted, so as to avoid the interference of complex growth conditions on the growth data. Then, data are regressed from the perspective of the reaction-diffusion system, and the crystallization kinetics in the form of partial differential equation (PDE) related to time and space is obtained. This equation is solved by finite difference method to simulate the complex snowflake morphology under different conditions, to analyse the influence of different reaction rate and diffusion rate on the crystal morphology. This work provides a new way to study how crystal morphology is impacted by process parameters, which could be a reference for the planning of crystallization experiments to obtain the specified crystal morphology.

Keywords: Crystal morphology; Reaction-diffusion system; Crystal growth kinetics; Finite difference method.

1. Introduction

Crystallization is an important unit operation in the chemical industry. Different growth conditions during the crystallization process, such as solvent, temperature, supersaturation, etc., will impact the crystal morphology, which is an important factor to determine the physical properties of crystals, including melting point, solubility and dissolution rate (Variankaval et al., 2008). These changes not only affect subsequent process operations, but also cause differences in crystal quality. Therefore, how to prepare crystals with specific morphology is becoming more and more important in academia and industry (Yang and Han, 2016).

Crystal growth kinetics can be employed to predict the crystal morphology by its growth conditions, so it is very helpful for exploring methods to control the morphology of crystals. However, it is difficult to find a suitable growth kinetics equation to describe it. With the development of computer technology and data-based methods, it provides a new way to find the kinetic equation for crystal growth, bypassing the complicated mechanism. Nonlinear dynamic partial differential equation function identification (PDE-FIND) is such a method proposed by Samuel H. Rudy in 2017. It finds the governing partial differential equation(s) of a given system by sparse regression of time series measurements in spatial domain and has been successfully applied to reaction-diffusion systems. From the perspective of systems engineering, the crystallization process can be regarded as a generalized reaction-diffusion system, as the process of solute attaching to the surface of the growing crystal and changing phase is considered as a reaction, and the process of solute moving randomly to the surface of crystal is treated as diffusion. Both are important factors affecting the crystal morphology (Wang et al., 2013). Therefore, this method can be used to explore the crystallization kinetic equation.

Snowflakes are ice crystals that grow from water vapor. They form a large number of exquisite, symmetrical, and complex patterns in atmosphere, with rich crystal morphologies (Libbrecht, 2017). At the same time, it is almost only affected by the density of water vapor in the atmosphere. Therefore, snowflake is a good research object for exploring crystallization conditions and crystal morphology. In order to obtain the crystallization kinetic equation, it is necessary to obtain the dynamic information of solid-liquid two phases, especially the accurate measurement of water vapor density data. However, there are several difficulties in obtaining the accurate real-time data because of the error of the crystal growth in the experiment, and the challenges of identifying the crystallinity, concentration and density in the solution from the actual images. With the fast development of computer technology, crystallization process can be simulated with appropriate computational tools. Therefore, it is possible to study and fit the kinetic equation based on simulated crystallization data. In our previous work, the snowflake growth process was simulated by cellular automata (CA) (Liu et al., 2020).

In this work, the relationship between crystal growth conditions and crystal morphology, especially the effects of diffusion coefficient and reaction coefficient on crystal morphology, is discussed by establishing a kinetic equation, which is obtained from the process data of snowflake growth. The paper is organized as follows: in the Section 2, how to obtain water vapor density data through simulated images is introduced, including a CA method for simulating an ideal snowflake formation system. Then, PDE-FIND algorithm is used to find the dynamic equation of this process, which is put into the Section 3. In the Section 4, the kinetic equation is divided into two parts: diffusion and reaction, and the effects of each part on snowflake morphology are discussed respectively. The last part is a summary of the full work.

2. Data for snowflake formation process: CA simulation

For the snowflake formation process, its morphology is closely related to the water vapor density. If the water vapor density during the formation of snowflakes is recorded as an image, a series of real images of crystal growth process can be used as data for studying crystal crystallization kinetics. However, it is still hard to experimentally obtain such images in time series with the required quality for model establishment according to literatures. Here, images of snowflake growth under ideal conditions simulated by CA were chosen as the substitute for the real picture of snowflake growth.

First of all, in order to establish the model, the ideal system is framed by selecting fewer variables and parameters for research. The system is assumed with a constant temperature and unaffected by impurities. Without considering the influence of stirring and the external environment of the crystallizer on the crystallization process and crystal dissolution, the ideal process of crystallization of water vapor into snowflakes is simulated. According to the conservation of mass and classical diffusion theory, the rules of CA are established. The model mainly consists of two variables, water vapor density α and three-dimensional increment β. References for specific simulation methods is available in the work of Liu (2020) et al. . Here, the model parameters $\alpha = 0.4$ and $\beta = 0$ are selected, and the result is shown in Fig.1.

N=0 N=975 N=1950 N=2925 N=3900

Fig.1. Simulated time series of snowflake growth by CA, where water vapor density $\alpha = 0.4$ and three-dimensional increment $\beta = 0$. N represents different simulation time steps.

Matlab is used to encode and integrate the data. By identifying the grey value of each point in the simulated image, the water vapor density and crystallization state are obtained. Therefore, an image can be converted into a two-dimensional data matrix, in which each element of the matrix corresponds to one pixel, and a large three-dimensional array is obtained from a series of images with equal time intervals. In order to show the snowflake growth mechanism better and remove the influence of the background on PDE recognition, only the 80 × 80 pixel area in the centre of images are identified, and the images here are 100 images with 10 time step interval.

3. The dynamic equation of snowflake formation based on data: the establishment of partial differential equations

In order to obtain the dynamical equations of the snowflake reaction-diffusion system, the PDE-FIND algorithm proposed by Rudy et al. is applied to obtain the partial differential equations of time and space. For this algorithm program, the data at each position in the time series are collected and numerically differentiated, and then the data are combined into a large matrix, and PDE candidates are also combined. Specific calculation methods and mathematical details are in Rudy 's work. The data in the three-dimensional array obtained in the second part are imported into the PDE-FIND algorithm program and got the PDEs as follows.

$$\frac{dn}{dt} = a_1 n^3 + b_1 n^2 + c_1 n + d_1 + (a_2 n^3 + b_2 n^2 + c_2 n + d_2)\left(\frac{\partial n}{\partial x} + \frac{\partial n}{\partial y}\right) + (a_3 n^3 + b_3 n^2 + c_3 n + d_3)\left(\frac{\partial^2 n}{\partial x^2} + \frac{\partial^2 n}{\partial y^2}\right) + (a_4 n^3 + b_4 n^2 + c_4 n + d_4)\left(\frac{\partial^2 n}{\partial x \partial y}\right) \quad (1)$$

$$a_1 = 1.3 \times 10^{-3} \quad a_2 = -9.2 \times 10^{-2} \quad a_3 = 8.4 \times 10^{-2} \quad a_4 = 3.7 \times 10^{-2}$$

$$b_1 = 1.5 \times 10^{-3} \quad b_2 = 2.0 \times 10^{-2} \quad b_3 = -2.0 \times 10^{-2} \quad b_4 = -1.2 \times 10^{-2}$$

$$c_1 = -1.5 \times 10^{-3} \quad c_2 = 7.3 \times 10^{-2} \quad c_3 = -7.6 \times 10^{-2} \quad c_4 = -2.5 \times 10^{-2}$$

$$d_1 = 0 \quad d_2 = 5.6 \times 10^{-6} \quad d_3 = -2.8 \times 10^{-6} \quad d_4 = -4.2 \times 10^{-6}$$

The finite difference (FD) method is used to discretize and approximate the partial differential term. For the Laplace operator, a five-point difference scheme is adopted, and forward difference method is adopted for the rest. The specific difference format is shown in the Eq.(2-5). The time interval θ is 0.01, and the space interval h is 0.5. The simulation results are shown in Fig.2.

$$\frac{dn}{dt} = \frac{\mathrm{n}(i,j)^{t+1} - \mathrm{n}(i,j)^{t}}{\theta} \tag{2}$$

$$\frac{\partial n}{\partial x} + \frac{\partial n}{\partial y} = \frac{\left(\mathrm{n}(i-1,j)^{t} - \mathrm{n}(i+1,j)^{t} + \mathrm{n}(i,j-1)^{t} - \mathrm{n}(i,j+1)^{t}\right)}{2h} \tag{3}$$

$$\frac{\partial^2 n}{\partial x^2} + \frac{\partial^2 n}{\partial y^2} = \frac{1}{h^2}\left(\mathrm{n}(i-1,j)^{t} + \mathrm{n}(i+1,j)^{t} + \mathrm{n}(i,j-1)^{t} + \mathrm{n}(i,j+1)^{t} - 4\mathrm{n}(i,j)^{t}\right) \tag{4}$$

$$\frac{\partial^2 n}{\partial x \partial y} = \frac{1}{h^2}\left(\mathrm{n}(i-1,j-1)^{t} + \mathrm{n}(i+1,j+1)^{t} + \mathrm{n}(i+1,j-1)^{t} + \mathrm{n}(i-1,j+1)^{t} - 4\mathrm{n}(i,j)^{t}\right) \tag{5}$$

a b1: N=100 b2: N=200 b3: N=300 b4: N=500

Fig.2. Simulated time series of snowflake growth by FD, where water vapor density $\alpha = 0.4$ and three-dimensional increment $\beta = 0$. N represents different simulation time steps. a: compare the pictures in Figure 1 and Figure 2 when the time step is 3900. b: Simulation results at different time steps.

It can be seen that the simulation result in Fig.2(b) is similar to those of CA method in Fig.1. Therefore, the equation is considered to be the kinetic equation of snowflake crystal growth when the water vapor density is 0.4.

4. The influence of process parameters on snowflakes morphology: from the perspective of the reaction-diffusion system

From a mathematical point of view, the reaction-diffusion system can be described by the following equation:

$$\frac{dx}{dt} = \nabla^2 x + F(x) \tag{6}$$

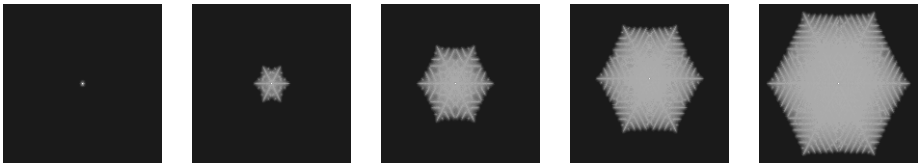
According to Eq.(6), the position-independent part in Eq.(1) is regarded as the reaction term denoted by $g(n)$, and the position-related part is regarded as the diffusion term denoted by $f(n)$. Then $f(n)$ and $g(n)$ are as Eq.(7) and Eq.(8), respectively.

$$f(n) = (a_2n^3 + b_1n^2 + c_1n + d_1)(\frac{\partial n}{\partial x} + \frac{\partial n}{\partial y}) + (a_3n^3 + b_1n^2 + c_1n + d_1)(\frac{\partial^2 n}{\partial x^2} + \frac{\partial^2 n}{\partial y^2}) + (a_4n^3 + b_1n^2 + c_1n + d_1)(\frac{\partial^2 n}{\partial x \partial y}) \quad (7)$$

$$g(n) = a_1n^3 + b_1n^2 + c_1n + d_1 \quad (8)$$

4.1. The influence of reaction parameters on the morphology of snowflakes

The reaction related terms were multiplied by 0.9 and 1.1 respectively, and simulated by the finite difference method to study the effect of the change of the reaction coefficient on the morphology of the snowflake. The simulation results are shown in Fig.3 and Fig.4.



N=0 N=140 N=280 N=420 N=560

Fig 3. Simulated time series of snowflake growth by FD, where g(n) × 0.9 and f(n) not changed.

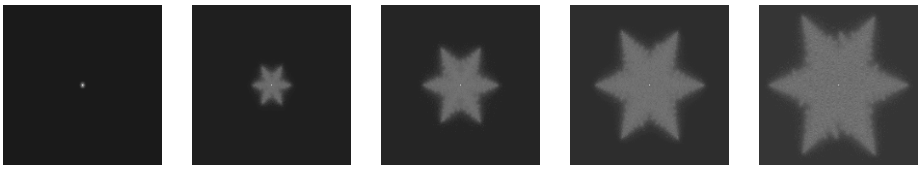
N=0 N=100 N=200 N=300 N=400

Fig 4. Simulated time series of snowflake growth by FD, where g(n) × 1.1 and f(n) not changed.

It can be seen that with the increase of reaction rate, snowflakes tend to form very dense hexagons, while when the reaction rate decreases, snowflakes tend to form hexagons with obvious branching.

4.2. The influence of diffusion parameters on the morphology of snowflakes

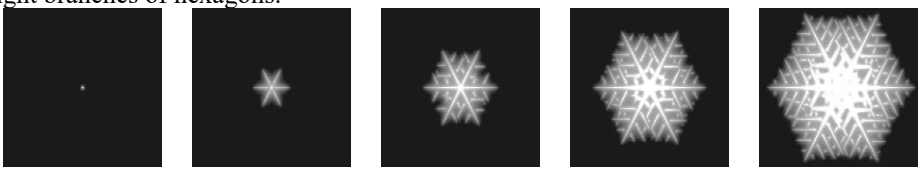
The diffusion related terms were multiplied by 0.99 and 1.01 respectively, and simulated by the finite difference method to study the effect of the change of the deffusion coefficient on the morphology of the snowflake. The simulation results are shown in Fig.5 and Fig.6.



N=0 N=130 N=260 N=390 N=520

Fig 5. Simulated time series of snowflake growth by FD, where f(n) × 0.99 and g(n) not changed. N represents different simulation time steps.

It can be seen that as the increase of diffusion rate, snowflakes tend to form hexagons with obvious branches, while when the diffusion rate decreases, snowflakes tend to form tight branches of hexagons.



N=0 N=120 N=240 N=360 N=480

Fig 6. Simulated time series of snowflake growth by FD, where $f(n) \times 1.01$ and $g(n)$ not changed. N represents different simulation time steps.

5. Conclusions

In this paper, the reaction-diffusion equation of snowflake formation is established by using the method of data regression, and the effects of different reaction rates and diffusion rates on the morphology of snowflakes are studied. The ideal state of the snowflake growth images simulated by CA is used as a substitute for the real image to obtain the time series data in the spatial domain. Meanwhile, by adjusting the diffusion and reaction coefficients in the kinetic equation, different snowflake morphologies are formed. With the increase of the reaction rate, snowflakes tend to form very compact hexagons, and on the contrary, dendrites become sparser. When the diffusion rate is high, snowflakes are more likely to form with distinct branches and buds, and conversely, the branches are denser. In other words, if a crystal product with a sparse network structure is a better choice, some methods could be used to achieve this, such as changing the medium to gel to increase the diffusion rate; If a crystal product with a dense block structure is a better choice, some methods could be used to realize this, such as increasing the degree of supersaturation and lowering the crystallization temperature to increasing the reaction rate. By studying the formation of different crystal morphologies in the crystallization system under the adjustment of diffusion and reaction rate, it is hoped to provide reference for the reasonable synthesis and structural design of materials.

Acknowledgements: The National Natural Science Foundation of China (Grant No. 21878012).

References

K. G. Libbrecht, 2017, Physical Dynamics of Ice Crystal Growth, Annual Reviews, 47, 271-295.

J. Liu, J. Dai, C. Han, J. Zhang, J. Ai, C. Zhai, X. Liu, W. Sun, 2020, Simulation of the Crystallization Process based on Cellular Automata--- Snowflake Formation from Pure Water System, 30th European Symposium on Computer Aided Process Engineering.

S. H. Rudy, S. L. Brunton, J. L. Proctor, J. N. Kutz, 2017, Data-driven discovery of partial differential equations, Science Advances, 3, 4, e1602614.

N. Variankaval, A. S. Cote, M. F. Doherty, 2008, From form to function: Crystallization of active pharmaceutical ingredients, AIChE Journal, 54, 7, 1682-1688.

H. Wang, Y. Han, J. Li, 2013, Dominant Role of Compromise between Diffusion and Reaction in the Formation of Snow-Shaped Vaterite, Crystal Growth & Design, 13, 5, 1820-1825.

T. Yang, Y. Han, 2016, Quantitatively Relating Diffusion and Reaction for Shaping Particles, Crystal Growth & Design, 16, 5, 2850-2859.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50184-6

Graphical user interface for development of dynamics model of fermentation process applying long short-term memory networks

Felipe M. M. Sousa[a], Rodolpho R. Fonseca[b], Flávio V. Silva[a*]

[a]*School of Chemical Engineering, University of Campinas, Albert Einstein Avenue 500, Campinas, 13083-852, Brazil*
[b]*Chemical Engineering Department, Federal University of Sergipe, Marechal Rondon Avenue, São Cristóvão, 49100-000, Brazil*
flaviovs@unicamp.br

Abstract

Fermentation is a complex process highly influenced by many variables whose effects usually are not accounted for in mathematical models, hindering their accuracy. As an alternative to surpass this issue, empirical models, especially deep learning algorithms, can extract abstract hidden information from real data, enhancing the model accuracy. We present a freeware graphical user interface that enables any user to develop a long short-term memory (LSTM) network to model the dynamics of a fermentation process based on macro variables usually monitored in industrial plants. The algorithm is coded in Python, ensuring platform independence, easy installation, and flexibility to include new variables if required. It also allows the use of the model as a software sensor, assisting in the decision-making of changes. The results are provided in graphics that can be saved as high-definition images or in spreadsheets.

Keywords: Fermentation, Long Short-Term Memory, Graphical User Interface.

1. Introduction

Fermentation is a biotechnological process widely employed in the pharmaceutical, food, and energy industries. Although fermentation and its phenomenological models have been extensively studied over the years, the continuous operation is highly susceptible to the effects of several unaccounted variables, making the phenomenological models less accurate due to simplifications. In this regard, empirical modeling based on deep learning algorithms represents a viable alternative to phenomenological modeling and permits the introduction of the effects of important operational factors. These deep learning models can be obtained from information commonly monitored through the supervisory control system and saved in highly dense databases, making the generated models more relatable to the real system.

Among the deep learning algorithms, Long Short-Term Memory (LSTM) networks represent a powerful tool for developing dynamic models, due to the introduction of recurrent patterns of the predictions through the network. Moreover, LSTM networks are less amenable to problems often observed during the optimization phase in comparison to other recurrent networks, guaranteeing a more reliable model. These networks were successfully applied for generating models in many fields such as the stock market (Althelaya, El-Alfy and Mohammed, 2018; Ghosh, Neufeld and Sahoo, 2021), rainfall-

runoff (Yin *et al.*, 2021), prediction and reconstruction of ocean wave heights (Jörges, Berkenbrink and Stumpe, 2021), COVID-19 transmission (Ibrahim *et al.*, 2021; Luo *et al.*, 2021) and fermentation process (Sousa, Fonseca and Silva, 2021), indicating their high applicability and accuracy in complex-forecasting processes. Although widely studied, this technique does not seem to be made available for larger audiences, especially in chemical and biochemical plants.

This work aimed at developing an intuitive graphical user interface (GUI) for users unfamiliar with deep learning techniques that permits the generation of LSTM models for the prediction of volume, cell content, and substrate and ethanol concentrations (referred to as endogenous variables from now on) in a fermentation process based on the user's available database concerning flowrate, cell content and substrate concentration (exogenous variables) in the inlet flow.

2. Program overview and technical description

The developed GUI focus on 1) facilitating the acquisition of transient models for the fermentation process employing LSTM networks, 2) applying the real data obtained from the process, 3) providing qualitative and quantitative performance analysis of the models, and 4) applying the model as a software sensor.

The software was developed using Python, an interpreted language, ensuring a free tool with easy installation and platform independence. The program is distributed as a source code, permitting the addition and change of the model variables if necessary. As an interpreted language, the changes can be written and tested in small section and does not require compilation.

The LSTM models are developed using Tensorflow and Keras, specialized and free libraries for deep learning applications. Besides these, a few other libraries are utilized during the complete run of the GUI and must be installed on the computer. A description of all required libraries is available in a READ.ME file that accompanies the source code.

3. The Graphic User Interface (GUI)

The main window of the program (Figure 1) is divided into two sections. At the top left corner, an entry bar requires the specification of the number of timesteps used in an already saved model employed that will be utilized as a software sensor. This information is needed before the selection of any other option concerning the software sensor application. A button bar to invoke the different program routines is placed right below the entry bar. These routines include data acquisition, model training, performance assessment, saved model upload, and software sensor prediction.

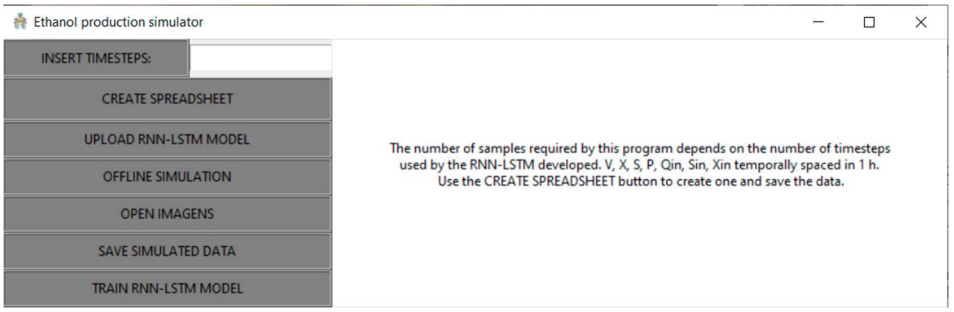


Figure 1 – The main window of the program.

The hyper-parameters for the development of a new LSTM model can be set in a pop-up window, as illustrated in Figure 2. This GUI permits to choose the number of timesteps for the new model, neurons in the output layer of each LSTM gate, weights for the L1 and L2 regularizations, dropout rate, initial weight distribution, percentage of training and test split, the maximum number of epochs for training, batch size, percentage of the training dataset saved for using in early-stopping algorithm and its maximum number of successive errors to cease the training.

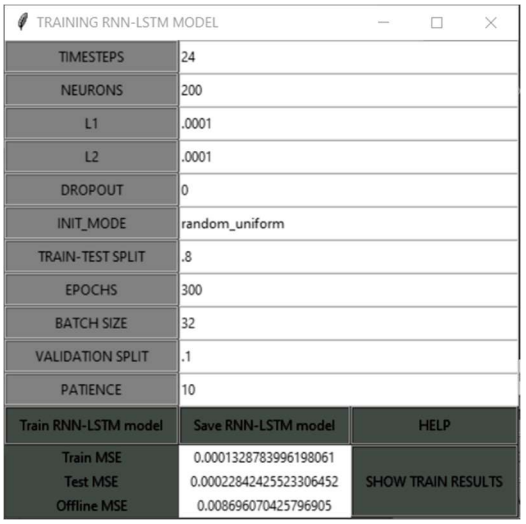


Figure 2 – Pop-up window for setting the network's hyper-parameters during the training stage.

The database for training must be provided in an Excel® spreadsheet and can have any number of vectors. Due to the transient nature model, the division between training and test datasets is sequential, as presented in the spreadsheet. The data is kept unshuffled during the training and test stages.

The program provides a help window (Figure 3) that briefly explains the function of each hyper-parameter aforementioned and their valid inputs. It also illustrates the structure that must be followed in the spreadsheet containing the complete database.

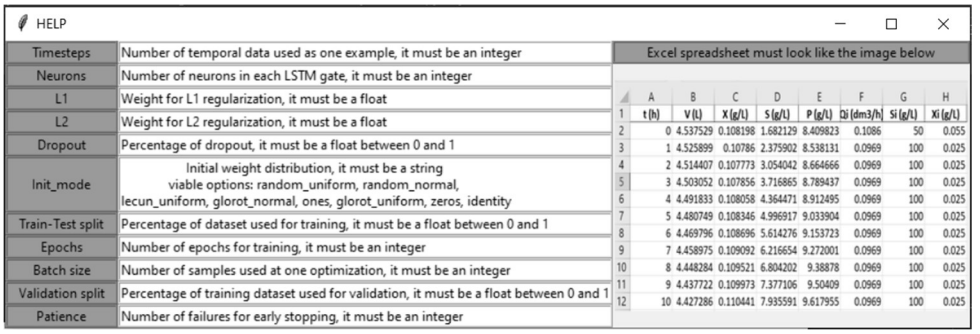


Figure 3 – Help menu for selection of the LSTM model's hyper-parameters.

The use of the GUI for software sensor applications demands an initial state that must be provided in a pop-up window (Figure 4). This window is accessed through the 'create spreadsheet' button available in the main menu (Figure 1). The number of rows is determined by the timestep value defined in the entry bar of the main window, and this information can be quickly acquired from an Excel® spreadsheet. The GUI also requires uploading files in the formats .json and .h5 specifying the LSTM structure and weight values. It must be mentioned that the GUI keeps the last values for the exogenous variables throughout the length of the prediction.

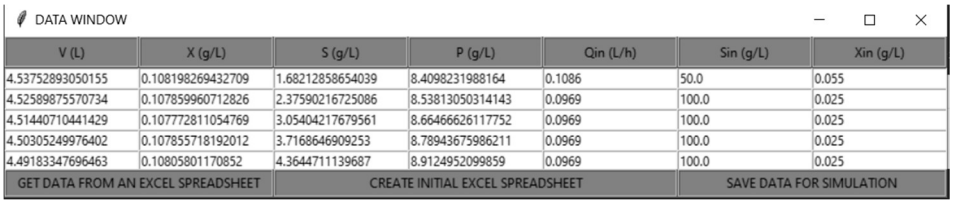


Figure 4 – Definition of the initial state for software sensor application.

4. Network performance and results visualization

The results from the training stage can be visualized in the window shown in Figure 5 and accessed through the 'show train results' button (Figure 2). The window generates comparative graphics showing the performance of the model for all four predicted variables using the training and test databases and the cost function using the training and validation sets. The test database can also be used and compared in the software sensor application. All graphics can be saved in PNG format and 600 dpi. The window also allows the user to zoom in and change specific graph settings.

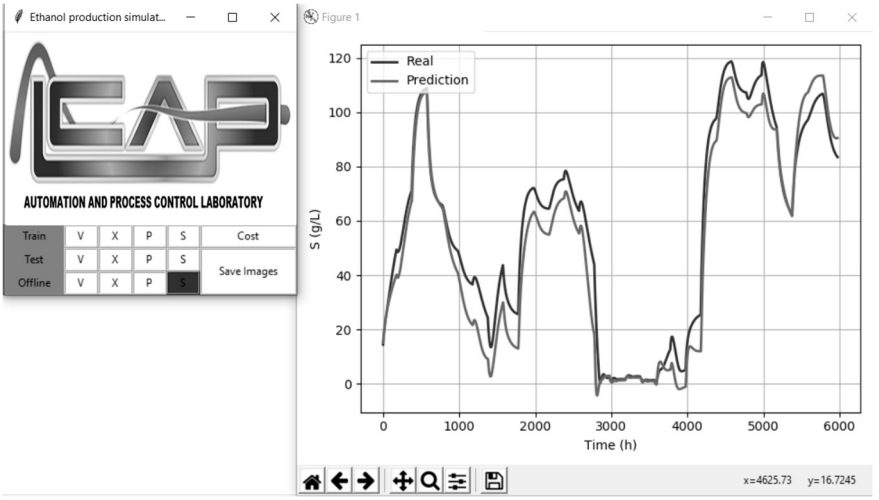


Figure 5 – Results window for training stage and generalization analysis.

Figure 6 illustrates the results window for the software sensor application. This window, accessed through the 'offline simulation' button (Figure 1), shows the predicted results for the endogenous variables using a previously saved model and new data acquired. The results can be viewed, zoomed and saved as PNG images or in an Excel® spreadsheet.

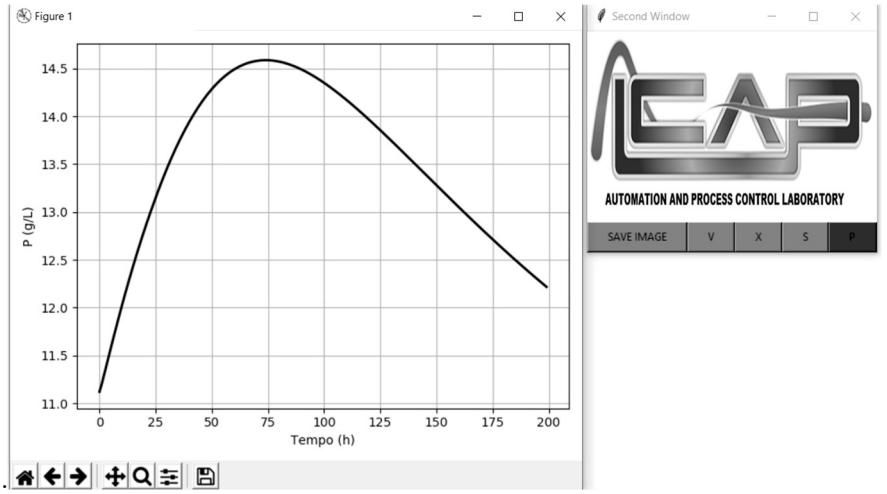


Figure 6 – Results window for prediction as software sensor.

5. Availability

The GUI described is available free of charge to all users and can be used in any operating system to run Python code. The program can be obtained at https://github.com/msousaeq/etGUI.

6. Conclusions

In this work, we developed an easily used Graphical User Interface employing Python code for obtaining the dynamics model of fermentation processes using long short-term memory networks.

The GUI allows the user to train a new model based on their database, defining the major hyperparameters of the model. Its performance is quickly assessed quantitative and qualitatively, saving the generated model for future uses. The rendered graphics for the qualitative analysis of all output variables and the cost function throughout the optimization stage can be saved, zoomed in, and changed.

The GUI also permits the use of the LSTM model as a software sensor for longer predictions, assisting in the decision-making of changes in operating conditions of the fermentation process from an initial state. The results obtained in this application can be graphically analyzed and saved in a spreadsheet for posterior uses.

Acknowledgments

The authors appreciate the financial support provided by the National Council for Scientific and Technological Development– CNPq (Grant number: 131783/2019-0).

References

Althelaya, K. A., El-Alfy, E.-S. M. and Mohammed, S. (2018) 'Stock Market Forecast Using Multivariate Analysis with Bidirectional and Stacked (LSTM, GRU)', in *2018 21st Saudi Computer Society National Computer Conference (NCC)*. IEEE, pp. 1–7. doi: 10.1109/NCG.2018.8593076.

Ghosh, P., Neufeld, A. and Sahoo, J. K. (2021) 'Forecasting directional movements of stock prices for intraday trading using LSTM and random forests', *Finance Research Letters*, p. 102280. doi: 10.1016/j.frl.2021.102280.

Ibrahim, M. R. *et al.* (2021) 'Variational-LSTM autoencoder to forecast the spread of coronavirus across the globe', *PLOS ONE*. Edited by C. Rozenblat, 16(1), p. e0246120. doi: 10.1371/journal.pone.0246120.

Jörges, C., Berkenbrink, C. and Stumpe, B. (2021) 'Prediction and reconstruction of ocean wave heights based on bathymetric data using LSTM neural networks', *Ocean Engineering*, 232, p. 109046. doi: 10.1016/j.oceaneng.2021.109046.

Luo, J. *et al.* (2021) 'Time series prediction of COVID-19 transmission in America using LSTM and XGBoost algorithms', *Results in Physics*, 27, p. 104462. doi: 10.1016/j.rinp.2021.104462.

Sousa, F. M. M., Fonseca, R. R. and da Silva, F. V. (2021) 'Empirical modeling of ethanol production dynamics using long short-term memory recurrent neural networks', *Bioresource Technology Reports*, 15, p. 100724. doi: 10.1016/j.biteb.2021.100724.

Yin, H. *et al.* (2021) 'Rainfall-runoff modeling using LSTM-based multi-state-vector sequence-to-sequence model', *Journal of Hydrology*, 598, p. 126378. doi: 10.1016/j.jhydrol.2021.126378.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50185-8

The biorefinery concept for the industrial valorization of pineapple leaves co-producing ethanol, citric acid, and xanthan gum: a techno-economic analysis

Juan Murcia[a,b*], Rolando Barrera[a], Alba Ardila[b], and Edwin Zondervan[c]

[a]*Department of Chemical Engineering, University of Antioquia, 67-53108, Medellin 050010, Colombia*
[b]*Faculty of Basic Social and Human Sciences, Jaime Isaza Cadavid Colombian Polytechnic, 48-7151, Medellin 4932, Colombia*
[c] *Faculty of Science and Technology, University of Twente, Drienerlolaan 5, Enschede 7522, The Netherlands*
fernando.murcia@udea.edu.co

Abstract

Processing pineapple field residues (PR), a significant waste from the pineapple industry, into multiple value-added products based on the biorefinery concept can reduce waste disposal and contribute to sustainable development goals. Although PR has been used to produce several added-value products, the production of ethanol, citric acid, and xanthan gum, with its economic feasibility at an industrial scale, has not been investigated so far. The present study evaluates the viability of a biorefinery process of PR by performing a comparative techno-economic analysis of three processing scenarios: (1) a whole biorefinery process that fully utilizes PR biomass and produces ethanol, citric acid, and xanthan gum (E+CA+XG) (2) a process that produces CA and XG, and (3) a process that produces only E. A plant capacity of 4 metric tons/hour was considered in the analysis to mimic an intensive pineapple plantation size. Each processing scenario was modeled using SuperPro Designer, and the economic performance was evaluated based on the internal rate of return (IRR), net present value (NPV), and payback period. Among the three scenarios, the whole biorefinery process (CA+XG+E) showed the highest techno-economic performance with the net present value (NPV), internal rate of return (IRR), and payback period of 122.3 million US-$, 34.8%, and 3.5 years, respectively, due to the diverse revenues and minimized waste disposal cost. On the other hand, the E plant showed the lowest economic performance with a negative NPV.

Keywords: Pineapple biorefinery; Modeling; Simulation; Techno-economic analysis.

1. Introduction

Globally nearly 76 million tons of pineapple field residues (PR) are generated every year and end as waste (Satyanarayana et al., 2007). Therefore, the vast number of PR calls for cost-effective and environmentally friendly management options, such as transforming them into multiple value-added products based on the concept of biorefinery (Elbersen, and Hengsdijk, 2019; Russland et al., 2017; Zhang et al., 2011). A growing body of research evidence that PR is full of valuable molecules such as cellulose, hemicellulose,

and soluble mono-sugars (glucose and xylose), amino acids, vitamins, and minerals, which are a potential feedstock for microbial conversion to produce fuels (i. e., alcohol) and value-added products (i.e., organic acids and hydrocolloids). Although PR has been used to produce single-cell protein, ethanol, energy (Chen et al., 2020), and substrates of fiber, among other end-products under the broad concept of biorefinery (Roda and Lambri, 2019; Banerjee et al., 2019), the production of ethanol, citric acid, and xanthan gum, with its economic feasibility at an industrial scale have not been investigated.

A biorefinery concept that integrates processes and technologies to produce multiple value-added products from pineapple stubble have the potential to manage this residual biomass due to the following advantages: 1) synergistic effects among processes for the reduced energy consumption and labor demands; 2) minimal waste generation because the waste discarded by one process could be the input for another process; 3) diverse revenues due to the generation of multiple products (Jin et al.,2018b). The techno-economic assessment of biorefinery concepts has been used to evaluate the bioprocess viability of lignocellulosic biomass conversion into several products such as ethanol, xylitol, furfural, and others (Giuliano et al., 2018, Giuliano et al., 2014).

The present study aims to evaluate the economics of a biorefinery process of PR by performing a comparative techno-economic analysis of three processing scenarios: (1) a whole biorefinery process that fully utilizes PR biomass and produces ethanol, citric acid, and xanthan gum (E + CA + XG), (2) a process that produces citric acid and xanthan gum (CA + XG), and (3) a process that produces only ethanol (E). Each processing scenario was modeled in detail using SuperPro Designer. The economic performance was evaluated based on the internal rate of return (IRR), net present value (NPV), and payback period. Results from this study will provide information to the pineapple industry to help direct their residual biomass management.

2. Biorefinery process description

Approximately 250 metric tons of fresh pineapple plant residues (wet basis, mainly leaves) are removed every two years per hectare (Elbersen and Hengsdijk, 2019). Therefore, the first step is collecting pineapple leaves from different plantations and transporting them to the processing plant. After receiving, pineapple leaves are transferred by a conveyor to a crusher, where the juice is extracted. After the extraction, two streams of juice (192.5 tons) and fibrous material (45 tons) were obtained (Chen et al., 2020). Although pineapple leaves bagasse, the solid fraction obtained from the juice extraction, can be further processed into its main components or sent directly to a thermal valorization route, these options were beyond the scope of this study.

The composition of pineapple stubble used in the simulations and reactor data assumed from the literature on citric acid, xanthan gum, and ethanol fermentation are presented in Table 1. A plant capacity of 4 metric tons/hour was considered in the analysis to mimic an intensive pineapple plantation size. The plant has an operation period of 330 days (11 months) per year and 24 h per day, corresponding to a processing capacity of 31,680 metric tons (equivalent to 35,000 tons) of pineapple field residues per year.

The process simulation of the biorefinery was carried out by implementing a process flowsheet in SuperPro Designer v12. Figure 1 reports the general simplified biorefinery

Block Flow Diagram (BFD), including possible process pathways. Each block corresponds to a SuperPro Designer flowsheet containing a specific process section of the plant.

The designed biorefinery process generates three significant products: ethanol, citric acid, and xanthan gum. The process block diagram, flow diagram, and critical parameters applied in the process are shown in Figures 1 and 2 and Table 1. The biorefinery process can fully utilize pineapple leaves to produce multiple products for maximal revenue. However, the large number of unit operations involved in producing various products increases the capital and operating costs, which may negatively affect economic performance. Therefore, the biorefinery process was compared with another two processes where only citric acid and xanthan gum were produced from pineapple leaves to understand economic competitiveness better. Thus, three processing scenarios were considered in this study (1) a whole biorefinery process that fully utilizes PR biomass and produces ethanol, citric acid, and xanthan gum (E + CA + XG), (2) a process that produces citric acid and xanthan gum (CA + XG), and (3) a process that produces only ethanol (E).

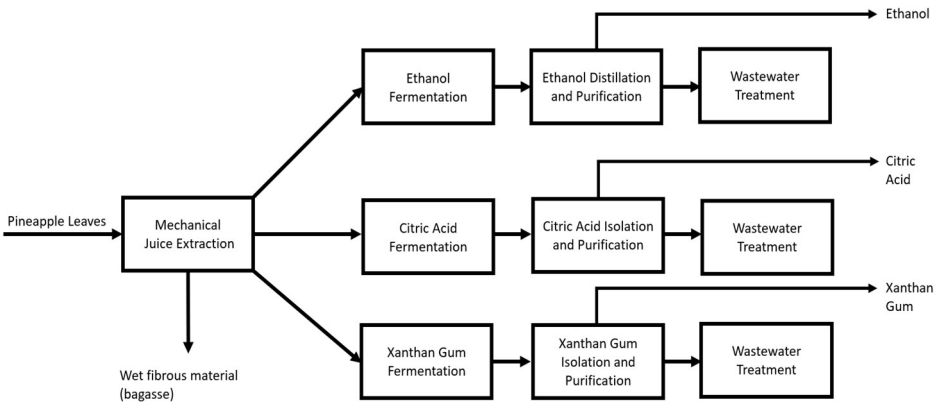


Figure 1 General Flowsheet of the multi-product pineapple biorefinery

2.1 Ethanol Production Steps

Glucose and xylose in the pineapple leaves juice to produce bioethanol with the *Zymomonas mobilis* bacteria, appropriate microorganisms for pentose and hexose fermentation. The fermentation process is carried out at 32°C for 48h in a bioreactor inoculated with 2% Z. mobilis (Aden and Foust, 2009). Following centrifugation of the fermentation medium, the produced bioethanol is distilled in a two-stage process using conventional distillation columns.

2.2 Xanthan Gum Production Steps

The fermentation process is carried out at the conditions of 0.01 vvm oxygen for 48 h and a temperature of 32 °C by inoculating the bioreactor with 2% *Xanthomonas campestris*. In downstream processing, the bacterial cells are inactivated by pasteurization and removed from the system in the solid phase by centrifugation. In the liquid phase, XG is obtained in solid form by precipitation with ethanol and separated by centrifugation. The solid XG is dried with a drum dryer, ground, and packaged. Finally, the liquid phase is distilled to recover and recycle the ethanol (Saydam et al., 2020).

2.3 Citric Acid Production Steps

The microorganism used to carry out the aerobic transformation of glucose and xylose into citric acid is *Aspergillus niger*. The fermentation is carried out under the following parameters: aeration rate of 1.25 vvm, stirring speed of 250 rpm, pH of 4.0, and a temperature of 30º C. When the fermentation is completed, the citric acid present in the fermentation product liquor is separated by two reaction-filtration stages, first with sulfuric acid and later with calcium hydroxide, to be purified in a crystallization stage. Finally, the separated citric acid crystals are dehydrated in a rotary dryer, obtaining the final product (granulated citric acid) with ~99% purity (Heinzle et al., 2006).

Table 1. Pineapple stubble data, conversion data of process simulations

Major components in the pineapple leaves, juice, and fibrous material	***Whole leaf***	***Juice***	***Wet fibrous material***
%Total solid (TS)	13.8	6.2	51.6
Cellulose (%TS)	22.6	---	36.8
Hemicellulose (%TS)	26.1	---	28.1
Lignin (%TS)	7.3	---	5.1
Glucose (g/L)	---	10.6	---
Xylose (g/L)	---	16.4	---
Crude Protein (%TS)	6.9	14	5.7
Crude Fat(%TS)	3.0	3.5	4.0
Ash (%TS)	6.1	10	1.7
Potassium (%TS)	2.6	3.8	0.6
Total carbohydrates (%TS)	---	72.5	---
Reactor's data	***Conversion (%)***	***residence time (h)***	
Xylose to ethanol	70	18	
Glucose to ethanol	93	18	
Xylose to citric acid	80	160	
Glucose to citric acid	85	160	
Xylose to xanthan gum	55	80	
Glucose to xanthan gum	60	80	

3. Economic Analysis

Economic analysis of the several cases was performed by estimating capital and operating costs. The purchased equipment, together with equipment installation, instrumentation and controls, processing piping, electrical systems, buildings, yard improvements, and service facilities, makes up the total plant direct cost. The total fixed capital investment consists of whole plant direct and complete plant indirect costs. With working capital and land, the fixed capital investment makes up the total capital investment (Peters et al., 2003).

Capital costs were estimated by power-law correlations based on unit capacity. In addition, data relevant to biorefinery sections were taken from the literature. Finally, the factorial method was applied to obtain the total investment cost. Raw materials mainly give operating costs. Therefore, a 40 US$/ton purchase cost for pineapple stubble was assumed.

Moreover, the prices of microorganisms, chemicals, and water required for pretreatment were also accounted for. In addition, the costs of salaries, maintenance, and insurance were assessed by the factorial method. The summary of the parameters used to carry out each process's technoeconomic assessment is presented in Table 2.

Revenues consist of products sales (ethanol, citric acid, and xanthan gum). Linear depreciation was assumed. A discounted cash flow analysis was carried out to obtain ethanol's Payback Selling Price (PSP) provided the price of the other possible by-products. The PSP was calculated as the value that makes the Present Net Value equal to zero after 20 years of plant life.

Table 2. Parameters for the techno-economic assessments

Economic data	
Plant life (y)	20
Ethanol Price (USD/L)	1
Citric Acid Price (USD/kg)	2
Xanthan gum Price (USD/kg)	9
Biomass Price (USD/ton)	40
Discount Rate (%)	7
Electricity cost (USD/MWhe)	100

4. Results

Based on the results of total capital costs, total operating costs, and revenues, the economic performances of the three plants were evaluated using NPV, IRR, and payback period as the indicators. Although the E + CA + XG plant had the highest total capital cost, it provided multiple products to generate revenue. From Table 3, we can see that the IRR for the E + CA + XG plant was 34.8%, with an NPV of 122.3 million US-$ and a payback period of 3.5 years. Although the biorefinery process of PR is new to the industry at the current stage, the products (ethanol, citric acid, and xanthan gum) from the process already have an established market. The positive NPV, the 34.8% of IRR, and the payback period of 3.5 years obtained from the techno-economic analysis suggested an economically feasible investment for this biorefinery process. Compared with the E + CA + XG plant, the CA + XG plant had lower revenue with the NPV, IRR, and payback period of 54.3 million US-$, 23.5%, and 4.3 years, respectively. The E plant produced ethanol as the sole product and had the worst economic performance among the three plants, showing a negative NPV of −15.6 million US-$, the lowest IRR of −2.1%, and the most extended payback period of 18.0 years, which was considered as economically unfavorable.

Table 3 Economic indicators (NPV, IRR, and payback period)

Item	**E+CA+XG**	**CA+XG**	**E**
NPV (million US-$)	122.3	54.3	-15.6
IRR (%)	34.8	23.5	-2.1
Payback period (year)	3.5	4.3	18

5. Conclusions

A techno-economic analysis was carried out on a biorefinery processing pineapple leaves to obtain ethanol, citric acid, and xanthan gum. The techno-economic model of pineapple stubble biorefinery process was developed with three scenarios: a single-product scenario (ethanol, E), a dual-product scenario (Citric Acid and Xanthan Gum, CA + XG), and an

integrated biorefinery scenario (Ethanol, Citric Acid, and Xanthan Gum, E + CA + XG). Among the three scenarios, the E + CA + XG scenario showed the best economic performance with the NPV, IRR, and payback period of 122.3million US-$, 34.8%, and 3.5 years, respectively, which was considered economically viable. Overall, this study demonstrated that the biorefinery of GP to produce multiple products is technically and financially feasible at a commercial scale. However, we are aware that this result might be significantly affected by the plant capacity, products price, and juice sugars content and deserves a sensitivity analysis, which will be addressed in future works.

References

A. Aden, & T. Foust, 2009, Techno-economic analysis of the dilute sulfuric acid and enzymatic hydrolysis process for the conversion of corn stover to ethanol. Cellulose, 16(4), 535-545.

S. Banerjee, A. Patti, V. Ranganathan, & Arora, 2019, A. Hemicellulose based biorefinery from pineapple peel waste: Xylan extraction and its conversion into xylooligosaccharides. Food and Bioproducts Processing, 117, 38-50.

A. Chen, Y. Guan, J. Bustamante, M. Uribe, L. Uribe-Lorío, M. Roos, & Y. Liu, 2020, production of renewable fuel and value-added bioproducts using pineapple leaves in Costa Rica. Biomass and Bioenergy, 141, 105675.

H. Elbersen, & H. Hengsdijk, 2019, Costa Rica pineapple field residue valorisation. In 27th European Biomass Conference and Exhibition, EUBCE 2019.

A. Giuliano, M. Poletto, & D. Barletta, 2015. Process design of a multi-product lignocellulosic biorefinery. In Computer Aided Chemical Engineering (Vol. 37, pp. 1313-1318). Elsevier.

A. Giuliano, D. Barletta, I. De Bari, & M. Poletto, 2018. Techno-economic assessment of a lignocellulosic biorefinery co-producing ethanol and xylitol or furfural. In Computer Aided Chemical Engineering (Vol. 43, pp. 585-590). Elsevier.

E. Heinzle, A. Biwer, & C. Cooney. 2006. Citric Acid-Alternative Process using Starch. Development of Sustainable Bioprocesses: Modeling and Assessment, 125-160.

Q. Jin, S. O'Keefe, A. Stewart, A. Neilson, Y. Kim, & H. Huang, 2021, Techno-economic analysis of a grape pomace biorefinery: Production of seed oil, polyphenols, and biochar. Food and Bioproducts Processing, 127, 139-151.

M. Peters, K. Timmerhaus & R. West. 2003. Plant design and economics for chemical engineers (Vol. 4). New York: McGraw-Hill.

A. Roda, & M. Lambri, 2019, Food uses of pineapple waste and by-products: a review. International Journal of Food Science & Technology, 54(4), 1009-1017 (2019).

N. Ruslan, N. Aris, N. Othman, A. Saili, M. Muhamad, & N. Aziz, 2017. A preliminary study on sustainable management of pineapple waste: the perspective of smallholders. Int J Acad Res Bus Soc Sci, 7(6):1-7.

K. Satyanarayana, J. Guimarães, & F. Wypych, 2007. Studies on lignocellulosic fibers of Brazil. Part I: Source, production, morphology, properties, and applications. Compos Part A Appl Sci Manuf, 38(7):1694-1709 (2007).

D. Saydam, A. Koulouris, & A. Dalgıç. 2020. Process simulation of integrated xanthan gum and sorbitol bioproduction: economic and sensitivity analysis with Taguchi approach. Process Integration and Optimization for Sustainability, 4(3), 279-295.

H. Zhang, D. Hu, J. Chen, X. Ye, S. Wang, J. Hao, & Z. An, 2011. Particle size distribution and polycyclic aromatic hydrocarbons emissions from agricultural crop residue burning. Environ Sci Technol, 45(13):5477-82.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50186-X

First Principles Based Development of Hybrid Models of Distillation Towers

Carlos Rodriguez[a], Prashant Mhaskar[a], Vladimir Mahalec[a]*

[a] *Department of Chemical Engineering, McMaster University, Hamilton, ON, L8S 4L7, Canada*
mahalec@mcmaster.ca

Abstract

In this work, we developed a simultaneous hybrid model with a novel error correction model for distillation columns that can be used in process monitoring, scheduling, and optimization. The proposed hybrid model consists of two parts a simplified first principles model (SFPM) and an error correction term. The SFPM is specific for distillation columns and consists of mass, energy, and equilibrium equations for all the trays of the column. For the error correction term, we used a traditional residual model that predicts the differences between the flow outputs of the SPFM and plant flow outputs, and we introduced a novel multiplicative correction model (MCM). The MCM corrects the output mass fractions predictions of the SFPM using a multiplicative factor. The advantage of using the MCM is that the predictions of the hybrid model are physically meaningful and more accurate than the residuals model when inputs are extrapolated beyond the training region. The proposed hybrid multiplicative model has a mean relative percentage error of less than 2% compared to rigorous simulations for interpolation and extrapolation inputs.

Keywords: Hybrid modeling; Distillation column; Process optimization; Process monitoring; Error correction models

1. Background

Optimal operation of refining and petrochemical plants requires that distillation towers meet product quality targets while minimizing energy consumption, which requires that accurate models of distillation towers be included in the plant-wide optimization models. Over the last decade, there has been a growing realization that rigorous plant models used in real-time optimization (RTO) lead to large-scale and challenging to maintain models. In addition, due to the model complexity, it is not possible to model an entire refinery via rigorous models and optimize such a model. In addition, if the distillation tower efficiency changes over time, the parameters of the rigorous models need to be adjusted by an expert. All of this indicates a need for simpler and yet accurate distillation models that can be used for inferential monitoring of plant operation and optimization.

Early work on approximate models of distillation towers employed partial least squares (PLS) models to predict product compositions from tray temperatures (Mejdel et al., 1991). Kano et al. (2000) introduced a dynamic PLS model which employed tray temperatures and manipulated variables. Even though these approaches lead to reasonable predictions in the region where training data is available, they do not extrapolate well if the tower moves to a region of increasingly nonlinear behavior.

Hybrid models of distillation tower employ first-principles mass and energy balances and approximate models of separation (Mahalec and Sanchez, 2012). In this work, we developed a simultaneous hybrid model for distillation columns consisting of a simplified

first principles model and an error correction model. The hybrid model can be used for process monitoring or optimization applications. We introduce a multiplicative correction term to demonstrate that it performs better than the traditionally used additive error correction term.

2. Case study

A butane splitter was selected as the study case for this work. A butane splitter is a standard unit in natural gas processing plants. It is the last distillation column of a separation train of natural gas liquid streams. The function of the butane splitter is to separate a binary mixture of isobutane iC4 and normal butane nC4. The relative volatility between these two components is close to one. Thereby, a distillation column with a high number of trays is required to achieve a successful separation.
Aspen plus steady-state simulations were used as a substitute for actual plant data. Due to the distillation column's components and pressure, the Peng-Robinson was selected as the property package in the simulations. The tower specifications are presented in Table 1 based on the work of (Aljuhani, 2016). The feed pressure, flow rate, and temperature were kept constant in the simulation, whereas the composition of the feed and product purities are varied in the simulation. Two datasets were created for the same butane splitter an interpolation and an extrapolation dataset, the difference being the feed mass fraction of iC4. The interpolation range of the feed mass fraction of iC4 is 0.20-0.35, whereas the extrapolation values of the feed mass fraction of iC4 are 0.15 and 0.4.

3. Modeling approach

Traditionally, series hybrid models have been used for modeling chemical units. We present a simultaneous hybrid model that combines the series approach (i.e., estimation of parameters of the model) with the parallel approach (i.e., correction model), as shown in Figure 1. The reason for combining both approaches is to obtain a hybrid model that is easy to train, update and compute while leveraging the knowledge and structure of fundamental equations and that can be updated for different tower conditions. The model's outputs are the mass fraction and mass flow for each of the components in the distillation tower.

Table 1 Butane splitter specifications

Tower conditions		Tower design		Operation range	
Feed pressure	12.1 bar	Number of trays	101	Feed $x_{F,iC4}$	0.15-0.40
Pressure drop	1 bar	Feed tray	50	Distillate $x_{D,iC4}$	0.96-0.99
Temperature	44 °C	Murphree efficiency	100 %	Bottoms $x_{B,nC4}$	0.96-0.99

3.1. Simplified first principles model

Distillation rigorous models consist of mass, equilibrium, summation, and heat (MESH equations). However, one of the main challenges of solving these equations is the non-linearity of the equilibrium equations (Biegler et al., 1997). Therefore, we introduce parameters estimated using data-driven models to reduce the complexity of the MESH equations while keeping the structure of mass, energy, and equilibrium equations. The SFPM can be applied to all distillation columns with the following characteristics: one feed, no sides streams, total condenser, and partial reboiler. The SFPM consists of two different matrices of linear equations. In the first matrix, the overall mass balances

equations Eq.(1) and energy equations Eq.(2) for each of the trays alongside the reflux ratio $RR = L_1/D$ and the specific reboiler duty $q_{reb} = Q_{reb}/F$ are used to find the total liquid and vapor flow in each of the distillation column trays. Where L_i denotes the total liquid flow in tray i numbered from top to bottom, V_i total vapor flow, H_i the liquid enthalpy, and the h_i vapor enthalpy. The temperature T of each tray is estimated using a data-driven model. It is assumed that it is possible to linearly approximate the enthalpies around one base operation as shown in and Eq.(3) and Eq.(4). Where the $^\circ$ denotes standard operation and C_p is the heat capacity.

$$F_i + V_{i+1} + L_{i-1} - L_i - V_i = 0 \tag{1}$$

$$\boldsymbol{F_i H_i + V_{i+1} h_{i+1} + L_{i-1} H_{i-1} - L_i H_i - V_i h_i = 0} \tag{2}$$

$$H_i = H_i^\circ + Cp_{i,liq}^\circ (T_i - T_i^\circ) \tag{3}$$

$$h_i = h_i^\circ + Cp_{i,vap}^\circ (T_i - T_i^\circ) \tag{4}$$

The second matrix consists of the equilibrium equations, the mass balances for each component, the reflux ratio for each of the components, and the summation equations (e.g., the sum of the flow of each of the components in the liquid has to be equal to the total liquid flow). This matrix is used to solve the mass flows of each of the components. The equilibrium equations are presented in Eq.(5) where the j subindex is the number of components and β is the equilibrium factor predicted with the empirical model. The mass balance equation for each of the components is presented in equation Eq.(6).

$$V_{i,j} = \frac{V_i}{L_i} \beta_{i,j} L_{i,j} \tag{5}$$

$$F_{i,j} + V_{i+1,j} + L_{i-1,j} = L_{i,j} + V_{i,j} \tag{6}$$

3.2. Error correction models

Parallel hybrid models have been used in the literature to correct the predictions of FPM. In this work, in addition to modeling the residuals (i.e., the difference between the actual flows for each of the components and the predictions of the SFPM), we present a multiplicative correction model.

The problem with using traditional residual models is that they don't have restrictions on feasible predictions. Therefore, it is possible in certain instances to obtain negative flows for some of the components that are in a lower proportion in the distillation column. For that reason, we introduced the multiplicative correction model (MCM). The objective of MCM is to predict the ratio between the actual mass fraction and the prediction mass fraction of the SFPM of each of the components.

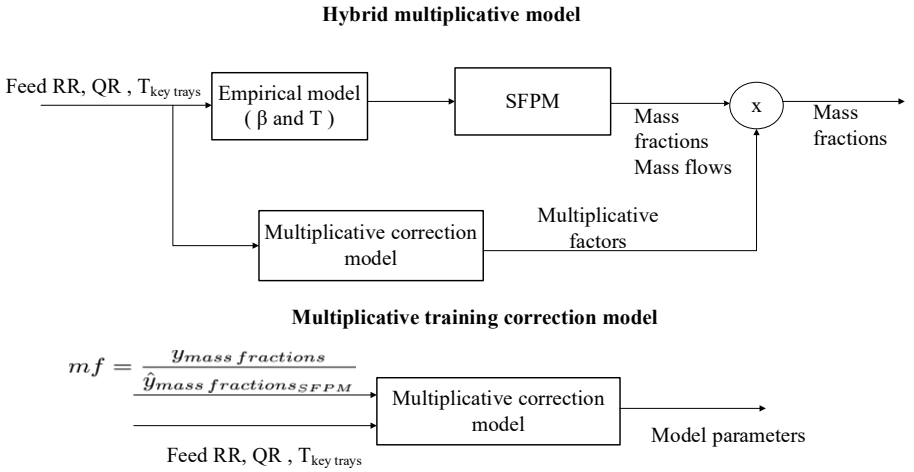


Figure 1 Hybrid multiplicative model structure for predicting and for training the multiplicative correction model

3.3. Training data-driven models

The temperature and equilibrium factor for each of the trays are estimated with empirical models. Neural networks (NN) and partial least squares (PLS) were used to predict those parameters. Although NN models are nonlinear mappings that can approximate any continuous function, they have several drawbacks, including the large amount of data to train them, local minima, and the overfitting problem. Therefore, linear models are often preferred as empirical models. The empirical models were trained using two different sets of inputs. The first set of inputs consists of the operating variables (reflux ratio and reboiler duty). The second set of inputs extends the first set with the addition of temperature measurements. The included temperatures are the temperatures next to the reboiler and condenser and the temperatures of the key trays. The key trays were determined using sensitivity analysis (Aljuhani, 2016). The models were trained using 80% of the interpolation dataset. The number of components for the PLS and the architecture of the NN was determined using the testing dataset, which is 10% of the interpolation dataset. The validation interpolation dataset determines the performance of the different hybrid models. The extrapolation performance of the models was evaluated using the extrapolation dataset. The metrics used to assess the models are the coefficient of determination R2, root mean squared error RMSE, mean relative percentage error MeanRE%, and the maximum relative percentage error MaxRE%.

4. Results and discussion

The first step of the hybrid model is the prediction of the parameters. The models for the parameters were trained using the set of inputs with temperatures and without temperatures of the key trays. The R2 of the prediction of the equilibrium factors evaluated using the validation dataset of the PLS models without the temperatures is 0.765 and 0.75 for the predictions of the temperatures. The R2 of the NN models without the temperatures of the key trays is 0.988 for the temperature and 0.984 for the equilibrium factors. The inclusion of the temperatures of the key trays in the inputs, increases the R2 of PLS models to 0.993 and 0.992 for the equilibrium factors and temperatures, respectively. Therefore, it is possible to estimate the temperature and

equilibrium factors of the trays only using linear models instead of nonlinear models. For that reason, PLS models were selected for the prediction of the parameters of the SFPM.

Table 2 Summary results predictions using different surrogate models for the mass fraction of nC4 in the distillate

	Interpolation Validation			Extrapolation		
	SFPM	Hybrid additive	Hybrid multiplicative	SFPM	Hybrid additive	Hybrid multiplicative
R2	0.998	0.9997	0.9996	0.988	0.9861	0.9984
RMSE	4.677e-4	1.915e-4	2.018e-4	1.095e-3	1.174e-3	3.956e-4
MeanRE%	1.69	0.694	0.694	3.5	4.91	1.41
MaxRE%	7.27	2.35	3.30	12.46	28.79	4.52

Table 3 Summary results predictions using different surrogate models for the mass fraction of iC4 in the bottoms

	Interpolation Validation			Extrapolation		
	SFPM	Hybrid additive	Hybrid multiplicative	SFPM	Hybrid additive	Hybrid multiplicative
R2	0.9962	0.9999	0.9987	0.9922	0.9996	0.9978
RMSE	6.132e-4	9.107e-5	3.558e-4	8.84e-4	2.08e-4	4.743e-4
MeanRE%	2.02	0.243	0.915	2.46	0.76	1.25
MaxRE%	5.36	0.84	2.37	11.74	5.26	9.16

A PLS model was developed for the additive correction model (i.e., correction of the mass flow of each of the components) and the multiplicative correction model. A summary of the results for the prediction of the mass fraction of the products of the butane splitter is presented in table 2 and table 3. For inputs within the training region (interpolation), both hybrid models perform better than only using the SFPM. However, when using the model outside of the training region, the predictions of the hybrid additive model are worse than the SFPM predictions for the nC4 in the distillate. In contrast, the hybrid multiplicative model corrects the predictions of the SFPM even for extrapolation inputs. Another advantage of hybrid multiplicative is that the predictions of concentrations can never become negative after the correction. The observed versus predicted plots for both components are shown in figure 2. We can observe from this figure that the differences between observed and predicted values are low for the validation interpolation inputs and the extrapolation inputs.

5. Conclusions

A novel simultaneous hybrid model for distillation columns is presented in this work. The hybrid model consists of a SFPM based on mass, approximate energy, and approximate equilibrium equations, and an error correction term. Separation factor and temperature profiles at the new operating state (which are predicted by using empirical models) are used in SFPM to predict product composition. The prediction of the SFPM is improved by using a multiplicative correction model. The presented hybrid model predicts product compositions with MeanRE% of less than 2% relative to the rigorous tower model, and it extrapolates well beyond its training region.

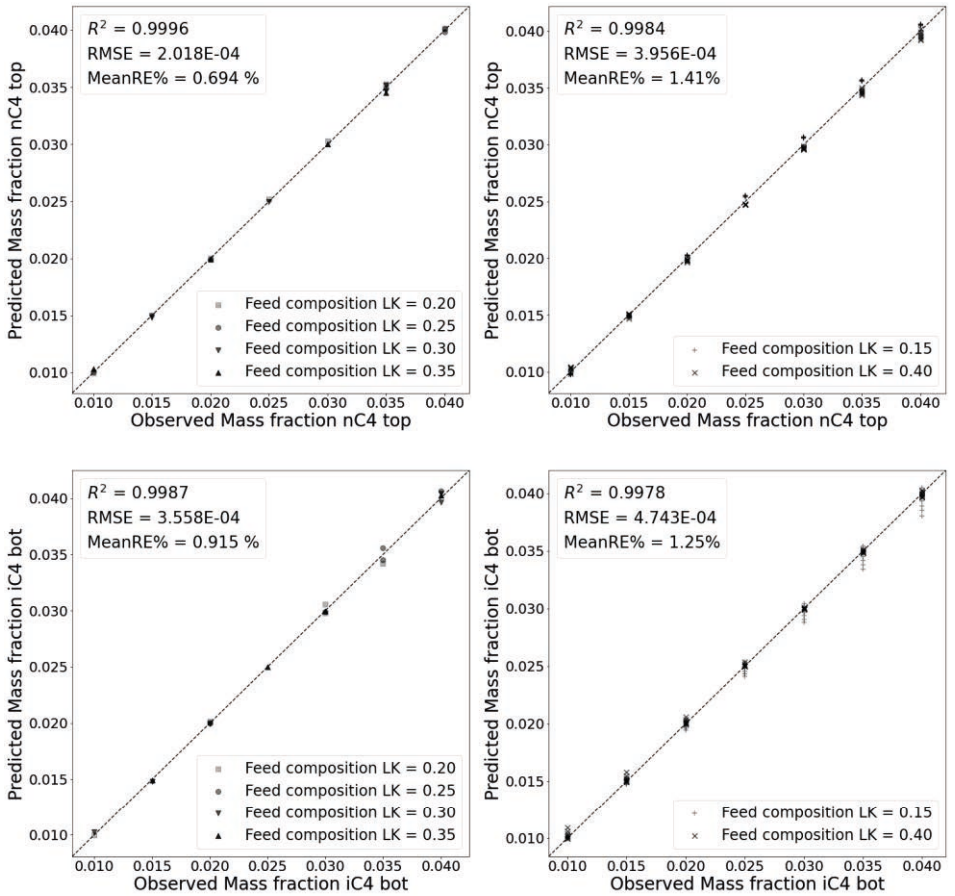


Figure 2 Observed versus predicted plots for the product composition of the distillates and bottoms. The results on the left represent the results of the validation dataset. The results on the right represent the results of the extrapolation dataset.

6. References

Aljuhani, F., 2016. Hybrid Model for Monitoring and Optimization of Distillation Columns ; McMaster University, M.A.Sc. Thesis

Biegler, L.T., Grossmann, I.E. and Westerberg, A.W., 1997. Systematic methods for chemical process design.

Kano, M., Miyazaki, K., Hasebe, S. and Hashimoto, I., 2000 Inferential control system of distillation compositions using dynamic partial least squares regression. J. Proc. Control, 10, 157-166

Mahalec, V. and Sanchez, Y., 2012. Inferential monitoring and optimization of crude separation units via hybrid models. Computers & chemical engineering, 45, pp.15-26.

Mejdell, T. and Skogestad, S. 1991 Estimation of distillation compositions from multiple temperature measurements using partial-least-squares regression. I&ECR, 30, pp. 2543-2555

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50187-1

Model-Based Development of Fuel Cell Stack and System Controllers

Shigeki Hasegawa[a*,b], Yusuke Miyamoto[b], Sanghong Kim[c], Yoshihiro Ikogi[b], Miho Kageyama[a], and Motoaki Kawase[a]

[a] *Department of Chemical Engineering, Kyoto University, Kyoto 615-8510, JAPAN*
[b] *Commercial ZEV Product Development Div., Toyota Motor Corporation, Aichi 471-8571, JAPAN*
[c] *Department of Applied Physics and Chemical Engineering, Tokyo University of Agriculture and Technology, Tokyo 184-8588, JAPAN*
s.hasegawa@cheme.kyoto-u.ac.jp

Abstract

Fuel cell (FC) system controllers for stack, air, H_2, and cooling systems are developed, and integrated with the system hardware models. These controllers allow flexible changes of system hardware specifications and operating conditions under a closed-loop simulation. The virtual FC-system hardware and controller design and system performance evaluation were demonstrated and confirmed to be beneficial for the actual product system development process.

Keywords: Fuel cell system; Model-based development; Physical modeling; Controller design

1. Introduction

The role of the FC-system controller is to ensure the optimal conditions in the FC stack by manipulating the system component considering the hardware limitations to achieve the maximum system performance such as the system efficiency and acceleration response. In many FC-system manufacturers, the controller design activities can be conducted only after the prototypes of system testbeds are manufactured. Since it takes a long lead-time as well as a great expense for the preparation of the system testbeds, the integrated FC-system simulator for the virtual controller development and evaluation has been strongly demanded. Though studies on control system are of such great importance, the fuel cell itself has been more intensively investigated (Weber et al., 2014), and much less research has been done on the FC stack, the FC system including the system components of air, H_2, and cooling systems, and the FC-system controllers. The authors have developed the integrated FC-system simulator with the physical models of the FC stack (Hasegawa et al., 2021a) and the system components of air, H_2, and cooling systems (Hasegawa et al., 2021b). This simulator can estimate the dynamic behavior of the entire FC system of 2nd-generation MIRAI, the latest commercial fuel cell electric vehicle (FCEV) shown in Fig. 1 (Takahashi et al., 2021), with the acceptable accuracy and allowable calculation time. The simulator can be utilized for the wide range of application system development of passenger vehicles, commercial vehicles of buses and tracks, railways, marine vessels, aviation, and stationary power generator purposes, though it consists of only hardware models of the FC stack and the FC-system components and the controllers are not included.

In this study, the entire FC-system controller with a streamlined configuration was developed, where the setpoints of the FC-system net power, the FC-stack state variables, and the actuation values of the FC-system components can be determined independently and it is remarkably easy to modify, replace, add, and remove the controller specifications. These controllers were integrated with the FC-system hardware model described above to obtain the closed-loop FC-system simulator. The simulator reduces the lead-time of the entire FC-system and controller development and the expenses for the prototypes of different applications systems. The effect of the coolant radiator size on the FC-system performance, the dynamic setpoint tracking performance of the FC-system net power and the FC-system net efficiency, were investigated by using the developed closed-loop simulator.

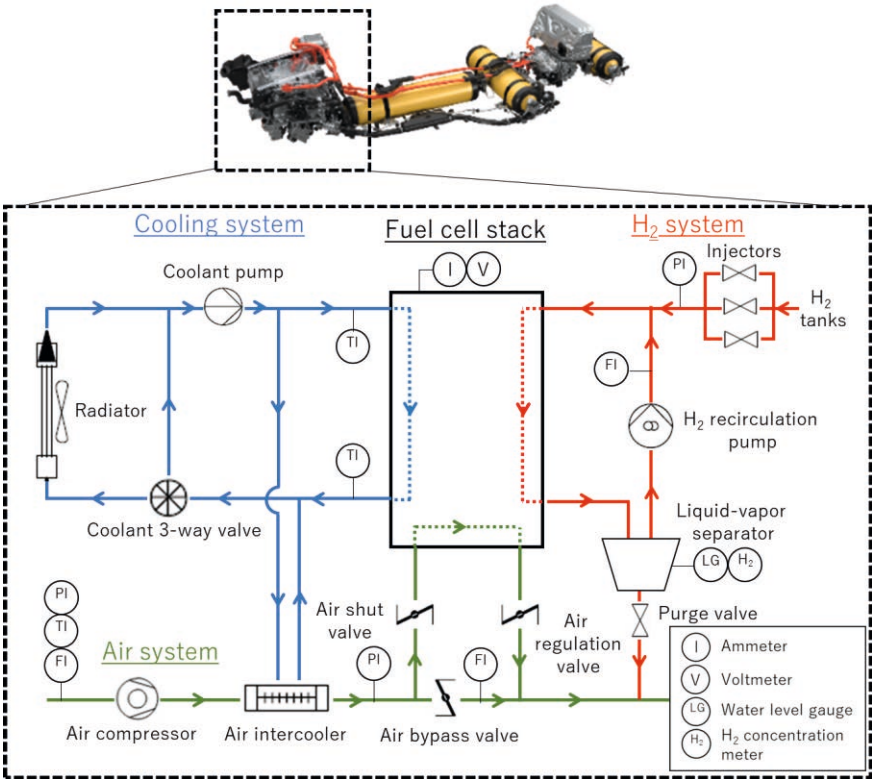


Fig. 1. Flow diagram of the FC system implemented in 2nd-generation MIRAI

In the FC-system applications, a setpoint of net power is given to the controller, and the setpoint tracking performance and high system net efficiency are the primary requirements.
The FC-system net power is defined by Eq. (1),

$$P_{\mathrm{FC}}^{\mathrm{net}} = P_{\mathrm{FC}}^{\mathrm{gross}} - (P_{\mathrm{ACP}} + P_{\mathrm{HP}} + P_{\mathrm{WP}} + P_{\mathrm{others}}) \tag{1}$$

where $P_{\mathrm{FC}}^{\mathrm{net}}$ is the FC-system net power [W], $P_{\mathrm{FC}}^{\mathrm{gross}}$ is the FC-stack gross power, P_{ACP}, P_{HP}, P_{WP} are the consumed power by the air compressor, hydrogen pump, and water

pump [W], and P_{others} is the consumed power by the other system components such as a DC–DC convertor [W]. The FC-system net efficiency is defined as Eq. (2)

$$\eta_{\text{FC}}^{\text{net}} = \frac{P_{\text{FC}}^{\text{net}}}{P_{\text{FC}}^{\text{gross}} + \dot{q}_{\text{FC}}} \tag{2}$$

where $\eta_{\text{FC}}^{\text{net}}$ is the FC-system net efficiency [-], and $\dot{q}_{\text{FC}}$ is the heat generation rate from the FC stack [W]. $P_{\text{FC}}^{\text{gross}}$ and $\dot{q}_{\text{FC}}$ are expressed by Eqs. (3) and (4),

$$P_{\text{FC}}^{\text{gross}} = I_{\text{FC}} V_{\text{FC}} \tag{3}$$

$$\dot{q}_{\text{FC}} = I_{\text{FC}} \left(V_{\text{FC}}^{\text{OC}} - V_{\text{FC}} \right) \tag{4}$$

where I_{FC}, V_{FC}, and $V_{\text{FC}}^{\text{OC}}$ are the FC-stack current [A], voltage [V], and open-circuit voltage [V]. They can be measured by the current and voltage sensors shown in Fig. 1.

2. FC-system controller and closed-loop simulation of the entire FC system

The configuration of the closed-loop FC-system simulator which consists of the FC-system controllers developed in this study and the FC-system hardware model is shown in Fig. 2. In this simulator, setpoint of the FC-system net power, the ambient temperature, and atmospheric pressure are inputs to the electric power controller. The calculated FC-system net power is output as well as the other calculated values in the FC system, the FC-stack polarization states such as current, voltage, and resistance, the distributions of state variables across the system such as pressure, flowrate, temperature, and gas composition, and the actuation values of the FC-system component such as pump speed and valve opening.

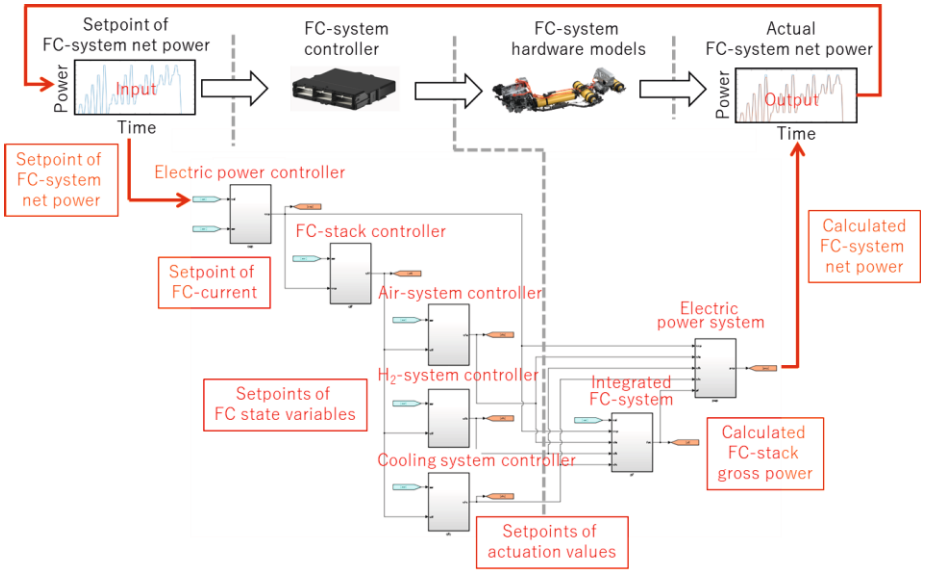


Fig. 2. Flow diagram of the closed-loop simulation of the entire FC system

The controllers in Fig. 2 consist of the electric power controller, the FC-stack controller, and the actuator controllers for the system components in air, H_2, and cooling systems. The electric power controller determines the setpoint of FC-current to minimize the error between the setpoint and calculated values of system net power. The FC-stack controller consists of air, H_2, and cooling state controllers. Each state controller determines the setpoints of the state variables such as pressure, flowrate, temperature, and gas composition at the inlet and outlet of the FC stack based on the FC-stack current and coolant outlet temperature measured by the sensors depicted with I and TI in Fig. 1 and pre-calibrated functions. The actuator controllers of air, H_2, and coolant determine the setpoints of actuation values such as the pump speed and valve opening to minimize the error between the setpoint of the state variables from the FC-stack controller and the actual value of state variables measured by the sensors shown in Fig. 1. Such a simple and streamlined controller configuration enables independent investigations of the optimal stack operating conditions and hardware specifications.

Fig. 3 is the algorithm of the electric power controller included in the Simulink flow diagram. Firstly, the setpoint of the FC-system net power is mediated with the upper limit, lower limit, minimum, and the gradient limitations determined by the system hardware limitations. Then the control error of the FC-system net power is compensated by determining the setpoint of the FC-stack current by a simple PI-control method including the supplemental functions of integral anti-windup and integral reset methods for the stable operation.

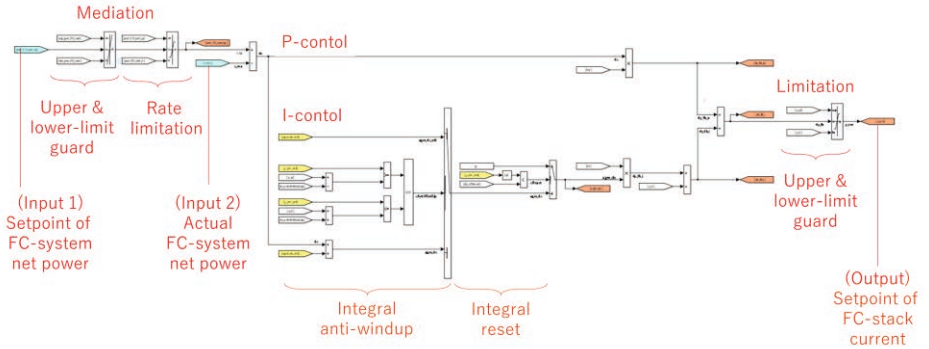


Fig. 3. Flow diagram of the electric power controller

Fig. 4 is the air pressure determination algorithm included in the Simulink flow diagram. The setpoints of the FC-stack current and temperature were given to the pre-determined functions, which were built based on experimental data to achieve the maximum FC-system net efficiency considering the electric power consumption and loss of air compressor. Then, the setpoint of air pressure is calculated. Pre-determined functions are used also for the determination of the setpoints of the other state variables. In the air, H_2, and cooling system controllers, the setpoints of the state variables outputted from the FC-stack controller are converted to the actuation values such as pump speed and valve opening by the same PI-control based algorithm in the electric power system controller shown in Fig. 3. By virtue of the closed loop simulator shown in Fig. 2, the pre-determined functions and the PI-controller parameters could be optimized and determined

without manufacturing the integrated FC-system hardware. This reduces the considerable effort and expense for the hardware preparation.

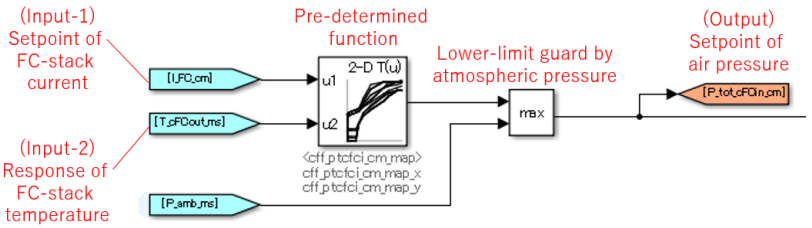


Fig. 4. Simulink flow diagram of air pressure setpoint determination in the FC-stack controller

3. Simulation results and discussion

Fig. 5 shows the results of the closed loop simulation shown in Fig. 2 to confirm the setpoint tracking performance of the FC-system net power when the vehicle speed is kept constant at 160 km/h and the coolant radiator size is varied. It was confirmed that the FC-system net power could trace the setpoint within an acceptable deviation less than 1 % in the case of the base-size radiator as shown in Fig. 5 (a). Fig. 5 (b) shows the 3 % deviation of the FC-system net power from the setpoint occurs during 660 – 670 s when the coolant radiator size is 70 % of the base size. In this period, the average coolant temperature is raised to 105 ℃ as shown in Fig. 5 (c), where IV performance of the FC stack deteriorated due to the severe dry condition as shown in Fig. 5 (d). From these results, it was suggested that the limitation of the radiator size reduction is around 85 % in the given operating condition.

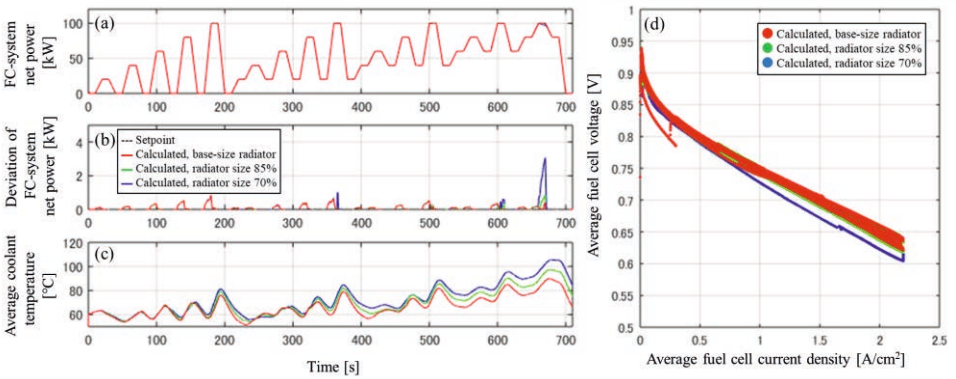


Fig. 5. Closed loop simulation results of the sensitivity of the coolant radiator size on (a) the setpoint and calculated value of the FC-system net power, (b) the deviation between the setpoint and calculated value of FC-net power, (c) the average coolant temperature, and (d) the FC-stack IV performance

Fig. 6 shows closed-loop simulation results of the FC-system efficiency described in Eqs. (1) and (2). The FC-system net efficiency was confirmed to be around 40–70 % in 20–90 kW with every radiator size, which was sufficiently high compared with conventional internal combustion engine systems.

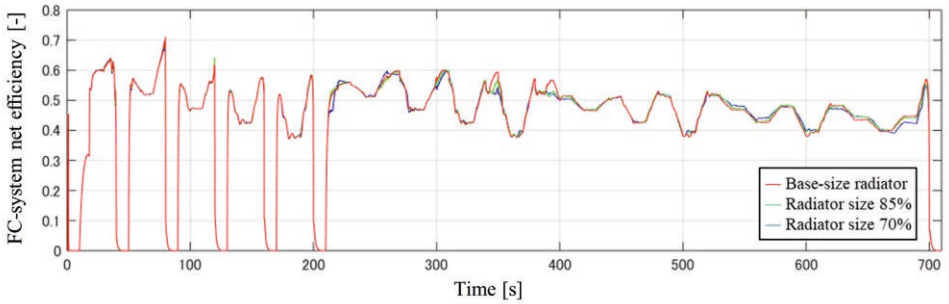


Fig. 6. Closed loop simulation outputs of the FC-system net efficiency

It was shown by the examples in Figs. 5 and 6 that the closed-loop simulation of the entire FC system in this study enabled the detailed system hardware and control investigation before manufacturing the costly FC-system testbed and test vehicles. On the other hand, control parameters, such as the gains in PI-control in Fig. 3 and the pre-determined function of target state variables in Fig. 4, were calibrated by the time-consuming trial-and-error and the maximum FC-system net efficiency cannot be ensured theoretically. The more efficient and sophisticated controller parameter determination process should be implemented by introducing the state-of-art optimization technologies such as model predictive control (MPC).

Conclusions

The controllers for the entire FC system were developed. These controllers were integrated with the FC-system hardware models to serve a closed-loop simulation. It enabled the design of the hardware and controllers in the entire FC system without manufacturing the costly system prototypes. The control parameters were calibrated on the developed simulator until the prospective controller performance was confirmed. The sensitivity of coolant radiator size on the setpoint tracking performance of the FC-system net power and the FC-system net efficiency were investigated with the calibrated controller as the demonstration of the model-based FC-system development.

Acknowledgement

This work was supported by the FC-Platform Program: Development of design-for-purpose numerical simulators for attaining long life and high performance project (FY 2020–2022) conducted by the New Energy and Industrial Technology Development Organization (NEDO), Japan.

References

A. Z. Weber et al., 2014, *Journal of The Electrochemical Society*, **161** (12) F1254-F1299

S. Hasegawa, M. Kimata, Y. Ikogi, M. Kageyama, S. Kim, and M. Kawase, 2021a, *ECS Transactions*, 104 (8) 3-26

S. Hasegawa, S. Matsumoto, Y. Ikogi, T. Takahashi, S. Kim, M. Kageyama, and M. Kawase, 2021b, *The 14th International Symposium on Process Systems Engineering* (*PSE 2021+*)

T. Takahashi and Y. Kakeno, 2021, *EVTeC 2021 Proceedings*, No. B1.1.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50188-3

A Nested Schur Decomposition Approach for Multiperiod Process Optimization

Noriyuki Yoshio[a], Lorenz T. Biegler[b*]

[a]*Kureha Corporation, 16, Ochiai, Nishiki-Machi,Iwaki, Fukushima 974-8686, Japan*
[b]*Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, USA*
biegler@cmu.edu

Abstract

This work develops an algorithm for solving multiperiod optimization (MPO) problem using a nested Schur decomposition (NSD) approach. The NSD approach decomposes MPO using a Schur complement and allows us to solve the decomposed nonlinear programming (NLP) problem in parallel. The NSD partitions the MPO into a two-level problem with individual NLPs at the lower level. The problem-level decomposition facilitates the flexible selection of the lower-level solver.

In this paper, the NSD approach is demonstrated with different process models for MPO in parallel computation. The solutions are also compared with the direct approach, which solves the entire MPO problem simultaneously. The demonstration shows IPOPT could be more efficient than CONOPT when the problem is well-conditioned. Moreover, it is noted that the NSD outperforms the direct approach when the size of the process model is large with CONOPT as the lower-level solver. From those results, we observe that NSD is well-suited to solve large MPO problems for chemical processes in an efficient, flexible, and robust manner.

Keywords: Optimization; Multiperiod; Applications.

1. Introduction

MPO problems are an important class of optimization problems that often consider design and operating plans for predicted demands over a given time horizon. For chemical processes, rigorous process models are highly integrated with multiple unit models and thermodynamic property models, which lead to a large system of nonlinear equations. MPO problems of such process models could be significantly larger and more complex than a single process model. The chemical industry highly demands strategies to extend a single process model to MPO in a robust, flexible, and efficient way over their process development.

To solve such a large-scale nonlinear problem in efficient and tractable ways, several decomposition approaches have been developed in the past. Those approaches can be classified into internal and external decompositions. The former internal decomposition approaches for an interior-point algorithm have been developed. The overview of the developments and the applicable problem structures can be seen in (Kang et al., 2015). The approach utilizes block-angular or general block-bordered structure in the solution of the augmented system for decomposition at the linear algebra level. Hence, the internal decomposition approach can fully utilize the advantage of interior-point methods in solving large-scale nonlinear problems. The methods, however, face difficulties when there are degeneracies in the problem constraints. In such case, we

often rely on active set methods such as CONOPT (Drud, 1985), which handles dependent constraints efficiently as part of the active set selection.

On the other hand, external decomposition approaches decompose the problem at the problem formulation level. The external approach could be more flexible and easier in the implementation than the internal one because the approach is less intrusive. For example, Benders decomposition and the alternating direction method of multipliers (ADMM) are one of the external decompositions. These methods utilize only the first order information of the decomposed problems. Thus, the methods are inefficient due to the lack of the exact Hessian information from the lower problem, which is fully utilized in the internal decomposition approach.

To address this problem, we develop the nested Schur decomposition (NSD) approach and demonstrate it with MPO problems for chemical processes. The approach applies an external decomposition so that we can flexibly utilize interior-point or active set solvers for the lower-level problem. On the other hand, as a Schur decomposition approach it has the same KKT structure and fast convergence properties which allows us to use the Hessian information of the lower-level problem as the internal Schur decomposition approach. Moreover, it retains the computational advantage of the parallelization in the external decomposition framework.

2. Problem statement

We consider the general MPO problem of the form

$$\begin{aligned} \min \quad & \bar{f}_0(x) + \sum_{i=1}^{N} \bar{f}_i(x, z_i) \\ \text{s.t.} \quad & h_0(x) = 0 \\ & h_i(z_i, d_i) = 0 \\ & g_i(z_i, d_i) + s_i = 0, \;\; s_i \geq 0 \\ & z^L \leq z_i \leq z^U, \;\; i = 1, \dots, N \\ & x^L \leq x \leq x^U \end{aligned} \tag{1}$$

where N is the number of periods, $x \in \mathbb{R}^{n_x}$ is the vector of global variables and $z_i \in \mathbb{R}^{n_{z,i}}$ are the decision variables in each period i. Some of the global variables are exposed to process constraints h_i and g_i of each period via $d_i \in \mathbb{R}^{n_{d,i}}$. The coupling (or complicating) variables d_i are also included in the global variable vector x. s_i are the slack variables for the inequality constraints g_i. The global constraints h_0 could be linking, common, and demand constraints, and the linking constraints use the information of adjacent periods to describe inventory updates. The common constraints enforce the same value for all periods such as design parameters. The demand constraints reflect the limitation of the sales for the demand with given demands or demand forecast models.

For the problem level decomposition, we reformulate Problem (1) as a two-level optimization problem by disaggregating the complicating variable d_i and introducing artificial variables δ_i and dummy constraints. The upper level problem is written as:

$$\begin{aligned} \min \quad & \Phi_0(x) \equiv f_0(x) + \sum_{i=1}^{N} f_i\big(z_i(x), \delta_i(x)\big) + \sum_{i=1}^{N} \beta q_i(x)^T e \\ \text{s.t.} \quad & h_0(x) = 0 \\ & h_i(z_i, d_i) = 0 \\ & x^L \le x \le x^U \end{aligned} \tag{2}$$

where $z_i(x), \delta_i(x)$ and $q_i(x)$ are solutions from each lower-level problem $i = 1, \dots, N$:

$$\begin{aligned} \min \quad & \Phi_i(x) \equiv f_i\big(z_i(x), \delta_i(x)\big) + \beta q_i(x)^T e \\ \text{s.t.} \quad & h_i(z_i, d_i) = 0 \\ & g_i(z_i, d_i) + s_i = q_{g,i}, \;\; s_i \ge 0 \\ & z^L \le z_i \le z^U, \;\; i = 1, \dots, N \\ & h_{d,i} = \delta_i - G_i x - C_{d,i}\big(q_{+,i} - q_{-,i}\big) = 0, \;\; q_i = \big[q_{+,i}^T, q_{-,i}^T, q_{g,i}^T\big]^T \ge 0 \end{aligned} \tag{3}$$

where $G_i \in \mathbb{R}^{n_{d,i} \times n_x}$ is a mapping matrix to assign the global variables x onto each period and $C_{d,i} \in \mathbb{R}^{n_{d,i} \times n_{d,i}}$ is a scaling matrix whose diagonal elements have scaling factors for δ_i. $q_{+,i} \in \mathbb{R}^{n_{d,i}}$ and $q_{-,i} \in \mathbb{R}^{n_{d,i}}$ represent positive and negative violation values for the relaxation of the dummy constraints. $q_{g,i} \in \mathbb{R}^{n_{g,i}}$ is also positive violation value for the relaxation of the inequality constraints. The values are penalized in the objective along with the penalty constant β and $e^T = [1,1,\dots,1]$. The purpose of the relaxation is to avoid an infeasible solution of the process model in each period. Furthermore, the original objective $\bar{f}$ is also modified to f so that the problem is separated into two parts.

3. Solution strategy

In order to collect the lower-level problems' information Φ_i into the upper-level problem, inequality constraints in the lower-level problems are replaced by barrier terms and the lower-level problems are assumed to be solved with interior point strategies. By considering the KKT conditions for each barrier problem of lower-level problem, the assembled Newton step is written as:

$$\left[\begin{array}{ccccc|ccccc|c} K_1 & & & & & E_1 & & & & & \\ & K_2 & & & & & E_2 & & & & \\ & & \ddots & & & & & \ddots & & & \\ & & & K_N & & & & & E_N & & \\ \hline E_1^T & & & & & & & & & & -G_1 \\ & E_2^T & & & & & & & & & -G_2 \\ & & \ddots & & & & & & & & \vdots \\ & & & E_N^T & & & & & & & -G_N \\ \hline & & & & & -G_1^T & -G_2^T & \cdots & -G_N^T & & 0 \end{array}\right] \left[\begin{array}{c} \Delta y_1 \\ \Delta y_2 \\ \vdots \\ \Delta y_N \\ \hline \Delta \gamma_1 \\ \Delta \gamma_2 \\ \vdots \\ \Delta \gamma_N \\ \hline \Delta x \end{array}\right] = -\left[\begin{array}{c} 0 \\ 0 \\ \vdots \\ 0 \\ \hline 0 \\ 0 \\ \vdots \\ 0 \\ \hline m \end{array}\right] \tag{4}$$

where $\gamma_i \in \mathbb{R}^{n_{d,i}}$ is the dual of the dummy constraints and the other primal and dual variables are represented $y_i \in \mathbb{R}^{n_y}$. $E_i = \nabla_{y_i} h_{d,i} \in \mathbb{R}^{n_{d,i} \times n_y}$. $K_i \in \mathbb{R}^{n_y \times n_y}$ is the KKT matrix where E_i is excluded. Furthermore, the right-hand side is set to zero except for m because we can solve problem (3) for each period individually. m is the gradient

information with respect to x and is written as $m = -\sum_{i=1}^{N} G_i^T \gamma_i$. By pivoting the block matrices in (4) we obtain:

$$\left[\begin{array}{cccc|cccc|c} K_1 & & & & E_1 & & & & \\ & K_2 & & & & E_2 & & & \\ & & \ddots & & & & \ddots & & \\ & & & K_N & & & & E_N & \\ \hline & & & & -P_1 & & & & -G_1 \\ & & & & & -P_2 & & & -G_2 \\ & & & & & & \ddots & & \vdots \\ & & & & & & & -P_N & -G_N \\ \hline & & & & -G_1^T & -G_2^T & \cdots & -G_N^T & M \end{array}\right] \left[\begin{array}{c} \Delta y_1 \\ \Delta y_2 \\ \vdots \\ \Delta y_N \\ \hline \Delta \gamma_1 \\ \Delta \gamma_2 \\ \vdots \\ \Delta \gamma_N \\ \hline \Delta x \end{array}\right] = - \left[\begin{array}{c} 0 \\ 0 \\ \vdots \\ 0 \\ \hline 0 \\ 0 \\ \vdots \\ 0 \\ \hline m \end{array}\right] \tag{5}$$

where $M = \sum_{i=1}^{N} G_i^T P_i^{-1} G_i$, $P_i = E^T K_i^{-1} E$. Note that $P_i^{-1} G_i$ can be generated directly by solving the following equation.

$$\begin{bmatrix} K_i & E_i \\ E_i^T & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} 0 \\ -G_i \end{bmatrix} \tag{6}$$

where $X_2 = P_i^{-1} G_i$ and $M = \sum_{i=1}^{N} G_i^T X_2$. M and m are the Hessian and gradient of the objective for the part of the lower-level problem. Thus, the augmented Hessian of the objective in the upper-level problem is $\nabla^2 \Phi_0(x) = \nabla^2 f_0(x) + M$. The gradient of the objective is $\nabla \Phi_0(x) = \nabla f(x) + m$.

When the lower-level problems' information augments the upper problem's objective, the upper-level problem becomes just a constrained nonlinear optimization problem. The upper problem can be solved with Newton-type solvers, an interior-point solver such as IPOPT, or a trust region solver such as *trust-constr* from Scipy by providing the augmented Hessian and gradient information. For the lower problem, each period is solved individually, as seen from the zero residuals on the right-hand side of Eq. (4). Although the approach is derived based on the interior point strategy with the barrier approach, any NLP solver such as IPOPT or CONOPT can be used, as long as the lower problem successfully converges and the KKT information can be retrieved. This is because the barrier parameter is equivalent to a Hessian projection upon convergence of the lower-level problem.

4. Implementation

As mentioned in Section 3, we can use any solver as long as the upper problem solver is a globally convergent Newton-type solver for constrained optimization. Here, we present the implementation scheme with *cyipopt* (Aides et al., 2021) as the solver of the upper problem, which is a python wrapper for IPOPT. The lower problems (3) are formulated in *Pyomo* (Hart et al., 2017), and solved with CONOPT or IPOPT. Then, the primal, dual values and the bound multipliers are extracted. The solution is utilized for the initial values of the next inner problem to use a warm start. After that, the part of the Hessian of the Lagrangian and the Jacobian of the constraints in the KKT system is evaluated with *Pynumero* (Laird et al., 2019). The Hessian of the bounded variables in the barrier term is separately evaluated in Python with the extracted primal and

multiplier values. The evaluation method follows the procedure of IPOPT (Wächter and Biegler, 2006).

5. Case study

In this section, we demonstrate the NSD approach with MPO problems for Williams-Otto (WO) process and benzene chlorination (BC) process. The upper level of the MPO problem consists of the inventory update and demand constraint which are set over the entire period. The lower level of the MPO problem consists of the process models for each time period to evaluate the production rate. The detailed problem formulation can be seen in (Yoshio and Biegler, 2021). Here, we focus on the computational performance. Different problem sizes are used from $N = 3$ to $N = 14$ periods and the problem size increases with N, as shown in Table 1 for the cases of both WO and BC processes. The WO process model is relatively small model. On the other hand, the BC process is considerably large model. In the performance test, the parallelized NSD approach uses N processes for solving each period in parallel. The computational time is compared with the direct solution that solves the entire MPO problem simultaneously without decomposition. For the lower-level solver, CONOPT is used for both the WO and BC processes, and IPOPT is used for the WO process. For comparison a direct solution of problem (1) is obtained with CONOPT.

The computational comparison can be seen in Figure 1. For the WO process, Figure 1(a) compares the computational time for the direct approach and NSD with both IPOPT and CONOPT as lower-level solvers. As seen in the figure, the direct approach is approximately 10-20 times faster than the NSD approaches. This is because the problem size of the WO process model is relatively small and the overhead of the NSD dominates any savings gained from the parallel solution of the lower-level problems. On the other hand, for parallel NSD, IPOPT outperforms CONOPT on the lower problems. This result indicates IPOPT for the lower problem solver could be more efficient than CONOPT for this well-conditioned problem. For the BC process, Figure 1(b) compares the computational time for the direct approach and NSD with CONOPT as a lower-level solver. The computational result shows that the computational time proportionally increases in the direct approach as N increases. For the parallel NSD approach, the wall clock time remains constant as the number of periods N increases. We note that parallel NSD outperforms the direct approach when N > 8.

Table 1 The number of variables in MPO problems

N	3	4	5	6	7	8	9	10	11	12	13	14
WO	149	198	247	296	345	394	443	492	541	590	639	688
BC ($\times 10^4$)	2.8	3.7	4.6	5.5	6.5	7.4	8.3	9.2	10	11	12	13

6. Conclusions

We have developed and demonstrated an NSD algorithm for solving large-scale nonlinear optimization problems, especially MPO problems. The NSD facilitates the parallelization of the decomposed problem within the algorithm by utilizing the structure of the KKT system. The approach was applied to MPO problems of the WO process and BC process. The studies show the capability of flexible solver choice for the inner problem. NSD is a reliable option to parallelize large-scale nonlinear problems because the flexible solver choice enables us to utilize robust active set solvers efficiently for degenerate problems.

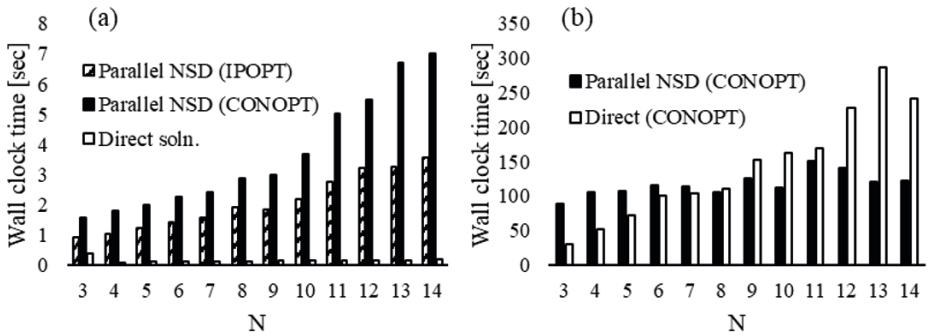


Figure 1 The comparison in the computational time for MPO of (a) WO process and (b) BC process.

Furthermore, the NSD approach has the computational advantage of the parallelization when the lower-level problem is large. This flexible, robust, and efficient approach could provide a number of benefits for current practical process development. In the future, NSD will be considered for a number of large-scale nonconvex NLP applications, including multiset and multi-effect parameter estimation problems, stochastic programming problems for process optimization under uncertainty, decomposition of integrated site-wide optimization problems with complex subsystems, and decomposition for optimization of spatially distributed energy networks.

References

A. Aides, A. Andreasen, M. Kümmerer, A. Toscano, W. Brendel, J.K. Moore, B. Bocklund, H. Wang, A. Pommel, A. De, S. Brockie, Y. Schälte, 2021. A Cython wrapper to the IPOPT optimization package. Technical Report. Accessed on Feb 6. https://pypi.org/project/ipopt/

A. Drud, 1985. Conopt: a grg code for large sparse dynamic nonlinear optimization problems. Math Program 31 (2), 153–191

W.E. Hart, C.D. Laird, J.-P. Watson, D.L. Woodruff, G.A. Hackebeil, B.L. Nicholson, J.D. Siirola, 2017. Pyomo–optimization modeling in python, vol. 67, 2nd Springer Science & Business Media . Hart, W.E. , Watson, J.-P. , Woodruff, D.L. , 2011. Pyomo: modeling and solving mathematical programs in python. Mathematical Programming Computation 3 (3), 219–260

J. Kang, N. Chiang, C.D. Laird, V.M. Zavala, 2015. Nonlinear programming strategies on high-performance computers. In: 2015 54th IEEE conference on decision and control (CDC). IEEE, pp. 4612–4620

C.D. Laird, B.L. Nicholson, J.S. Rodriguez, 2019. Parallel Numerical Algorithms for Structured Problems with PyNumero and Pyomo. Technical Report. Sandia National Lab.(SNL-NM), Albuquerque, NM (United States)

A. Wächter, L.T. Biegler, 2006. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Math Program 106 (1), 25–57

N. Yoshio, and L. T. Biegler. "A Nested Schur Decomposition Approach for Multiperiod Optimization of Chemical Processes." Computers & Chemical Engineering (2021): 107509

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50189-5

Design and Optimisation of Boil-off Gas Recycling Strategy in Liquefied Natural Gas Production

Ahmed AlNouss, Saad A. Al-Sobhi*

Department of Chemical Engineering, College of Engineering, Qatar University. Doha, Qatar
saad.al-sobhi@qu.edu.qa

Abstract

The process of liquefying natural gas is associated with many design and optimisation challenges. The formation of so-called boil-off gas (BOG) due to the difference in the bubble point of LNG and the surrounding ambient temperature around storage tanks, is one of the most challenging issues. Industries around the world are investigating the recovery of BOG to enhance the economy of design and reduce environmental impacts due to flaring. The BOG composition forms a crucial aspect of the recovery process where the nitrogen and methane content can affect the location of recovery in addition to the economy and mechanism of the recovery process. This study addresses the design and optimisation of the BOG recycling flowsheet configuration in the LNG production process. Multiple sensitivity analyses are considered to study the effect of changing temperature and mass flow of recycled BOG, in addition to the temperature and pressure of LNG feed. The observed parameters include final methane content, nitrogen content, mass flow and temperature of LNG product and fuel gas along with the Wobbe Index of LNG product. The BOG utilisation strategy focuses on recycling BOG prior the liquefaction unit with three different nitrogen rejection unit (NRU) configurations, namely, the double distillation columns, single distillation column and flash drum. To show the significance of the proposed recycling strategy, an illustrative case study is analysed and evaluated for recycling 52,000 metric tonnes of BOG annually while optimising the design conditions of the BOG and LNG streams. The results indicate the huge effect of BOG to natural gas blending ratio on the purity of the LNG product. Hence, the increase in BOG mass flow rate demonstrates to increase the methane content on the produced LNG and reduce the nitrogen content for the flash drum and double columns configurations. This highlights the potential of BOG recovery on enhancing the performance of the LNG process despite the need for an additional economic and environmental investigation.

Keywords: LNG, BOG, Flare management, Simulation, Optimisation, NRU.

1. Introduction

The global consumption of natural gas (NG) continuous to rise with the growing forecasted demand on energy and the increase in economic growth and world population. The low environmental impacts of NG relative to other fossil fuels place it as the favourable energy alternative. The liquefaction of NG in the well-known Liquefied Natural Gas (LNG) process dominates the energy market despite the intensive energy requirement. The LNG process is essential to ensure easier transportation and shipping of liquefied NG worldwide. The process comprise treatment of sour field NG to remove the presence of CO_2, H_2S and water, separation and liquefaction of sweet NG from the

associated heavier hydrocarbons, extraction of left-over nitrogen and helium components, and finally the storage, loading and transportation of LNG product. The volume of the LNG product constitute 1/600 of its gas-form volume. The common employed liquefaction technology in LNG process is C3-MR consisting of propane and mixed refrigerant cryogenic cooling loops. The A-PX technology has emerged as an improvement from C3-MR. LNG process is associated with a number of challenges; one of which is the formation of boil-off gas (BOG) during LNG storage loading and shipping operations. The large difference between LNG storage conditions (-160 ^{o}C) and ambient surrounding (~25^{o}C) in addition to sudden pressure changes and unloading operation act as driving force for the heat transfer despite advanced insulation mechanism and the partial evaporation of stored LNG producing BOG. Therefore, BOG consist mainly of methane (93wt%) and nitrogen (7wt%) and accounts for 1-3 vol% of LNG product. Venting and flaring of BOG is the common mitigation in global industries that is associated with environmental and economic concerns (Al-Sobhi et al., 2021).

Scholars have approached the aspects of BOG management from various perspectives including minimisation, recycling and quantification. For example, Kurle et al. (2015) explored different BOG recycling strategies in C3-MR process to establish optimum temperature for BOG minimisation and recycle. An optimal temperature of -166 ^{o}C found to reduce the total cost associated with the minimisation and recovery of BOG. Similarly Bao et al. (2019) compared three BOG management strategies from power output perspective while examining the effect of electricity cost, interest factor, and BOG content on the net present value (NPV). The results demonstrated an increasing trend for NPV with BOG content and a decreasing trend with interest rate. Shin et al. (2007) studied from a safety-driven objective the optimisation of BOG compression system to verify adequacy through a mixed-integer linear model. The results revealed an energy recovery potential of 11.5% from the current operating compressors.

Despite the high literature studies on BOG management, there is still a necessity to establish the optimum pathway of BOG recovery and utilisation within A-PX LNG process. This study explore the optimisation of key design parameters involved in the recovery and utilisation of BOG within different recycling configurations in the LNG production process. The liquefaction flowsheet model with recycled BOG stream is established to study the effect of changing temperature, pressure and mass flow of recycled BOG, in addition to the temperature and pressure of LNG feed on the final methane content, nitrogen content, mass flow and temperature of LNG product and fuel gas along with the Wobbe Index of LNG product. Traditional LNG plants practice the flare of BOG streams due to the intensive energy associated with liquefaction and recycling. However, flaring is not a sustainable mitigation due to the high associated NO_X and CO_2 emissions and the global burden to apply stringent environmental regulations. Hence, this study aims to establish an optimum pathway for the recovery of BOG.

2. Methodology

The overall methodology to design and optimise the BOG recycling strategy in LNG production line consist of 1) estimation of BOG rate of formation, 2) Simulation of LNG/BOG recycling scenarios, 3) execution of key sensitivity analyses, 4) formulation of optimisation problem, and 5) conclusion of optimal parameters. The BOG utilisation strategy focuses on recycling BOG prior the liquefaction section with three different nitrogen rejection unit (NRU) configurations, namely, the double distillation columns, single distillation column and flash drum. The base models have been emphasised in an

earlier work (Al-Sobhi et al., 2021) with the detailed description of process models and economic and environmental benefits of BOG recycling. This study forms an extension to optimise the BOG utilisation with the variation of key design parameters. To show the significance of the proposed recycling strategy, an illustrative case study representing the state of Qatar LNG plants is analysed and evaluated for various design alternatives of recycling 52,000 metric tonnes of BOG annually while optimising the conditions of the BOG and LNG streams. BOG is assumed to be generated from five 300,000 m^3 storage tanks at a vaporisation rate of 0.5vol% per day. Aspen HYSYS is utilised to construct the integrated LNG-BOG flowsheets of different schemes. Qatar is currently producing approximately 77 million metric tonnes per annum (MMTPA) with planned future expansion up to 110 MMTPA in 2025 placing Qatar as the largest global LNG producer and exporter (Qatargas, 2019). Multiple sensitivity analyses are considered to study the effect of changing temperature and mass flow of recycled BOG, in addition to temperature and pressure of LNG feed. The observed parameters the final methane content, nitrogen content, mass flow and temperature of LNG product and fuel gas along Wobbe Index (WI) of LNG product. WI is used to correlate the high heating value (HHV) of a gas mixture to the square root of the gas specific gravity; density of mixture relative to air, as expressed in Eq. (1). It is utilised in this study as a representative of the significance of BOG recycling in enhancing energy content.

$$WI = \frac{HHV}{\sqrt{\frac{\rho}{\rho_{air}}}} \tag{1}$$

Aspen HYSYS is used to assess the variations in each operating parameter. Results of the sensitivity analyses are used to construct regression models relating WI of LNG product and methane content of fuel gas to changes in each operating parameter. The functions of the WI and methane content of LNG product for each operating parameter are employed where the weighted average of these functions is calculated to characterise the objective functions of the proposed optimisation model. The singular objective function of WI and methane content of LNG product, expressed in Eqs. (1 and 2), is then maximised to produce the optimal decision variables for each objective. Later, optimal solutions for the simultaneous computing objectives of WI and methane content of LNG product are achieved using Matlab Genetic Algorithm tool and a Pareto front is generated for each recycling configurations.

Variables:

$\dot{T}_{fuel\ gas}$: Fuel gas final temperature (°C)
$\dot{T}_{LNG}$: LNG final temperature (°C)
$\dot{m}_{fuel\ gas}$: Fuel gas production rate (t/h)
$\dot{m}_{LNG}$: LNG production rate (t/h)
$\dot{m}_{Nitrogen,fuel\ gas}$: Nitrogen content in fuel gas (t/h)
$\dot{m}_{Nitrogen,LNG}$: Nitrogen content in LNG (t/h)
$\dot{m}_{Methane,fuel\ gas}$: Methane content in fuel gas (t/h)
$\dot{m}_{Methane,LNG}$: Methane content in LNG (t/h)
WI: Wobbe Index (WI) of LNG product (MJ/Nm^3)

Decision variables:

$\dot{T}_{BOG}$: BOG inlet temperature (°C)
$\dot{T}_{LNG}$: LNG feed temperature (°C)
$\dot{m}_{BOG}$: BOG inlet rate (t/h)
$\dot{P}_{BOG}$: BOG inlet pressure (bar)

$\dot{P}_{LNG}$: LNG feed pressure (bar)

Objective function:

Maximise: WI (2)

Maximise: $\dot{m}_{Methane,LNG}$ (3)

Constraints: *As illustrated in Table 1*

Table 1. Optimisation problem raw data

Parameter	Flow (t/h)	Variation (t/h)	T (°C)	Variation (°C)	P (bar)	Variation (bar)
NG Feed	890.4	-	-84	-100:-200	41	1:20
BOG Feed	5.94	0.5:20	-162	-100:-200	1	1:20

3. Results

The output of sensitivity analyses and singular optimisation of WI and methane content of LNG product demonstrated a variation in the resulted key decision variables. Table 1 illustrates the complete set of decision variables of the two optimisation problems for the different configurations.

Table 2. Singular optimisation results

Parameter	$\dot{T}_{BOG}$(°C)	$\dot{T}_{LNG}$(°C)	$\dot{m}_{BOG}$(t/h)	$\dot{P}_{BOG}$(bar)	$\dot{P}_{LNG}$(bar)	Objective function
Objective			Flash drum configuration			
WI	-100	-100	20	20	20	80.28
CH_4 content	-200	-200	20	2	3	884 (t/h)
Objective			Single column configuration			
WI	-165	-199	20	1	20	53.00
CH_4 content	-102	-166	0.5	19	2	578 (t/h)
Objective			Double columns configuration			
WI	-200	-167	0.5	20	2	52.61
CH_4 content	-168	-166	20	2	20	705 (t/h)

The flash drum configuration demonstrated the highest in terms of WI and CH_4 content results followed by double columns configuration and lastly single column configuration. The optimal parameters for the flash drum configurations demonstrate the requirement to feed the NG and recycled BOG at the highest temperature, pressure and flowrate of the range, -100 °C, 20 and 20, respectively, to achieve the highest value of WI and a flowrate of BOG at 0.5 t/h. Vase versa, the highest CH_4 content is achieved at the lowest temperature and pressure of the range, -200 °C, 2 and 3, respectively and the highest flowrate of BOG at 20 t/h. This represents the importance to maintain the recycle process at high pressure and low temperature to ensure achieving highest content of methane in the LNG product with highest WI. Whereas, the optimal parameters for the single column configurations demonstrate the requirement to feed the NG and recycled BOG at the temperature, pressure and flowrate of -165 °C, -199 °C, 20 and 1, respectively, to achieve the highest WI. Vase versa, the highest CH_4 content for the single column configurations is achieved at the temperature and pressure of -102 °C, -166, 19 and 2, respectively and a flowrate of BOG at 0.5 t/h. The results indicate the possibility to feed the BOG at atmospheric pressure condition with -165 °C to achieve the carryover of nitrogen content in the fuel gas while increasing these conditions will ensure the maximum methane content in the LNG product. The last configuration of double columns demonstrate the requirement to feed the NG and recycled BOG at the temperature, pressure and flowrate

of -200 ºC, -167 ºC, 20 and 0.5, respectively, to achieve the highest WI. Vase versa, the highest CH_4 content for the double columns configurations is achieved at the temperature and pressure of -168 ºC, -166, 2 and 20, respectively and a flowrate of BOG at 20 t/h. The main reason behind these results is the fact that both columns of the configuration are operated at atmospheric pressure. Hence, feeding the BOG at cold temperatures and low pressure ensures to separate most of the methane out to the LNG stream. Opposite of that ensures to have the highest WI.

Utilising the formulation to solve the two computing objectives simultaneous results on the trade-off between WI and methane content of LNG product. The optimal solutions for each recycling configurations, generated using Matlab Genetic Algorithm tool, are presented as Pareto fronts in Figures (1-3). The three figures demonstrate the trade-off between WI and methane content of LNG product where the increase in one leads to the decrease of other. These solutions represent the possible operating points of the integrated LNG-BOG system to achieve BOG recycling while maintain the process design specs.

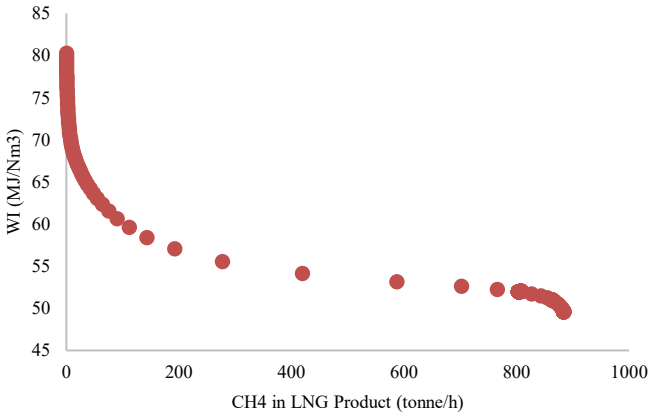


Figure 1: Pareto curve of WI and methane content objectives for flash drum configuration.

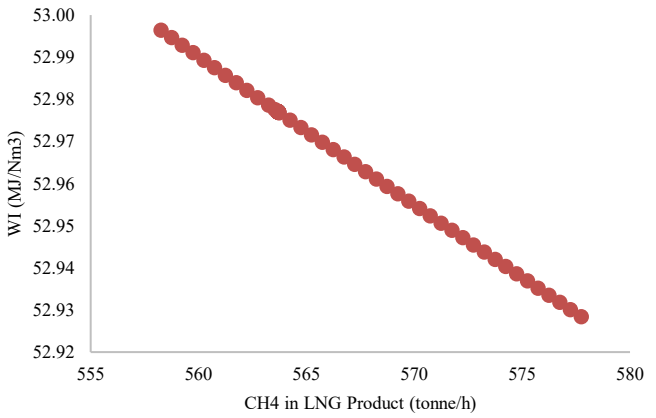


Figure 2: Pareto curve of WI and methane content objectives for single column configuration.

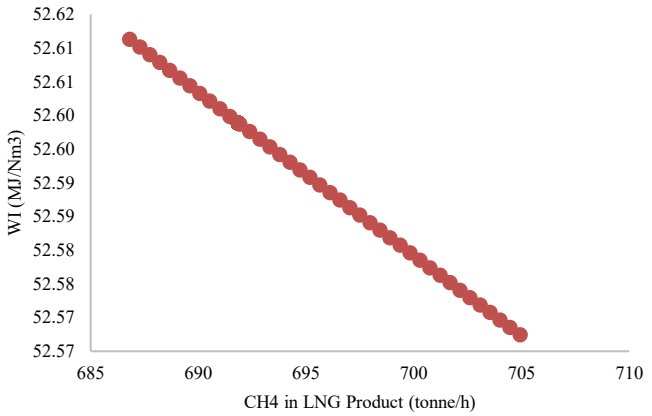


Figure 3: Pareto curve of WI and methane content objectives for double columns configuration.

4. Conclusions

The liquefaction of NG in the well-known Liquefied Natural Gas (LNG) process dominates the energy market despite the intensive energy requirement. LNG process is associated with a number of challenges; one of which is the formation of boil-off gas (BOG) during LNG storage loading and shipping operations. This study explore the optimisation of key design parameters involved in the recovery and utilisation of BOG within different recycling configurations in the LNG production process. The flowsheet models are established to study the effect of changing temperature, pressure and mass flow of recycled BOG, in addition to the temperature and pressure of LNG feed on the final methane content, nitrogen content, mass flow and temperature of LNG product and fuel gas along with the WI of LNG product. The Pareto fronts generated from the multi objective optimization demonstrate the trade-off between WI and methane content of LNG product where the increase in one leads to the decrease of other. These solutions represent the possible operating points of the integrated LNG-BOG system to achieve BOG recycling while maintain the process design specs

References

S.A. Al-Sobhi, A. AlNouss, M. Shamlooh, K. Al-Nuaimi, A. AlMulla, and M. Khraisheh, 2021, Sustainable boil-off gas utilization in liquefied natural gas production: Economic and environmental benefits, Journal of Cleaner Production, 296, 126563.

J. Bao, T. Yuan, L. Zhang, N. Zhang, and X. Zhang, 2019, Comparative study of three boil-off gas treatment schemes: From an economic perspective, Energy Conversion and Management, 201, 112185.

Y.M. Kurle, S. Wang, and Q. Xu, 2015, Simulation study on boil-off gas minimization and recovery strategies at LNG exporting terminals, Applied Energy, 156, 628-41.

Qatargas, 2019, North Field Expansion project,

M.W. Shin, D. Shin, S.H. Choi, E.S. Yoon, and C. Han, 2007, Optimization of the Operation of Boil-Off Gas Compressors at a Liquified Natural Gas Gasification Plant, Industrial & Engineering Chemistry Research, 46, 20, 6540-45.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50190-1

An Implicit Function Formulation for Nonlinear Programming with Index-1 Differential Algebraic Equation Systems

Robert B. Parker[a], Bethany L. Nicholson[b], John D. Siirola[b], Carl D. Laird[a], Lorenz T. Biegler[a*]

[a]*Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, United States*
[b]*Sandia National Laboratories, P.O. Box 5800, Albuquerque, NM 87123, United States*

**biegler@cmu.edu*

Abstract

A reduced space formulation for optimization of index-1 differential algebraic equation systems (DAEs) is described and implemented using PyNumero. The formulation defines implicit functions from algebraic equations and uses them to remove algebraic variables and equations from the optimization problem. The formulation is used to solve dynamic optimization problems with the Ipopt solver. In challenging case studies involving the simulation and optimization of a chemical looping combustion reactor, the reduced space formulation is more robust, solving 115 out of 125 problem instances in the first case and 52 out of 72 instances in the second case. The full space simultaneous formulation solves 50 and 24 problem instances, respectively. The results indicate the potential of this formulation to be more reliable than the full space formulation for challenging nonlinear DAE optimization problems.

Keywords: Dynamic, Optimization, Software

1. Introduction

Differential algebraic equation systems (DAEs) are an expressive class of equations for chemical and process systems engineers. They are capable of representing processes that evolve over continuous time and/or space domains and involve nonlinear physical and chemical phenomena. Complicated interactions among differential states are often described by large numbers of nonlinear algebraic equations that represent the thermodynamics, chemical reactions, and transport phenomena of the system. Due to their number and complexity, the ability to simulate or optimize a large-scale DAE relies on the ability to converge these algebraic equations without excessive problem-specific initialization.

Two common approaches for optimization of DAE systems are the sequential approach described by Goh & Teo (1988) and the simultaneous approach described by Cuthrell & Biegler (1987). Hybrid approaches have been proposed, including the quasi-sequential approach of Hong et al. (2006), which eliminates all equality constraints from the optimization algorithm but retains inequality constraints that may involve state variables. We propose a hybrid approach for DAEs that are index-1 in which only the algebraic variables and equations are eliminated from the nonlinear optimization problem using implicit functions. Our approach is simultaneous in the sense that the continuous domain is discretized and that all discretization points are considered simultaneously by the optimizer, but has the characteristic of sequential approaches that square problems are solved by

an embedded solver within each iteration of the optimization solve. This is a reduced-space formulation as only the differential, input, and derivative variables are seen by the optimization algorithm. The implicit function subproblems admit exact first and second derivatives in terms of these variables and are thus compatible with second-order optimization algorithms. Our approach is conceptually similar to that of Bongartz & Mitsos (2017), where external functions are used to solve global flowsheet optimization problems in a reduced space. Rather than implementing a fully implicit function, however, they use a sequential modular function evaluation for which they compute convex relaxations.

2. Background and Implementation

A discretized DAE has the form given by Equation (1). Here F describes the differential equations, G describes the algebraic equations, d describes the discretization equations, and x_0 are the initial conditions. The equations are in terms of differential variables x, algebraic variables y, input variables u, and derivatives $\dot{x}$ with respect to a continuous domain. We consider the discretized DAE, in which variables are expressed at discrete points that evolve from point t^- to t, within the continuous domain. With the exception of initial conditions and discretization equations at the first point in the domain, these variables and equations are repeated at every point along the domain. Functions F, G, and d are vector valued and assumed to be twice continuously differentiable in all arguments.

$$\dot{x}(t) = F(x(t), y(t), u),\ G(x(t), y(t), u) = 0,\ d(x(t), x(t^-), \dot{x}) = 0,\ x(0) = x_0 \tag{1}$$

We say that a DAE is index-1 if the Jacobian of algebraic equations with respect to algebraic variables, $\nabla_y G$, is nonsingular for all values of differential, algebraic, and input variables. In this case, by the implicit function theorem, there exists a function G_y which maps the vector (x, u) to y such that $0 = G(x, y, u)$ is satisfied.

A simultaneous nonlinear programming (NLP) formulation for the optimization of a DAE model has the form given by Equation (2). We refer to this as the full space formulation. Here, the DAE is fully discretized and equations are repeated at every point in time. We omit the time indexing for brevity.

$$\min_{(\dot{x},x,y,u)\in\mathcal{W}} \varphi(x, u),\ \text{s.t. } \dot{x} - F(x, y, u) = 0,\ G(x, y, u) = 0,\ d(x, \dot{x}) = 0,\ x(0) = x_0 \tag{2}$$

where $\mathcal{W}$ indicates the bounds for the optimization variables. A dynamic optimization problem of this form can be constructed in an algebraic modeling language such as Pyomo (Bynum et al. (2021)) if the constraints and objective are explicit functions of the variables. The optimization problem may then be solved with a generic nonlinear programming solver. In contrast, our reduced space formulation takes advantage of the index-1 property to eliminate algebraic variables and equations from the NLP. That is, $G(x, y, u) = 0$ and the index-1 property imply that $y = G_y(x, u)$, where G_y is an implicit function. The formulation is given by Equation (3).

$$\min_{(\dot{x},x,u)\in\mathcal{W}} \varphi(x, u), \quad \text{s.t.} \quad \dot{x} - F(x, G_y(x, u), u) = 0, \quad d(x, \dot{x}) = 0, \quad x(0) = x_0 \tag{3}$$

An optimization problem of this form cannot be constructed directly in an algebraic modeling environment because G_y is an implicit function. This formulation can still be solved

by a nonlinear programming algorithm, however, if G_y may be evaluated and admits sufficient derivative information. We implement an interface that evaluates such an implicit function and computes first and second derivatives via the implicit function theorem. The interface makes use of the PyNumero extension to the Pyomo algebraic modeling environment (Rodriguez et al. (2021)), and considers NLPs of the form shown by Equation (4), in which f describes the residual equations, which are exposed to the NLP solver and g describes the external equations, which are removed as an implicit function.

$$\min_{(a,b)} \ \varphi(a) \ \text{ s.t. } \ f(a,b) = 0, \ \ g(a,b) = 0, \ \ a \geq 0 \tag{4}$$

Assuming the Jacobian of external equations with respect to external variables, $\nabla_b g$, is nonsingular, the external equations may be removed from the NLP and used to solve for b as a function of a. The system seen by the NLP solver is given by Equation (5).

$$\min_{a} \ \varphi(a) \ \text{ s.t. } \ \bar{f}(a) = f(a, b(a)) = 0, \ \ a \geq 0 \tag{5}$$

To solve this NLP with a second-order optimization algorithm, the Hessian of the Lagrangian of Equation (5), $\nabla^2_{aa}\mathcal{L}$, and the Jacobian of $\bar{f}$, $\nabla_a \bar{f}$, are necessary. These derivatives are given by Equation (6).

$$\begin{aligned}
\nabla_a \bar{f} = \ & \nabla_a f + \nabla_a b \nabla_b f \\
\nabla^2_{aa}\mathcal{L} = \ & \nabla^2_{aa}\varphi + A^{-1}Z + \textstyle\sum_i \nabla^2_{aa} f_i \lambda_{f_i} + \sum_j \nabla^2_{aa} g_j \lambda_{g_j} \\
& + \left(\textstyle\sum_i \nabla^2_{ab} f_i \lambda_{f_i} + \sum_j \nabla^2_{ab} g_j \lambda_{g_j}\right) \nabla_a b^T \\
& + \nabla_a b \left(\textstyle\sum_i \nabla^2_{ba} f_i \lambda_{f_i} + \sum_j \nabla^2_{ba} g_j \lambda_{g_j}\right) \\
& + \nabla_a b \left(\textstyle\sum_i \nabla^2_{bb} f_i \lambda_{f_i} + \sum_j \nabla^2_{bb} g_j \lambda_{g_j}\right) \nabla_a b^T
\end{aligned} \tag{6}$$

Output coordinates of vector function f are indexed by i and coordinates of g are indexed by j. For the barrier term $A^{-1}Z$, $A = \text{diag}(a)$, $Z = \text{diag}(z)$, and z and λ_f are the bound and equality multipliers, respectively, in (5). Jacobian $\nabla_a b$ and vector λ_g are given by $\nabla_a b = -\nabla_a g \nabla_b g^{-1}$ and $\lambda_g = -\nabla_b g^{-1} \nabla_b f \lambda_f$. The derivatives in (6) are derived via the implicit function theorem and the chain rule using an approach similar to that of Heinkenschloss (2008). All Jacobian and Hessian matrices are calculated by PyNumero's interface to the Ampl Solver Library (ASL) (Gay (1997)). The matrices $\nabla_a \bar{f}$ and $\nabla^2_{aa}\mathcal{L}$ are sent to CyIpopt via the PyNumero interface, allowing us to solve NLPs in the form of Equation (3) with the Ipopt algorithm.

With an implicit function interface for systems in the form of Equation (4), the application to DAE systems in the form of Equation (1) is straightforward. Residual constraints are differential and discretization equations, external constraints are algebraic equations, external variables are algebraic variables y, and the optimization problem is solved in the space of differential variables, x, their time derivatives $\dot{x}$, and DAE inputs u. A separate implicit function is used to describe the differential and algebraic equations at each point in the discretized domain.

3. Case Studies

3.1. Chemical Looping Combustion Simulation

A chemical looping combustion (CLC) reactor is a gas-solid hydrocarbon reactor in which fuel and air reactions occur in separate chambers, with a metal-oxide oxygen carrier loop-

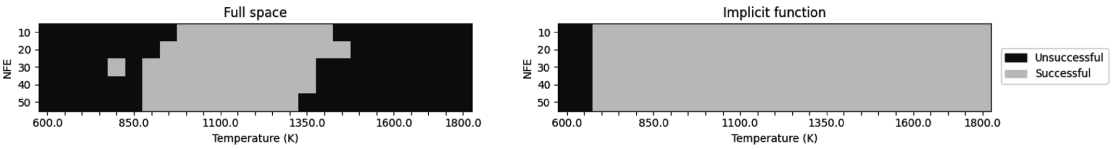


Figure 1: Convergence status of the simulation problem for each combination of NFE and gas inlet temperature for the full space and implicit function formulations.

ing between the two. Here we present results for simulation of the fuel reactor involving methane and iron oxide operating at steady state in a counter-current moving bed configuration. The reduction reaction is $CH_4 + 12Fe_2O_3 \rightarrow 2H_2O + CO_2 + 8Fe_3O_4$. The model equations form a DAE in which the continuous domain is distance along the length of the reactor. Differential variables are flow rates and pressure, and algebraic variables are thermodynamic, hydrodynamic, and kinetic properties. In the current simulation case study, the model has no inputs; inlet conditions are fixed. Further details are given by Ostace et al. (2018) and Okoli et al. (2020).

To compare robustness and solve times of the full and reduced space formulations, we perform a parameter sweep, varying the number of finite elements (NFE) in the spatial discretization and the gas phase inlet temperature. To initialize each simulation or optimization, we initialize differential variables to their inlet values, deactivate discretization equations, and solve the square problem at each discretization point individually. In addition, we apply the following scaling factors: 100 for material flows, $1e3$ for gas enthalpy flow, and $1e5$ for solid enthalpy flow. Figure 1 shows the convergence status of each combination of parameters for the full space and implicit function formulations. An unsuccessful simulation may be due to timeout, iteration limit, function evaluation error, or convergence to an infeasible point. Among problem instances that converge in both formulations, full space solves in less than 0.1 s, while implicit function solves take an average of 30 s. The data indicate that while the implicit function formulation takes more time for the solver to converge, it can also solve significantly more instances of the simulation problem than the full space formulation. Of the 125 problem instances, the implicit function formulation can solve 115, while the full space formulation can solve only 50, with the same initialization. While additional problem-specific tuning on initialization and scaling may improve convergence for some instances, the implicit function formulation has the advantage of fewer variables to initialize and scale.

3.2. Chemical Looping Combustion Optimization

We now present results for optimization of operating conditions of the chemical looping combustion reactor. The model is the same as in Section 3.1., but now inlet conditions are degrees of freedom and outlet conditions participate in an objective function penalizing their deviation from target conditions. The target outlet conditions are taken from the moving bed process described in Table 5 of Okoli et al. (2020).

Because inlet gas temperature is no longer a fixed parameter, we now vary the gas temperature used for model initialization, as well as the number of finite elements, in a parameter sweep. Four different NFEs and 18 temperatures are used for 72 total problem instances.

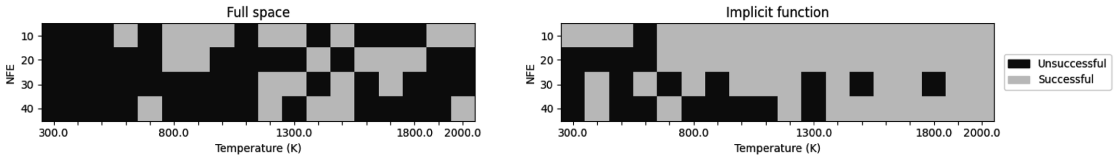


Figure 2: Convergence status of the optimization problem for each combination of NFE and initialization gas inlet temperature for both formulations

Figure 2 shows the convergence status for all combinations of parameters. Among problem instances for which both formulations converge, the average full space solve is much faster than the implicit function solve. A breakdown of solve times for the instance with 30 finite elements, initialized to 1200 K, is shown in Table 1. In this instance, the implicit function formulation takes 203 s and converges in 47 Ipopt iterations, while the full space formulation takes 1.3 s and converges in 95 Ipopt iterations. The "Interface" category refers to time spent in function and derivative evaluations, and is broken down more finely for the implicit function formulation. These calculations take 92 % of the time in this formulation, are performed independently at each point in the discretized domain, and may be done in parallel. In this instance, up to 30 processors can be utilized, in which case the entire implicit function solve would be performed in only 22 s with perfect speedup.

Table 1: Percentage of solve time spent in each activity for full and reduced space formulations of the CLC optimization problem with 30 finite elements initialized to 1200 K

Formulation	I/O	Ipopt	Interface	Function	Jacobian	Hessian	other
Full	25 %	50 %	21 %	–	–	–	4 %
Implicit	7 %	< 1 %	–	86 %	4 %	2 %	1 %

The data again indicate that although the reduced space formulation takes longer to converge than the full space formulation, there are many instances for which only the reduced space formulation converges. Out of 72 problem instances, the reduced space formulation can solve 52, and the full space formulation can solve only 24 with the same initialization. We hypothesize that the improved robustness of the reduced space formulation is because the reduced space and external Jacobians, $\nabla_a \bar{f}$ and $\nabla_a b$ in (6), are less likely to be poorly conditioned at points where the algebraic equations have been converged than at arbitrary infeasible points where the algebraic equations may have large residuals. While the implicit function formulation is slower than the full space formulation, we note that much more can be done to accelerate the implicit function formulation. The implicit function and derivative evaluations for each point in the discretized domain can be performed in parallel, implicit function computations can be performed entirely in compiled code by operating on a PyNumero `PyomoNLP` object rather than directly in Pyomo, and sparsity in the derivative matrices can be exploited. We defer these improvements to future work.

4. Conclusions

We have implemented an implicit function formulation for the simulation and optimization of index-1 DAE systems via nonlinear programming. In the application of our formulation

to the simulation and optimization of a chemical looping combustion reactor, our formulation solves over twice as many problem instances as the full space formulation, illustrating its potential to be much more robust. When it converges, however, the full space formulation is significantly faster than our current implementation.

Future extensions to this work include improving the computational performance, alternative formulations removing only a subset of the algebraic equations from the optimization problem, and applying implicit functions with exact first and second derivatives to other types of decomposable optimization problems. A comparison with a fully sequential dynamic optimization formulation with exact first and second derivatives, performed with our current Ipopt/PyNumero computational framework, will also be valuable.

Disclaimer

This work was supported through the Simulation-Based Engineering, Crosscutting Research Program within the U.S. Department of Energy's Office of Fossil Energy. This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

References

Bongartz, D. & Mitsos, A. (2017), 'Deterministic Global Optimization of Process Flowsheets in a Reduced Space Using McCormick Relaxations', *J. Glob. Opt.* **69**, 761–796.

Bynum, M. L., Hackebeil, G. A., Hart, W. E., Laird, C. D., Nicholson, B. L., Siirola, J. D., Watson, J.-P. & Woodruff, D. L. (2021), *Pyomo–Optimization Modeling in Python*, Vol. 67, third edn, Springer Science & Business Media.

Cuthrell, J. E. & Biegler, L. T. (1987), 'On the Optimization of Differential-Algebraic Process Systems', *AIChE J.* **33**(8), 1257–1270.

Gay, D. M. (1997), Hooking Your Solver to AMPL, Technical report, Bell Laboratories.

Goh, C. & Teo, K. (1988), 'Control Parametrization: A Unified Approach to Optimal Control Problems with General Constraints', *Automatica* **24**(1), 3–18.

Heinkenschloss, M. (2008), Numerical Solution of Implicitly Constrained Optimization Problems, Technical report, Rice University.

Hong, W., Wang, S., Li, P., Wozny, G. & Biegler, L. T. (2006), 'A Quasi-Sequential Approach to Large-Scale Dynamic Optimization Problems', *AIChE J.* **52**(1), 255–268.

Okoli, C. O., Ostace, A., Nadgouda, S., Lee, A., Tong, A., Burgard, A. P., Bhattacharyya, D. & Miller, D. C. (2020), 'A Framework for the Optimization of Chemical Looping Combustion Processes', *Powder Technology* pp. 149–162.

Ostace, A., Lee, A., Okoli, C. O., Burgard, A. P., Miller, D. C. & Bhattacharyya, D. (2018), Mathematical Modeling of a Moving-Bed Reactor for Chemical Looping Combustion of Methane, *in* 'Proc. 13th International Symposium on Process Systems Engineering (PSE 2018)', Computer-Aided Chemical Engineering, Elsevier, pp. 325–330.

Rodriguez, J., Parker, R., Laird, C., Nicholson, B., Siirola, J. & Bynum, M. (2021), 'Scalable Parallel Nonlinear Optimization with PyNumero and Parapint'. Preprint at `http://www.optimization-online.org/DB_HTML/2021/09/8596.html`.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50191-3

Multi-objective optimization of NH_3 and CO_2 separation with ionic liquid process

Yuanmeng Duan[a,b], Guoxiong Zhan[b], Fei Chang[b], Sensen Shi[b], Jens Abildskov[a], Jakob Kjøbsted Huusom[a,*], and Xiangping Zhang[b,*]

[a] *Department of Chemical and Biochemical Engineering, Technical University of Denmark, Lyngby 2800, Denmark*
[b] *Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100190, China*
jkh@kt.dtu.dk
xpzhang@ipe.ac.cn

Abstract

Melamine production produces tail gas with a significant amount of both NH_3 and CO_2. Benefiting from the development of green solvents, using ionic liquids instead of traditional solvents as water for NH_3 and CO_2 separation has attracted wide attention. Multi-objective optimization (MOO) is employed to optimize the ionic liquid-based process in this work. In order to decrease the energy consumption in the ionic liquid-based separation process, a MOO research was carried out using Aspen Plus and Matlab software in this work.

In this work, one ionic liquid-based process was simulated and optimized by the nondominated sorting genetic algorithm II (NSGA-II) algorithm. The minimum total separation cost (TSC) and total process CO_2 emission (TPCOE) were set as two objective functions. With the constraints and several operational parameters optimized, the Pareto front displays a set of nondominated, optimal design parameters that satisfy the specification of the NH_3 concentration standard in outlet gas.

The results show that the effect of desorption pressure and the ratio of lean solvent to total solvent are critical for both TSC and TPCOE. After the MOO, the TSC of the ionic liquid process can be decreased by 5%, and TPCOE is reduced by 12% compared with the base case. The optimization results support the optimal design and operation of the NH_3 and CO_2 separation process with ionic liquids considering environmental and economic objectives.

Keywords: Ionic liquid; Process Design; Multi-objective optimization.

1. Introduction

Melamine is an important industrial raw material that can be prepared by condensation polymerization of melamine resin, widely used in tableware, heat insulation materials, coatings, adhesives, etc. However, for the treatment of melamine tail gas containing NH_3 and CO_2, the reaction of these two gases leads to ammonium bicarbonate crystals that cause operational difficulties. Water scrubbing and some co-productions technologies such as urea co-production, ammonia bicarbonate co-production are the primary gas treatment method employed in the industry. However, they have disadvantages such as high energy and water consumption, equipment corrosion, secondary pollution, and difficulty reaching standard discharge. Ionic liquids is a new green material and medium for the industry's gas separation and purification process. Ionic liquids have the advantage

of low vapor pressure, being structurally designable, good stability. Therefore, it is interesting to use its advantage to develop new technology. Based on the functional IL [Bim][NTf_2], it can achieve the target of high absorption capacity, good selectivity of NH_3 and CO_2, no water discharge, low energy consumption, and high purity of recovery gas.

A posteriori techniques using multi-objective approach are mostly population-based stochastic optimization techniques (C.A. Coello Coello, 2007). These include NSGA-II (elitist non-dominated sorting GA), SPEA2 (strength Pareto evolutionary algorithm-2), MOEA (multi-objective evolutionary algorithm), and MOPSO (multi-objective PSO) (K. Deb, 2001). Each iteration ranks the solutions based on objective values to obtain many Pareto-optimal solutions at the end of the given stopping criterion (usually, the maximum number of iterations), all in one run of the optimizer/program. A simplified flow chart of NSGA-II is presented in Figure 1(Rangaiah et al, 2015). In the process of optimization, non-repetitive and dominant solutions and frontier sets are found in the iteration operation to retain the optimal results.

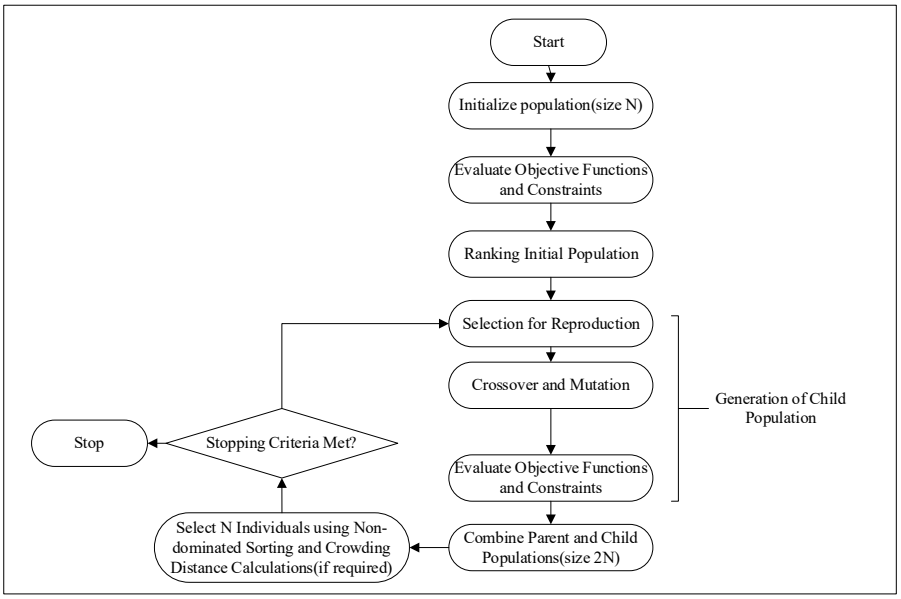


Figure 1. A flowchart of NSGA-II for MOO.

To lower the energy consumption and analyze relationships of operational parameters with TSC and TPCOE of the ionic liquid process, MOO research was carried out in the separation of NH_3 and CO_2 with the ionic liquid process using Aspen Plus and Matlab software in this work. Because of the internal relationship with different operating parameters in the ionic liquid process, the optimization may go into the local optimization solution instead of the global optimization solution. In applying chemical engineering, most practical problems usually seek to satisfy this under multiple design goals—the MOO problem of the best design scheme for these objectives. On the one hand, MOO can give full play to the role of decision-making participants, and on the other hand, it is more in accordance with reality(C.A. Coello Coello, 2007). Researchers usually integrate mathematical software and process simulation software to solve the optimization problem of multiple objective functions.

Recently, many researchers have used MOO techniques for the analysis of chemical processes. Li et al. (2021) used a MOO strategy to integrate the design and control of ionic liquid-based extractive distillation processes. The Pareto front shows the optimization results, which are based on trade-offs between controllability and economics. Zhan et al. (2021) investigated the ammonia-containing gas separation and ammonia recovery with ionic liquids. MOO was employed to finish the optimization and the result showed that using functional ionic liquids could greatly decrease purification cost and energy consumption.

However, there is limited research on MOO for the separation of NH_3 and CO_2 by the ionic liquid process. In this work, the MOO of the ionic liquid-based NH_3 and CO_2 separation process was made, using the nondominated sorting Genetic algorithm II(NSGA-II).

2. Methods

2.1 Process flow chart of NH_3/CO_2 separation with ionic liquids

The base case has been previously established and the flow chart of the NH_3/CO_2 separation process with ionic liquids is shown in Figure 2(Duan et al, 2021). The melamine tail gas contains 7.6 % nitrogen (N_2), 0.4 % water (H_2O), 55 % ammonia (NH_3) and 37 % carbon dioxide (CO_2) in mole fractions. The conceptual process consists of two sections, which are the NH_3 absorption section and absorbent desorption section. The tail gas flows to the bottom of the absorber and meets the ionic liquid in counter current flow and then NH_3 is released due to increased pressure and decreased temperature in Flash 1 and Flash 2. After tail gas separation, the NH_3 concentration below 6000 ppm (mole fraction) in the purified gas. The NH_3 product purity is set over 0.996 mass fraction.

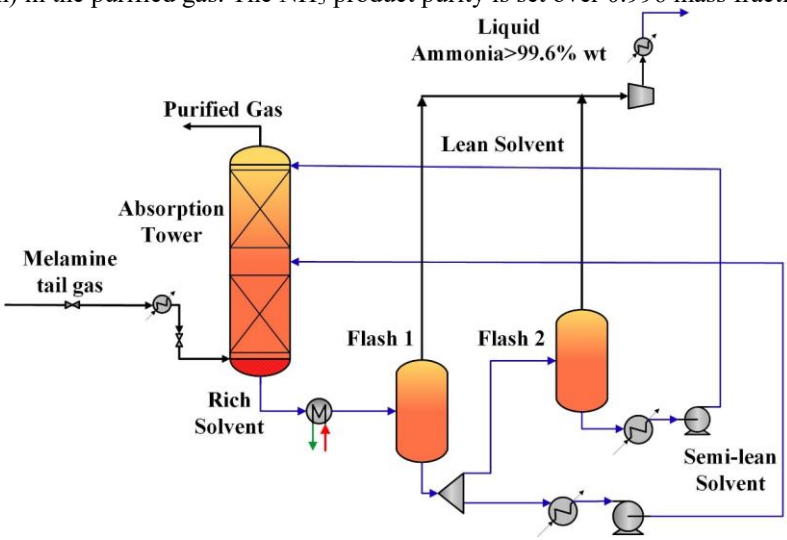


Figure 2. The flow chart of NH_3/CO_2 separation process with ionic liquids (Duan et al, 2022).

2.2 MOO of NH_3/CO_2 separation with ionic liquids

In order to investigate the internal relationship among different parameters and TSC of the ionic liquid-based process, MOO was carried out. In this work, the NSGA-II is employed in the MOO of the NH_3 and CO_2 separation process. MOGA can be employed to solve MOO to avoid falling into a local optimal solution when sequential iteration was

used in the iterative process. As a bridge between Aspen Plus and Matlab, ActiveX is used in the MOO process. Six design parameters are simultaneously optimized, including pressure of absorption, ratio of lean solvent to total solvent, total usage of solvent, temperature of Flash1, pressure of Flash1 and 2 (P, R_{LT}, F, T , P_{F1}, P_{F2}), the value range of each parameter is shown in Table 1. In this study, the tray number of absorption is not used as a decision variable, it is fixed at 8.

Table 1. Range of parameters of NH_3/CO_2 separation with ionic liquid.

Operation parameters	P (kPa)	R_{LT}	F (t/h)	T (K)	P_{F1} (kPa)	P_{F2} (kPa)
Lower bound	101	0.2	10	363.15	10	1
Higher bound	405	0.8	100	383.15	50	10

In process optimization, objective functions consist of the minimization of TSC(Total Separation Cost) and TPCOE(Total process emission of CO_2), the constraints are that $y_{NH3,\ purified} \leq 6000$ ppm, $x_{NH3,\ wt} \geq 0.996$, TSC ≥ 0, TPCOE ≥ 0, the ratio of gas and liquid 1 (GL1) ≤ 1000, the ratio of gas and liquid 1(GL2) ≤ 500. The MOO procedure of ionic liquid-based process is shown in Figure 2. The selection type is binary tournament selection. The crossover probability is 0.8. The variation method is Gaussian variation, and the variation probability is 0.1, the population size is set as 100, and the number of generations is set as 15.

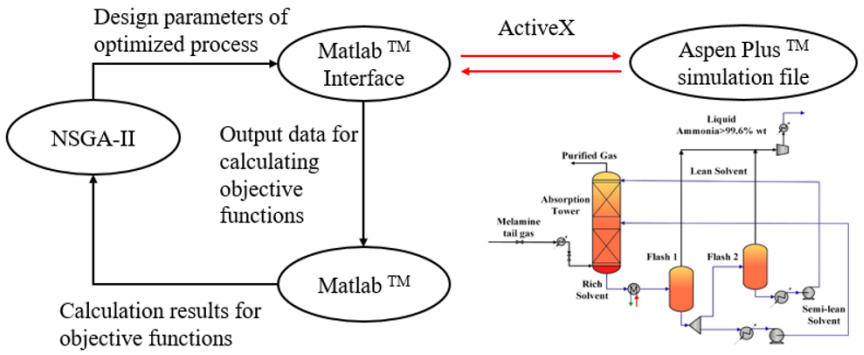


Figure 3. The MOO procedure of ionic liquid-based process.

3. Results and discussion

Based on the sensitivity analysis of the ionic liquid-based process, the MOO is conducted to obtain the solution for this chemical process, which can realize the two optimal objective functions: TSC and TPCOE. Figure 4(a) shows the relationship between TSC and TPCOE. The Pareto optimal solution of the NH_3 and CO_2 separation process indicates that TSC rises with the increasing of TPCOE. In the separation process, low TSC means low energy consumption, which leads to lower TPCOE.

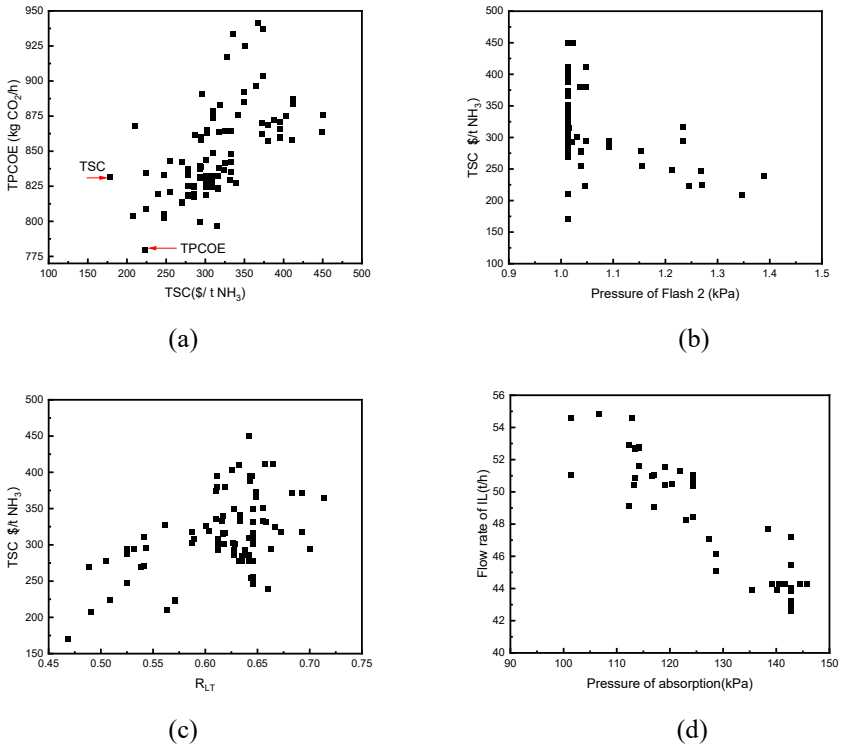


Figure 4. The MOO results. (a) The Pareto optimal result of NH_3 and CO_2 separation of TSC and TPCOE. (b) Effects of key variable ratio of lean solvent to total solvent (R_{LT}). (c) Effects of key variable pressure of Flash 2(P_{F2}). (d) Relationship between Pressure of absorption (P) and Flow rate of ionic liquids (F).

Figure 4(b) and (c) show the effects of crucial input variables (P_{F2}, R_{LT}) on TSC. From Figure 4(b), the pressure change of Flash 2 will lead to the diversification of objective function TSC. From Figure 4(c), TSC change with R_{LT} because higher R_{LT} means more lean solvent will flow into the Flash 2 and give more burden to Flash 2 with lower pressure.

For effect among the input parameters, Figure 4(d) shows the relationship between absorption pressure and flow rate of ionic liquids. With the increase of pressure, more NH_3 will be absorbed by ionic liquids, then less ionic liquids will be used in this process. Although increasing flow rate of ionic liquid can gain the solvent cost, it is observed that the operation can reduce the absorption pressure. These two parameters have a balance with each other about the effect on TSC.

Because there are two objective functions, we select the designs in the Pareto optimal solution named case 1 and 2 with the lowest TSC and TPCOE. Table 1 offers TSC and TPCOE of all configurations of the ionic liquid process. As the calculation, case 1 shows the lowest TSC as 170.87 \$/t NH_3 which decreased by 5% of the base case, and case 2 has the lowest TPCOE decreased by 12% as 779.85 kg/h of the base case. In summary, the optimized results could give more detail information of ionic liquid-based NH_3/CO_2 separation process, which provides a perspective for separation technology for the future.

Table 2. Optimization results of NH_3/CO_2 separation with ionic liquid.

Operation parameters	Case 1	Case 2	Unit
Theoretical stage of absorption	8	8	-
Feed stage position of semi-lean	4	4	-
Total ionic liquid flow rate	47.09	43.90	$t \cdot h^{-1}$
Absorption pressure	128	142	kPa
Pressure of Flash 1	10	10	kPa
Temperature of Flash 1	366.10	363.98	K
Pressure of Flash 2	1	2	kPa
Temperature of Flash 2	366.10	363.98	K
R_{LT}	0.47	0.57	-
TSC	170.87	222.91	$\$ \cdot t^{-1}$ NH_3
TPCOE	831.29	779.85	$kg \cdot h^{-1}$

4. Conclusion

In this work, MOO was carried to separate NH_3 and CO_2 using the ionic liquid. The result showed that desorption pressure and the ratio of lean solvent to total solvent affects TSC. In addition, the pressure of absorption and flow rate of ionic liquid have conflict and balance with each other. After optimization, the TSC and TPCOE decreased 5% and 12% compared with the base cases, respectively. The optimization results provide information for optimal design and operation of NH_3 and CO_2 separation processes with ionic liquids considering environmental and economic objectives. In next step work, the different algorithm will be employed to compare the effect to the identification of critical parameters.

References

C.A. Coello Coello, G.B. Lamont, D.A. Van Veldhuizen. Multi-objective Optimization and Multi-criteria Decision Making, Evolutionary Algorithms for Solving Multi-Objective Problems (2nd ed.), Springer, Berlin, Heidelberg (2007), 220-221.

K. Deb. Multi-objective Optimisation in Manufacturing Supply Chain Systems Design, Multi-Objective Optimization Using Evolutionary Algorithm. John Wiley & Sons, Chichester, UK (2001), 53-54.

GP. Rangaiah, S. Sharma, B. Krishna Sreepathi,2015, Multi-objective optimization for the design and operation of energy efficient chemical processes and power generation, Current Opinion in Chemical Engineering, 10, 49-62.

J. Li, L. Li, R. Li, Z.Yang, Z. Ma, L. Sun, N. Zhang, 2021, Investigation of multi-objective optimization for integrating design and control of ionic liquid-based extractive distillation, Chemical Engineering Research and Design, 170, 134-146.

G. Zhan, F. Cao, L. Bai, F. Chang, B. Zhou, Y. Duan, S. Zeng, H. Dong, Z. Li, X. Zhang. 2021, Process simulaiton and optimization of aoomnia-containing gas separation and ammonia recovery with ionic liquids, ACS Sustainable Chem. Eng, 9, 312-325.

Y. Duan, G. Zhan, F. Chang, S, Shi, S. Zeng, H. Dong, J. Abildskov, J. Huusom, X. Zhang. 2022,process simulation and evaluation for NH_3/CO_2 separation in melamine tal gas with ionic liquids, Separation & Purification technology, submitted.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50192-5

Primal-dual Feedback-optimizing Control with Direct Constraint Control

Risvan Dirza, Dinesh Krishnamoorthy, and Sigurd Skogestad*

Dept. of Chemical Engineering, Norwegian Univ. of Science & Technology (NTNU), NO-7491 Trondheim, Norway
**sigurd.skogestad@ntnu.no*

Abstract

This work proposes a strategy to track steady-state changes in active constraints and minimize dynamic constraint violations in order to achieve system-wide optimal operation using simple feedback control structures and logic blocks. The strategy is based on the recently proposed primal-dual feedback-optimizing control scheme that optimally handles steady-state changes in active constraints. However, the constraints are controlled in a slow time scale by updating the dual variables (Lagrange multipliers). To reduce dynamic constraint violations, we propose a "fix-up" to the primal-dual scheme with direct control of hard constraints. We show that the improved method can reduce profit loss in the long run by allowing for smaller back-off from hard constraints. The application is to coordinated control of gas-lifted oil wells.

Keywords: Distributed feedback-optimizing control, Oil/gas, Production optimization.

1. Introduction

The optimal process operation involves making decisions in real-time to meet production goals. This is typically done in the context of real-time optimization (RTO) using mathematical concepts, process models, and real-time measurements. In the 80s, there was an increasing interest in replacing model-based numerical solvers with a simple feedback loop, named feedback-optimizing control. The idea is to translate the economic objective into a process control objective by finding a function of the controlled variables (CVs), and when it is held constant, it leads to the optimal adjustment of the manipulated variables (MVs). These MVs drives the process to optimal operating condition (Morari et al., 1980). Twenty years later, Skogestad (2000) suggested replacing the term "optimal adjustments" with "acceptable adjustments" (in terms of the loss). This idea is known as self-optimizing control (SOC). In SOC, *"when the optimum lies at some constraints, we use active constraint control where the available MVs tightly control the constrained variables"*. When the optimum may be unconstrained, the self-optimizing CVs are measured variables or combinations of them. We need a good model to determine (offline) an accurate self-optimizing CV, and it can be time-consuming if we have a complex and large-scale system. Not considering noise, the ideal self-optimizing CV is the gradient of the cost function w.r.t. the control input, that when we keep at a constant setpoint of zero, it satisfies the necessary conditions of optimality (Halvorsen et al., 2003). In constrained cases, the process reaches ideal optimal operating conditions when the gradient of the Lagrange function w.r.t. to control input is kept at a constant setpoint of zero. If the objective function is additively separable, we can decompose the problem and let each local system controls its local gradients of the Lagrange function w.r.t. local

control input. *In this framework, we need a central coordinator to update the shadow price of shared constraints and broadcast it to every subsystem (Wenzel et al., 2016).*

2. Recent Works and Problem Statement

Consider the following steady-state optimization problem of N different subsystems.

$$\min_{\mathbf{u}} J(\mathbf{u},\mathbf{d}) = \sum_{i=1}^{N} J_i(\mathbf{u}_i,\mathbf{d}_i) \tag{1a}$$

$$s.t.\ \ \mathbf{g_s}(\mathbf{u},\mathbf{d}) \leq \mathbf{0} \tag{1b}$$

where $\mathbf{u}_i \in \mathbb{R}^{n_{\mathbf{u}i}}$ denotes the MVs for subsystem i, $n_{\mathbf{u}_i}$ is the number of MVs in subsystem i, and $\mathbf{u} = [\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_N]^{\mathrm{T}}$, $\mathbf{d}_i \in \mathbb{R}^{n_{\mathbf{d}i}}$ denotes the disturbances in subsystem i, $n_{\mathbf{d}_i}$ is the number of disturbances in subsystem i, and $\mathbf{d} = [\mathbf{d}_1 \quad \cdots \quad \mathbf{d}_N]^{\mathrm{T}}$, $J_i: \mathbb{R}^{n_{\mathbf{u}i}} \times \mathbb{R}^{n_{\mathbf{d}i}} \rightarrow \mathbb{R}$ is a function that denotes the local objective of subsystem i, $\mathbf{g_s}: \mathbb{R}^{n_{\mathbf{u}}} \times \mathbb{R}^{n_{\mathbf{d}}} \rightarrow \mathbb{R}^{n_{\mathbf{gs}}}$is a function that denotes the inequality (shared) constraints. $n_{\mathbf{g_s}}$ is the number of constraints. The Lagrangian function of problem (1) is $\mathcal{L}(\mathbf{u},\mathbf{d},\boldsymbol{\lambda}_{g,s}) = \sum_{i=1}^{N} J_i(\mathbf{u}_i,\mathbf{d}_i) + \boldsymbol{\lambda}_{g,s}^{T}\mathbf{g_s}(\mathbf{u},\mathbf{d})$, where $\boldsymbol{\lambda}_{g,s} \in \mathbb{R}^{n_{\mathbf{gs}}}$ is the shadow price of the (shared) resource constraints. The goal of problem (1) is to determine optimal MVs to achieve system-wide steady-state optimal operation. Our motivation is to solve problem (1) using a feedback control structure that handles changing active constraints.

One possible approach is the reduced gradient approach or region-based control (Jäschke and Skogestad, 2012). *This method is easy to implement for a simple case with a few regions, and the result usually converges faster than the decomposed one for a large-scale problem. However, this method can be problematic for a complex and large case* as the number of the region is equal to $2^{n_{\mathbf{gs}}}$. Another attractive framework is distributed feedback-based real-time optimization, which is also known as primal-dual feedback-optimizing control (Dirza et al., 2021; Krishnamoorthy, 2021). This method can avoid solving numerical optimization problems online by having real-time iteration of dual/Lagrange decomposition (e.g., Wenzel et al. (2016)). Consequently, it has a central coordinator acting as a "slow" central constraint controller. This structure makes primal-dual flexible in the presence of changing active constraints. The problem with this method is that the constraint is controlled only on the slow time scale through the manipulation of the shadow prices, which is only indirectly through the unconstrained optimization layer that affects the (physical) MVs. This causes the shadow prices (broadcasted to the actual plant) to be suboptimal during the transient. This condition may lead to dynamic violation during the transient, and later lead to an infeasible operation. This violation is unacceptable when we have a hard constraint. Thus, it is necessary to introduce a “back-off” from that constraint. *Note that this back-off will also apply at a steady-state condition, and it may then result in a considerable economic penalty, which can lead to profit loss. This work addresses this violation issue and aims to minimize the profit loss.*

3. Proposed Control Structure

Mathematically, the profit loss scale is linear with the back-off parameter. One can express this as $Loss = -\boldsymbol{\lambda}_{g,s}{}^{T}\boldsymbol{\zeta}_{bo}$, where $\boldsymbol{\zeta}_{bo}$ is the back-off parameter, which means that by reducing the back-off parameter, one can reduce the profit loss in the long run.

Therefore, *this paper proposes an additional structure to control a hard constraint tightly in the primal-dual framework to minimize the back-off parameter*.

Because the primal-dual approach only has a central constraint controller that control the constraints on a slow time scale, we introduce direct constraint control as a ”fix-up” to reduce dynamic constraint violation. The direct constraint control is based on pairing the constraint with a particular MV using a selector. This tightly controls any active (shared-) hard constraints on a fast time scale. This structure automatically switches back to the unconstrained mode when none of those existing constraints turns active. *We introduce this proposed control structure as primal-dual feedback-optimizing control with direct constraint control*. The implementation is discussed in detail in Section 4 (see Fig. 2(b)).

Selectors, which are well-known tools in the industries, are used for active constraint switching (Krishnamoorthy and Skogestad, 2020). The switching determines the assigned value to the specified MV. When using selectors, only one of some control actions is the actual input to the plant at any given time. For the ones that are not selected, the feedback loop is “broken”. Consequently, the integral term is possibly building up. Thus, it is essential to implement anti-windup using a back-calculation scheme.

4. Implementation in Subsea Oil Production Network

We consider a subsea gas-lifted oil well production system, consisting of N clusters, that lift oil from the different reservoirs, completed with a fixed shared gas-lift compressor with limited available power. The production system model is like the one used in Dirza et al. (2021) and an additional model to calculate power consumption of the compressor as a linear function: $Pow_{gl} = \theta \sum_{i=1}^{N} \sum_{j=1}^{N_i} w_{gl,i,j}$, where θ is a function of a fixed ratio of outlet and inlet pressure of the compressor. Further, N_i is the total number of wells in cluster i, and $w_{gl,i,j}$ is the gas-lift rate injected to well j in cluster i.

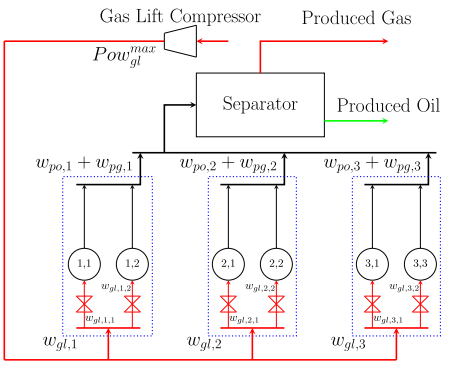


Figure 1: Field illustration

The objective function is to maximize the oil production income while minimizing the cost of the gas lift. The optimization problem is as follows.

$$\min_{\mathbf{w}_{gl}} \sum_{i=1}^{N} \left(- p_{o,i} \sum_{j=1}^{N_i} w_{po,i,j} + p_{gl,i} \sum_{j=1}^{N_i} w_{gl,i,j} \right) \tag{2a}$$

$$s.t. \quad \mathbf{f}(\mathbf{x}, \mathbf{w}_{gl}, \mathbf{d}) = \mathbf{0} \tag{2b}$$

$$\mathbf{g}(\mathbf{x}, \mathbf{w}_{gl}, \mathbf{d}) \leq \mathbf{0} \tag{2c}$$

$$\mathbf{g}_s(\mathbf{x}, \mathbf{w}_{gl}, \mathbf{d}) = Pow_{gl} - Pow_{gl}^{max} \leq 0 \tag{2d}$$

where $p_{o,i}$, $p_{gl,i}$, and $w_{po,i,j}$ are the price of produced oil, the cost of gas-lift, and the produced oil rate of well j in cluster i, respectively. Pow_{gl} is the total power consumed by the fixed compressor to inject the total gas-lift rate i, and Pow_{gl}^{max} is the maximum available power, which can also be a function of back-off parameter, ζ_{bo}. Further, $\mathbf{x} \in$

$\mathbb{R}^{n_{\mathbf{x}}}$, and $\mathbf{d} \in \mathbb{R}^{n_{\mathbf{d}}}$ are the vectors of states, and disturbance (i.e., gas-oil-ratio) for the entire system. $\mathbf{w}_{gl} \in \mathbb{R}^{n_{\mathbf{w}_{gl}}}$ is the vector of inputs for the entire system, which can be seen as a vector of gas-lift rate from each well, $\mathbf{w}_{gl} = [w_{gl,1,1} \quad \cdots \quad w_{gl,N,N_N}]^{\mathrm{T}}$.

Constraint (2b) and (2c) represent model and physical constraints, respectively. We assume that one locally manages constraint (2c) to maintain the focus of the discussion. The objective function (2a) is additively separable. Moreover, Eq. (2d) is a linear and hard constraint. Thus, we can decompose the problem into N subproblems. This case study considers $N = 3$ subsea clusters, where each cluster has two production wells (see Fig. 1) and has different oil prices to indicate the type of oil produced by each reservoir is different.

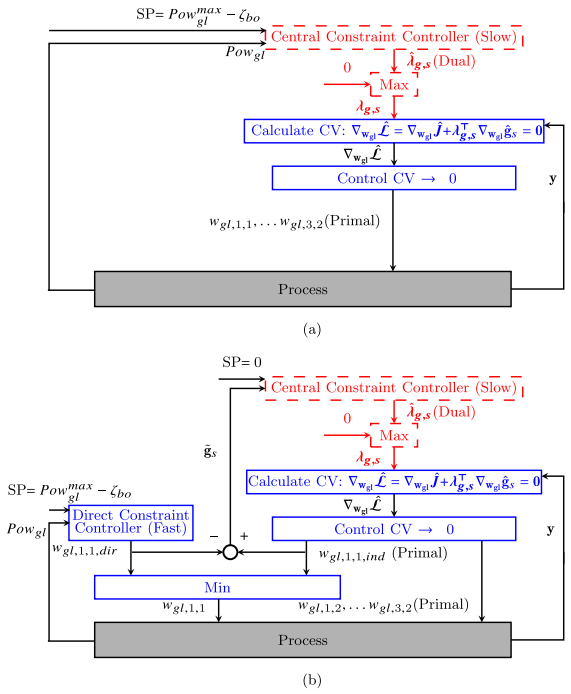


Figure 2: (a) Primal-dual control structure; (b) Proposed control structure which combines primal-dual optimization with direct constraint control.

As primal-dual can converge to optimal steady-state conditions (Dirza et al., 2021; Krishnamoorthy, 2021), *this simulation compares primal-dual (as shown in Fig. 2(a)) with the proposed control structures (as shown in Fig. 2(b))*. Note that **y** indicates the real-time measurements set. The grey boxes represent the physical system. The white boxes with solid blue lines represent a faster timescale computation block, and the white boxes with dashed red lines represent the slower ones.

Table 1: Controlled Variables, Setpoints, and Manipulated Variables

Well	CV	CV^{SP}	Calculated MV	Physical MV
1,1 (indirect)	$CV_{1,1,ind} = \nabla_{w_{gl,1,1}} \mathcal{L}_{1,1}$	0	$w_{gl,1,1,ind}$	$w_{gl,1,1}$
1,1 (direct)	$CV_{1,1,dir} = Pow_{gl}$	Pow_{gl}^{max}	$w_{gl,1,1,dir}$	$w_{gl,1,1}$
i, j^*	$CV_{i,j,ind} = \nabla_{w_{gl,i,j}} \mathcal{L}_{i,j}$	0	$w_{gl,i,j,ind}$	$w_{gl,i,j}$

*: For the remaining well j in cluster i

In the proposed control structure, we assume that well 1 of cluster 1 is technically more feasible to control hard constrained variables tightly. We have the original constraint $\mathbf{g}_s \leq 0$, and by using step response, we obtain that $\frac{d\mathbf{g}_s}{dw_{gl,11}} > 0$. This means that a small value of $w_{gl,1,1}$ is good in terms of satisfying the constraint and a min selector is needed,

$w_{gl,1,1} = min(w_{gl,1,1,dir}, w_{gl,1,1,ind})$, where $w_{gl,1,1,dir}$ is the MV computed by the direct constraint controller, and $w_{gl,1,1,ind}$ is the primal MV by the optimization block. Note that we, at the optimal steady-state, must have $w_{gl,1,1,dir} \geq w_{gl,1,1,ind}$ or equivalently $\tilde{\mathbf{g}}_s = w_{gl,1,1,ind} - w_{gl,1,1,dir} \leq \mathbf{0}$. This is the constraint controlled in the proposed new structure. Table 1 shows the CVs, Setpoints, and the MVs in this case study, where $\nabla_{w_{gl,i,j}} \mathcal{L}_{i,j} = \nabla_{w_{gl,i,j}} \boldsymbol{J} + \lambda_{\boldsymbol{g},s}^T \nabla_{w_{gl,i,j}} \mathbf{g}_s$. Additionally, we use the same method as Dirza et al. (2021) to estimate steady-state cost and constraint gradient, labelled by $\nabla_{w_{gl,i,j}} \hat{\boldsymbol{J}}$ and $\nabla_{w_{gl,i,j}} \hat{\mathbf{g}}_s$, respectively.

The key idea is that we adjust the shadow price $\lambda_{g,s}$ so that on the long run the value of the MV computed by the direct constraint control (when it is active) is equal to the optimal primal value computed by the layer above (see Fig. 2(b)). To determine the applied $\lambda_{g,s}$, one can use a PI controller as a central constraint controller equipped with a max selector that gives 0 when the constraint is no longer active. The anti-windup is necessary to avoid $\lambda_{g,s}$ keeps changing in this case. Thus, this selector selects either 0 or the calculated shadow price $\hat{\lambda}_{g,s}$. Further, that shadow price $\hat{\lambda}_{g,s}$ at iteration k is as follows.

$$\hat{\lambda}_{g,s} = \lambda_{g,s}^k + K_p \tilde{\mathbf{g}}_s{}^k + \sum_{\tau=k-1}^{k} \left(K_I \tilde{\mathbf{g}}_s{}^\tau + K_{aw} \left(\lambda_{g,s} - \hat{\lambda}_{g,s} \right)^\tau \right) \tag{3}$$

where K_p, K_I, and K_{aw} are proportional, integral, and anti-wind-up gain, respectively.

PI controllers are tuned using the SIMC tuning method introduced by Skogestad (2003). The local controllers and the direct constraint controller have a sampling time of 1 sec. The central constraint controller updates the shadow price every 2.5 min because it may take more time to gather information from every cluster.

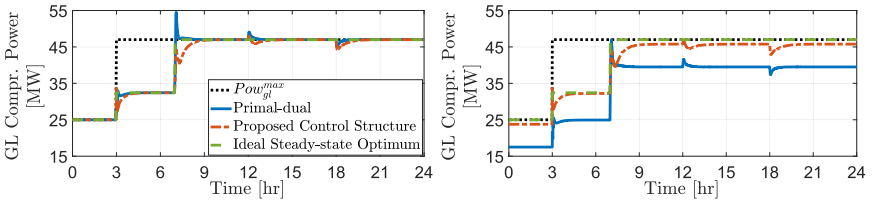


Figure 3: Left: Both Primal-dual and Proposed structure reach optimal steady-state conditions, but the constraint violation is much smaller for the proposed structure. Right: After implementing back-off from the power constraint.

Fig. 3 (left subplot) shows the simulation results when we consider $\zeta_{bo} = 0$. At time t = 3 hrs, the available power increases, and the shared constraint becomes inactive. Consequently, the gas-lift injection rates respond accordingly to achieve the optimal total available gas-lift allocation. Both primal-dual and proposed structure result in no dynamic violation at this time. At time t = 7 hrs, GOR dramatically decreases in all wells and causes extreme responses by the associated PI controllers. As a result, primal-dual significantly violates the constraint during the transient. We obtain a different result when applying the proposed control structure where Fig. 3 shows no dynamic constraint violation. At time t = 12 hrs, the GOR in most wells decreases, and the constraint is still active. The primal-dual has significant constraint violation during the transient for some time. As a comparison, the proposed structure responds accordingly and can reduce the magnitude and duration of that violation. At time t = 18 hrs, the GOR in most wells

increases, and the constraint is still active. Both methods have no dynamic violation at this time. In general, the proposed structure can reduce those dynamic violations (in constrained cases) because (conceptually) direct constraint control selects the calculated direct constraint control input instead of the indirect one, which is calculated based on suboptimal shadow price during the transient.

In terms of dynamic violation magnitude, primal-dual and proposed structure can reach 7.4955 MW and 1.0329 MW, respectively. *When the maximum available power is a hard constraint, the proposed structure outperforms primal-dual as it can reduce more 'required' back-off, ζ_{bo} or even eliminate it (see Fig.3).* Fig. 4 shows the profit obtained by both methods in this simulation. The result indicates that the proposed one can reduce profit loss as much as 22,207.00 price unit *(0.18 %)* in 24 hours when one implements a back-off strategy for the same case and duration.

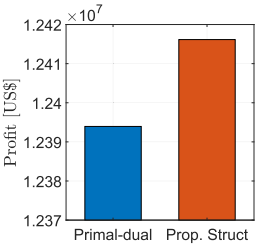


Figure 4: Profit $(-J)$ obtained by primal-dual and proposed structure in 24 hrs.

5. Conclusions

This work shows that the *proposed structure* with direct constraint control and primal-dual decomposition for optimization, is able to provide both tight constraint control on a fast timescale and optimal operation on a slow timescale. This strategy *offers* the possibility to reduce the back-off from constraints, which can give a large economic benefit.

6. Acknowledgment

The authors gratefully acknowledge the financial support from SUBPRO, financed by the Research Council of Norway, major industry partners, and NTNU.

References

R. Dirza, S. Skogestad, D. Krishnamoorthy, 2021. Optimal resource allocation using distributed feedback-based real-time optimization. IFAC-PapersOnLine 54 (3), 706–711, 16th IFAC Symposium on Advanced Control of Chemical Processes ADCHEM 2021.

I. J. Halvorsen, S. Skogestad, J. C. Morud, V. Alstad, 2003. Optimal selection of controlled variables. Industrial & Engineering Chemistry Research 42 (14), 3273–3284.

J. Jäschke, S. Skogestad, 2012. Optimal controlled variables for polynomial systems. Journal of Process Control 22 (1), 167–179.

D. Krishnamoorthy, 2021. A distributed feedback-based online process optimization framework for optimal resource sharing. Journal of Process Control (97), 72-83.

D. Krishnamoorthy, S. Skogestad, 2020. Systematic design of active constraint switching using selectors. Computers & Chemical Engineering 143, 107106.

M. Morari, Y. Arkun, G. Stephanopoulos, 1980. Studies in the synthesis of control structures for chemical processes: Part i: Formulation of the problem. process decomposition and the classification of the control tasks analysis of the optimizing control structures. AIChE Journal 26, 220 – 232.

S. Skogestad, 2000. Plantwide control: the search for the self-optimizing control structure. Journal of process control 10, 487 – 507.

S. Skogestad, 2003. Simple analytic rules for model reduction and pid controller tuning. Journal of Process Control 13, 291–309.

S.Wenzel, R. Paulen, G. Stojanovski, S. Kramer, B. Beisheim, S. Engell, 2016. Optimal resource allocation inindustrial complexes by distributed optimization and dynamic pricing. Automatisierungtechnik 64(6), 428–442.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50193-7

Data-driven coordination of expensive black-boxes

Damien van de Berg[a], Panagiotis Petsagkourakis[b], Nilay Shah[a], Ehecatl Antonio del Rio-Chanona[a,*]
[a]*Sargent Centre for Process Systems Engineering, Roderic Hill Building South Kensington Campus. London, SW7 2AZ, United Kingdom*
[b] *Sargent Centre for Process Systems Engineering (CPSE), Department of Chemical Engineering, University College London, Torrington Place, London, WC1E 7JE, United Kingdom*
a.del-rio-chanona@imperial.ac.uk

Abstract

Coordinating decision-making capacities using optimization is a key factor in the success of chemical companies. However, this coordination is often inhibited by expensive, legally-constrained, or proprietary subproblem models. We propose two variations on how model-based (surrogate) derivative-free optimization (DFO) methods can be used to coordinate subproblems with few connecting variables. When these surrogates are convex quadratic, they can be efficiently exploited using semidefinite programming techniques. We compare the performance of these two variations with a distributed optimization solver (ADMM), a model-based, and a direct DFO solver (Py-BOBYQA and DIRECT-L). This comparison is done on four variations of an economic-environmental feedstock blending optimization case study. While ADMM seems to display faster initial convergence, explorative DFO optimization solvers seem promising in escaping local minimizers, especially in lower dimensions.

Keywords: Expensive black-box; Surrogate optimization; Derivative-free optimization

1. Introduction

Model-based optimization of operations is key for chemical enterprises to remain competitive in an environment of increasingly complex economical, sustainability, and safety considerations. In enterprise-wide optimization, previously disconnected chemical engineering optimization models are integrated into a single model, wherein subproblems are coupled via few complicating variables and constraints. When the complicating, also called shared or global, variables are sparse compared to the number of local, or private, subproblem variables, these applications lend themselves well to distributed optimization and decomposition techniques (Tang and Daoutidis, 2019). Examples include the planning of supply chains or the operation of interconnected processing units where regional agents decide on the shared material streams that minimize a local, private cost.

Distributed optimization is a powerful tool that allows for the solution of large-scale nonlinear problems with significant computational savings using only limited information exchange. Augmented Lagrangian methods, such as the Alternating Direction Method of Multipliers (ADMM), have garnered special attention (Boyd et al., 2010). These methods iterate between a coordination step and the parallel solution of local subproblems. While these methods have proven convergence properties on convex problems, they lose convergence guarantees on nonconvex problems. Despite this, they can often be applied in practice (Rodriguez et al., 2021).

As such, Houska et al. (2016) have proposed an Augmented Lagrangian based algorithm for distributed nonconvex optimization (ALADIN) algorithm and conditions for convergence to local minimizers on nonconvex problems. ALADIN iterates between the parallel optimization of subproblems and a sequential quadratic programming (SQP) step for the coordination. This has been applied to nonconvex model predictive control and optimal power flow (Engelmann et al., 2020), as well as sensor localization problems (Houska et al., 2016).

Both algorithms display drawbacks that impede practical applicability: ADMM, as a subgradient method, requires many iterations to converge. ALADIN requires cheap gradient expressions and an approximation of the Hessian of the subproblems for the SQP step. Derivative information might not be available if the optimal solution of the subproblems requires 'expensive black-box' evaluations. This is the case in many process systems engineering (PSE) applications when the subproblems sample the output of proprietary simulation queries, or if expressions for the local objectives and constraints are not available for security, privacy, organizational, or other reasons: multiple business entities having to coordinate on the design of a supply chain while respecting local constraints and privacy; or in multi-objective optimization, where each objective is given by a different black-box simulation or optimization model.

Black boxes are often optimized using derivative-free optimization (DFO). DFO algorithms do not require gradient information and can be broadly classified into two classes (Larson et al., 2019): direct methods which approach the optimum by handling function evaluations directly; and model-based methods which rely on the intermediate construction of surrogates. In a coordination problem, we can then use DFO to find the shared variables that minimize the sum of local subproblem objectives. The key distinction between this DFO-based coordination and distributed optimization algorithms (like ADMM) lies in how the shared variables are updated in the coordination step.

In this work, we consider the coordination of black-box subproblems using no model information except for input-output data. We propose 1) the use of DFO algorithms for the coordination step instead of distributed optimization, and 2) two purely `data-driven' DFO strategies that rely on quadratic surrogates used within a trust region framework.. These methods are compared with ADMM, as well as a model-based, and direct DFO algorithm on all four combinations (lower- and higher-dimensional, convex and nonconvex) of a coordination problem consisting of an an economic blending and an environmental impact simulation subproblem.

2. Methodology

We are interested in solving problems of the following generic form:

$$\min_{\mathbf{z}\in \mathrm{Z},\, \boldsymbol{x}\in X} \sum_i f_i(\boldsymbol{x_i},\, \boldsymbol{z}) \tag{1}$$

where $\boldsymbol{z} \in R^{n_z}$ is the vector of global, shared, variables within the feasibility set Z, and $\boldsymbol{x} \in R^{n_x}$ is the vector of local, private variables within the feasibility set X. This can be reformulated as the following min-min problem:

$$\min_{\mathbf{z}\in\mathrm{Z}} \min_{\mathbf{x}\in\mathrm{X}} \sum_i f_i(\boldsymbol{x_i},\, \boldsymbol{z}) \tag{2}$$

After fixing $\boldsymbol{z}$, the problem becomes block separable, which makes the problem amenable to primal decomposition and distributed optimization. In fact, Eq. (2) is equivalent to the following constrained (bi-level) optimization problem:

$$\min_{\mathbf{z}\in\mathrm{Z}} \mathrm{F}(\boldsymbol{z}) \quad \text{subject to: } \mathrm{F}(\boldsymbol{z}) = \min_{\mathbf{x}\in\mathrm{X}} \sum_i f_i(\boldsymbol{x_i},\, \boldsymbol{z}) \tag{3}$$

Decomposition techniques rely on iteratively updating $\boldsymbol{x}$ and $\boldsymbol{z}$ with the other set of variables fixed (Gauss-Seidel sweeps) (Palomar and Chiang, 2006). In the Alternating Direction Method of Multipliers (ADMM) – a distributed optimization technique - local copies of $\boldsymbol{z}$ are introduced $\boldsymbol{y_i}$. These are then penalized in the objective of (2) as added Lagrangian and Augmented Lagrangian terms. This problem can then be solved iteratively in its consensus form:

$$\boldsymbol{x_i^{k+1}} = \underset{\mathbf{x_i}\in\mathrm{X_i},\ \boldsymbol{y_i}\in Z}{\operatorname{argmin}} f_i(\boldsymbol{x_i},\, \boldsymbol{y_i}) + \frac{\rho}{2}\left\|\boldsymbol{x_i} - \boldsymbol{z^k} + \boldsymbol{u_i^k}\right\|_2^2 \tag{4a}$$

$$\boldsymbol{u_i^{k+1}} = \boldsymbol{u_i^k} + \boldsymbol{x_i^{k+1}} - \boldsymbol{z^{k+1}} \tag{4a}$$

where $\boldsymbol{z^{k+1}}$ is the average of $\boldsymbol{x_i^{k+1}}$, and $\boldsymbol{u_i^k} \in R^{n_z}$ are the scaled dual variables of iteration k and subproblem i. This method is used as a base case to compare our proposed algorithms to. What we suggest is the use of derivative-free optimization (DFO) to solve for $\boldsymbol{z}$ that minimizes the objective $\mathrm{F}(\boldsymbol{z})$ in Eq. (3) where the latter is extended with an Augmented Lagrangian term to ensure convergence of the subproblem: when there is no feasible $\boldsymbol{x_i}$ for the proposed $\boldsymbol{z}$, the solution converges to the nearest feasible $\boldsymbol{x_i}$ and the associated deviation from $\boldsymbol{z}$ is penalized in the $\frac{\rho}{2}\|\boldsymbol{x_i} - \boldsymbol{z}\|_2^2$ term:

$$\mathrm{F_i}(\boldsymbol{z}) = \min_{\mathbf{x}\in\mathrm{X}} f_i(\boldsymbol{x_i},\, \boldsymbol{z}) + \frac{\rho}{2}\|\boldsymbol{x_i} - \boldsymbol{z}\|_2^2 \tag{5}$$

For this work, we choose Py-BOBYQA and DIRECT-L as promising model-based and direct DFO solvers respectively. The interested reader is referred to van de Berg et al. (2021) where we have described these solvers and benchmarked them on a handful of chemical engineering applications. We have also introduced CUATRO as a trust region-based convex quadratic surrogate DFO algorithm. There are two different ways that surrogate models (such as CUATRO) could be used for the derivative-free optimization of problem (3). In both ways, a set of shared variables is broadcast to the subproblems, where the associated local objective functions (Eq. (5)) are evaluated in private. All evaluations within the trust region are then used to fit a quadratic surrogate to the evaluations. The difference lies in whether one single surrogate is fitted over the sum of the subproblem evaluations as in Eq. (6a), or whether each subproblem fits its own surrogate Eq. (6b), whose sum is then coordinated in the objective. The approach used in Eq. (6a) is similar to (Li et al., 2021), and could be loosely referred to as a `Data-driven ADMM', while the other version, with one surrogate per subproblem, resembles a `Data-driven ALADIN'.

$$\min_{\mathbf{z}\in Z} \hat{\mathrm{F}}(\boldsymbol{z}) \quad \text{subject to: } \hat{\mathrm{F}}(\boldsymbol{z}) = \boldsymbol{z}^{\boldsymbol{T}} A\boldsymbol{z} + \boldsymbol{b}^{\boldsymbol{T}}\, \boldsymbol{z} + c \;\approx \sum_i F_i(\boldsymbol{z}) \tag{6a}$$

$$\min_{\boldsymbol{z}\in Z} \sum_i^N \hat{F}_i(\boldsymbol{z}) \quad \text{subject to: } \hat{F}_i(\boldsymbol{z}) = \boldsymbol{z}^{\boldsymbol{T}} A_i\boldsymbol{z} + \boldsymbol{b}_i^{\boldsymbol{T}}\, \boldsymbol{z} + c_i \;\approx F_i(\boldsymbol{z}) \tag{6b}$$

These surrogates are then optimized within a trust region, and the best iterate and trust region radius are updated as in van de Berg et al. (2021) for the CUATRO algorithm. When A, $\boldsymbol{b}$, and c are fitted to be convex, semidefinite programming can be leveraged for the training and exploitation of the surrogates. This way, the algorithm overhead remains tractable compared to the expensive subproblem evaluations.

Our synthetic case study is inspired by an industrial problem where the composition of chemical feedstock needs to be optimized based on cost and environmental impact. The decision variables represent feedstock composition. The solution also needs to satisfy product performance constraints. The practitioners however only have access to the two separate problems as the output of expensive, proprietary black-boxes, preventing the two models from being integrated and exploited using conventional mixed-integer solvers. We present here a simplified case study involving 10 or 28 decision feedstocks as the shared variables **z**. The economic optimization problem takes the following form:

$$\min_{\mathbf{y}\in\{0,1\}^{n_x};\, \mathbf{z},\boldsymbol{x}\in R^{n_x}} \sum_i (x_i Cost_i + \rho\,(x_i - z_i)^2) \tag{7a}$$

$$s.t. \quad x_i \le y_i, \quad \sum_i y_i \le N_{integer}\,, \quad \sum_i x_i = 1, \quad L_i \le \sum_i x_i\, A_i \le U_i \tag{7b) - (7e)}$$

where the objective function (Eq. (7a)) constitutes of linear cost terms and an Augmented Lagrangian term to ensure convergence of the subproblem for all proposed **z**. Eq. (7e) includes lower and upper bounds on blending quality; Eq. (7d) ensures that the total composition adds up to one; Eq. (7b) and (7c) ensure that only $N_{integer}$ variables are non-zero if there is a cap on the number of feeds in the mix. The latter constraints can be relaxed to make the problem convex. The environmental impact subproblem involves a simulation whose output is the result of linear and square terms in the composition variables with the optional addition of bilinear terms to introduce nonconvexity. The two objective evaluations are then scalarized into a single objective using a scalarization factor for the environmental impact output. We take ρ to be $1 \cdot 10^6$ and start from an initial guess where all components weigh the same, 0.0357 and 0.1 respectively.

The next section compares the convergence of the two versions of DFO solvers CUATRO, Py-BOBYQA, and DIRECT-L with ADMM in its consensus form. Each comparison is made on all four combinations of lower- and higher-dimensional as well as convex and nonconvex versions of the coordination problem.

3. Results and Discussion

Figure 1 shows the convergence of the best function evaluation with respect to the number of function evaluations of all methods on the four variations of the coordination problem. ADMM stands out as the method that makes the most progress in the least number of evaluations. Within ten evaluations, ADMM manages to approach a solution that is feasible (to an acceptable numerical tolerance) and near-optimal in all four cases. The next-best performing method, Py-BOBYQA, takes around 50 evaluations to get as close

in the convex lower-dimensional case, and around 200 in the higher-dimensional nonconvex case. Otherwise, the final convergence of Py-BOBYQA is close to that of ADMM in the lower-dimensional case. It is slightly worse in the higher-dimensional convex, and slightly better in the higher-dimensional nonconvex case. It is expected that these methods perform better in the first evaluations as they are exploitative without explicit exploration.

In the lower-dimensional convex case, DIRECT-L and CUATRO with subproblem surrogates converge to the same optimum as ADMM and Py-BOBYQA, but require more evaluations. In the two higher-dimensional cases, DIRECT-L converges to a slightly worse optimum than ADMM and Py-BOBYQA in significantly more evaluations. For most cases, CUATRO constructing subproblem surrogates performs better than its single `coordinated surrogate' counterpart. However, the lower-dimensional nonconvex case stands out, as both DIRECT-L and CUATRO employing a single coordinated surrogate perform best, converging to a better optimum than both Py-BOBYQA and ADMM.

This observation attempts to answer the central question of this work. Generally. while subgradient methods (ADMM) are slow to converge in terms of iterations, in higher dimensions, subgradient information still leads to quicker convergence than using data on its own. However, the lower-dimensional nonconvex case study suggests that derivative-free optimization solvers could shine for nonconvex applications with very few decision variables, where exploration is encouraged, and additional evaluations required to escape local minimizers. While the discussion of the relative performance of DFO algorithms is in line with that of van de Berg et al. (2021), we are not aware of any previous literature that directly compares DFO and distributed optimization coordination approaches. Additionally, since CUATRO has the capacity to explicitly handle black-box

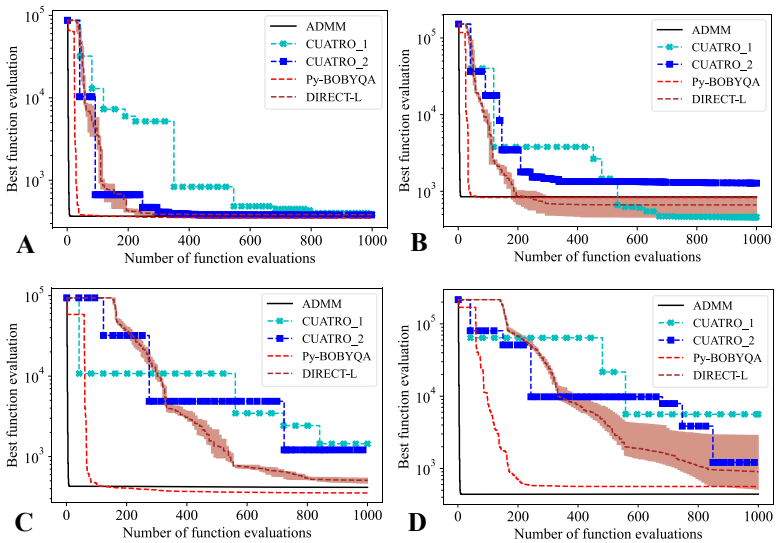


Figure 1. Convergence plots: Best function evaluation versus number of evaluations on the 10-d convex (A), 10-d nonconvex (B), 28-d convex (C), and 28-d nonconvex (D) case study using CUATRO_1 with a single coordinated surrogate (cyan dotted lines with X markers), CUATRO_2 with local subproblem surrogates over each subproblem (dark blue dotted lines with squares), ADMM (solid black line), Py-BOBYQA (dotted orange line), and DIRECT-L (dotted red line with shaded min-max range over 10 runs)

constraints, CUATRO could become useful for the navigation of more complex private constraints, e.g. when the subproblems can only access binary constraint information – is this set of global variables feasible or not? Finally, CUATRO, as a regression- rather than interpolation-model (such as Py-BOBYQA), could lead to better convergence when the subproblems involve stochasticity.

4. Conclusion

We have proposed alternatives to using distributed optimization in the case where subproblems need to be coordinated using input-output data as the only accessible information from the subproblem. It is possible to use DFO solvers that converge to the same optimum as the distributed optimization solver ADMM. However, only under specific conditions is the use of DFO solvers encouraged over that of ADMM, namely when the subproblems are highly nonconvex and connected by few shared variables. As for the DFO solvers, Py-BOBYQA displays the quickest and most reliable convergence, but as an exploitative method also regularly gets stuck in the same local minima as ADMM. DIRECT-L usually displays slow but still consistent convergence, escaping many local minima due to its partition-based nature. Finally, CUATRO should not be dismissed as a potential solver for more complex coordination problems involving stochasticity, and black-box constraints.

Acknowledgements

D.v.d.B. gratefully acknowledges financial support from the Bansal bursary. D.v.d.B. would also like to thank all BASF S.E. collaborators, especially his co-supervisors Dr. Debora Morgenstern and Dr. Inga-Lena Darkow, scientific exchange coordinator Dr. Christian Holtze, and Dr. Sangbum Lee for their invaluable support in conceptualizing this work.

References

S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, 2010, Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. Trends R Mach. Learn. 3, 1–122

A. Engelmann, Y. Jiang, B. Houska, T. Faulwasser, 2020, Decomposition of Nonconvex Optimization via Bi-Level Distributed ALADIN. IEEE Trans. Control Netw. Syst. 7, 1848–1858

B. Houska, J. Frasch, M. Diehl, 2016, An augmented Lagrangian based algorithm for distributed nonconvex optimization. SIAM J. Optim.

J. Larson, M. Menickelly, S. Wild, 2019, Derivative-free optimization methods. Acta Numerica, 28, 287-404

Z. Li, Z. Dong, Z. Liang, Z. Ding, 2021, Surrogate-based distributed optimisation for expensive black-box functions. Automatica 125, 109407

D. P. Palomar, M. Chiang, 2006, A tutorial on decomposition methods for network utility maximization, IEEE Journal on Selected Areas in Communications, 24, 8, 1439-1451

J. S. Rodriguez, B. Nicholson, C. Laird, V. M. Zavala, 2018, Benchmarking ADMM in nonconvex NLPs. *Comput. Chem. Eng.* 119, 315–325

W. Tang, P. Daoutidis, 2019, Distributed control and optimization of process system networks: A review and perspective. Chinese J. Chem. Eng. 27, 1461–1473

D. van de Berg, T. Savage, P. Petsagkourakis, D. Zhang, N. Shah, E.A. del Rio-Chanona, 2021, Data-driven optimization for process systems engineering applications, Chemical Engineering Science, 117135

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50194-9

Data-Driven Adaptive Robust Unit Commitment Assisted by Machine Learning Techniques

Ning Zhao[a*], Fengqi You[a]

[a]*Cornell University, Ithaca, New York, 14853, USA*
nz225@cornell.edu

Abstract

In this paper, we propose a novel robust unit commitment (UC) framework with data-driven disjunctive uncertainty sets for volatile wind power outputs, assisted by machine learning techniques. To flexibly identify the uncertainty space for wind power forecast error data with disjunctive structures, the uncertainty data are grouped using K-means and density-based spatial clustering of applications with noise. The disjunctive uncertainty sets are constructed accordingly as the union of multiple basic uncertainty sets, including conventional uncertainty sets, and data-driven uncertainty sets using Dirichlet process mixture model, principal component analysis coupled with kernel density estimation, and support vector clustering. The problem is formulated into a two-stage robust UC model with data-driven disjunctive uncertainty sets and with a multi-level optimization structure. To facilitate the solution process, a decomposition-based optimization algorithm is developed. The effectiveness of the proposed framework is illustrated using a case study based on the IEEE 39-bus system.

Keywords: Unit commitment, robust optimization, machine learning, disjunction.

1. Introduction

The United States is planning to use wind energy to provide 20% electricity in 2030 and 35% in 2050. Considering the intermittent nature of wind power (Ning and You, 2022), including the forecast uncertainties in the unit commitment (UC) decision-making process can help ensure power systems reliability and reduce the economic costs (Padhy, 2004). Robust UC has gained attention in recent years, owing to its robustness, flexible conservatism control, computational efficiency (Qiu et al., 2020), and effective utilization of large-scale uncertainty data (Ning and You, 2019). The robust UC problem is generally formulated into a two-stage adaptive robust optimization (ARO) model. The first stage determines the on-off decisions, and the second stage provides power output and dispatch decisions according to the worst case (Bertsimas et al., 2013). Note that the wind power forecast errors can have complex and disjunctive data structures, and studies found that it could be more appropriate to depict such uncertainty using more general distributions, such as the Gaussian mixture model (Wang et al., 2017). The finding indicates that the traditional "one-set-fits-all" approach for constructing a single uncertainty set based on all uncertainty data may not flexibly and accurately capture the uncertainty space. From a machine learning perspective, clustering has been a useful tool in detecting the disjunctive structure of a data set. To the best of our knowledge, there is no existing literature on robust optimization-based UC that systematically and effectively captures the uncertainty space based on uncertainty data with disjunctive structures by integrating clustering techniques with the ARO models. To fill the knowledge gap, this study aims to propose a novel two-stage

adaptive robust UC framework that has the potential of leading to better optimization solutions by incorporating data-driven disjunctive uncertainty sets to deal with the uncertain wind power forecast error data, which are independent and identically distributed (Bludszuweit et al., 2008) and have disjunctive structures. A tailored decomposition-based optimization algorithm is developed to iteratively solve the resulting robust UC problems. To illustrate the effectiveness of the proposed framework, an application on the IEEE 39-bus system is presented.

2. Data-driven robust unit commitment with disjunctive uncertainty sets

A two-stage ARO framework with data-driven disjunctive uncertainty sets is proposed in this work for UC with uncertain wind power forecast errors, as shown in Figure 1. First, the optimal number of uncertainty data clusters is determined using the Calinski-Harabasz index. Next, the uncertainty data are clustered according to the optimal number of clusters using machine learning techniques. After integrating the uncertainty data with the clustering results, the data-driven disjunctive uncertainty sets are then constructed. Subsequently, the ARO problem is formulated by incorporating the data-driven disjunctive uncertainty sets. Lastly, the proposed ARO problem is solved iteratively using a tailored decomposition-based optimization algorithm.

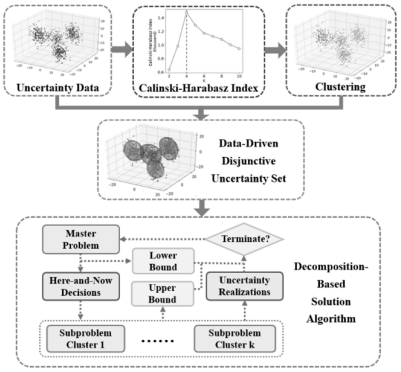


Figure 1. Flowchart of the proposed ARO framework with disjunctive uncertainty sets.

The objective function is minimizing the total UC cost under uncertainty. Note that the optimization model has a two-stage structure. Specifically, the "here-and-now" decisions for the first stage represent the commitments of generators that are determined 24 hours ahead, including online status, start-up status, and shutdown status of generators. The "wait-and-see" decisions determined after the uncertainty is realized in the second stage depict the economic dispatch process, including the power output of generators, power dispatch of a wind farm, and slack variables for balance constraints. The constraints of the robust UC model include the logic relations of generators, the minimum uptime and downtime of generators, the minimum and maximum power outputs of a generator, the ramping rates, the energy balance of the system, the capacities of transmission lines, the maximum outputs from wind farms that equal to the summation of forecasted wind power generation and the uncertain forecast errors, the initial commitment status and power outputs of generators, and the feasible region of decision variables. The maximum output constraint for wind farms holds for all potential realizations of the uncertainty, which guarantees the robustness of solutions.

A novel approach for developing data-driven disjunctive uncertainty sets is developed in this study. The conventional robust optimization adopts the "one-set-fits-all" approach that constructs a single uncertainty set to depict the uncertainty space. In contrast, the disjunctive uncertainty sets consist of multiple basic uncertainty sets (Ning and You, 2018), and the uncertainty space is represented by the union of these uncertainty sets, as shown in Eq.(1). L is the set of uncertainty data clusters.

$$\mathbf{U} = \bigcup_{l} \mathbf{U}_l \tag{1}$$

To determine the optimal number of clusters, the Calinski-Harabasz index is calculated, which finds the optimal number of clusters by pursuing a balance between the compactness of a data group and the distribution of all data groups. Specifically, the uncertainty data are first grouped using K-means for a range of k, and the one with the highest Calinski-Harabasz index represents the optimal number of clusters. For data clustering, two machine learning techniques, namely K-means and density-based spatial clustering of applications with noise (DBSCAN), are used to group the uncertainty data according to the optimal number of clusters. K-means is an unsupervised, non-deterministic, iterative, and centroid-based clustering algorithm, while DBSCAN is a density-based clustering algorithm.

Based on the resulting uncertainty data clusters, five types of basic uncertainty set can be constructed, including the traditional box and budget uncertainty sets, and three types of data-driven uncertainty sets using a variational inference algorithm for the Dirichlet process mixture model (DPMM) (Ning and You, 2017), principal component analysis (PCA) coupled with kernel density estimation (KDE) (Ning and You, 2018), and support vector clustering (SVC) (Shang and You, 2017, 2019). These data-driven uncertainty sets are widely adopted in the robust optimization community because of their remarkable performances in capturing the uncertainty space (Ning and You, 2019).

The two-stage robust UC with the proposed disjunctive uncertainty sets have an objective function with a multi-level structure, semi-infinite constraints, and non-convex uncertainty sets. Therefore, to facilitate the solution process, a decomposition-based optimization algorithm is developed and applied. Specifically, a master problem and a set of subproblems are iteratively solved in the decomposition-based optimization algorithm. The master problem optimizes the UC decisions under multiple optimality cuts that correspond to a partial enumeration of the extreme points of the basic uncertainty sets and provides a lower bound to the original robust UC problem. Subsequently, we fix the values of first-stage decision variables following the optimal solutions of the master problem and develop a set of subproblems to investigate economic dispatch under the worst case. Note that each basic uncertainty set $\boldsymbol{U}_l$ corresponds to an individual subproblem. To reformulate the subproblem into a single-level maximization problem that can be solved directly by off-the-shelf solvers, the classical Karush-Kuhn-Tucker (KKT) conditions and the big-M method are used. The solutions of subproblems provide upper bounds of the original problem. A set of additional optimality cuts is then generated based on the uncertainty realization in the worst case and is updated in the master problem for the next iteration. Eventually, the algorithm terminates when the relative optimality gap is below the tolerance level ξ. Furthermore, to improve the computational efficiency, the master problem and subproblem are constructed only once during the solution process, and a part of the variables and constraints are updated in the following iterations.

3. Case study based on IEEE 39-bus systems

To illustrate the effectiveness of the proposed framework with disjunctive uncertainty sets, a case study based on the IEEE 39-bus system is investigated. The system includes 39 buses, 10 generators, 46 lines, and 3 wind farms. 800 uncertainty data samples are generated from a Gaussian mixture model. The two-stage robust UC problem is coded using Pyomo in Python on a PC with an Intel i7-8700 @ 3.20 GHz CPU and 32GB RAM, running on a 64-bit Windows 10 Enterprise operating system. The reformulated master problems and subproblems are solved using Gurobi 9.1.

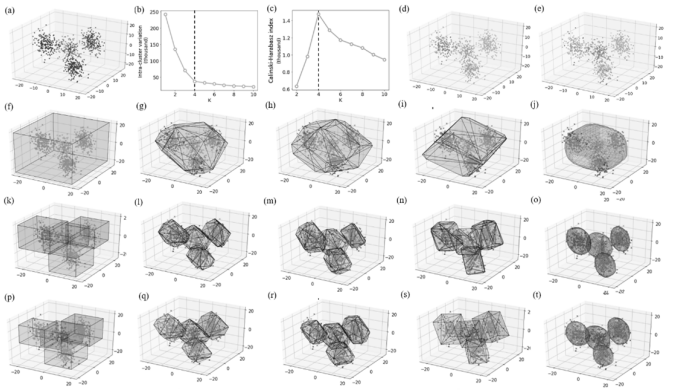


Figure 2. Uncertainty data, determination of optimal group number, clustering results, and uncertainty sets for application on IEEE 39-bus system. (a) Uncertainty data samples. (b) Elbow method. (c) Calinski-Harabasz index. (d) K-means results. (e) DBSCAN results. (f) Conventional box. (g) Conventional budget. (h) Conventional DPMM. (i) Conventional PCA & KDE. (j) Conventional SVC. (k) K-means + box. (l) K-means + budget. (m) K-means + DPMM. (n) K-means + PCA & KDE. (o) K-means + SVC. (p) DBSCAN + box. (q) DBSCAN + budget. (r) DBSCAN + DPMM. (s) DBSCAN + PCA & KDE. (t) DBSCAN + SVC.

Figure 2 presents the uncertainty data, the data clustering results, and the uncertainty sets. Wind forecast error data are shown in Figure 2(a). According to the elbow method and the Calinski-Harabasz index method in Figure 2(b)-(c), the optimal number of data groups is 4. Next, K-means and DBSCAN are applied individually to group the data into 4 data groups in Figure 2(d)-(e). Conventional and disjunctive uncertainty sets are shown in Figure 2(f)-(t). All types of data-driven uncertainty sets have the same level of conservativeness, which can be chosen by the decision makers to balance risk and robustness. The level of conservativeness is set to be 90%, represented by the same data coverage level of 90% across all types of data-driven uncertainty sets. The UC problem is solved using the proposed decomposition-based optimization algorithm. The solution time for the conventional uncertainty sets without clustering ranges from 46 to 182 seconds, and the problems with the proposed disjunctive uncertainty sets take 60-85 seconds to solve. Compared to the conventional approach that updates the entire models during the iterative solution process, the proposed algorithm can significantly improve the computational efficiency by reducing the solution time by around 70%. The optimal objective values, namely the minimum UC costs, are listed in Table 1. Higher values correspond to more conservative solutions. For comparison, the minimum cost for the deterministic case with no uncertainties is $426,443. To investigate the effectiveness of the proposed approach, we apply the price of robustness (PoR) to measure the level of additional cost for the robust optimization cases compared to the deterministic case. The

proposed approach reduces PoR by 28-38%, 7.6%, 23%-29%, 27%-31%, and 21%-22% for the problems with box, budget, DPMM, PCA coupled with KDE, and SVC uncertainty sets, respectively, compared to the conventional approach. Also, the problems with disjunctive uncertainty sets constructed using DBSCAN tend to have lower optimal costs than the problems with K-means-based disjunctive uncertainty sets, showing that DBSCAN may handle the outliers of the uncertainty data more efficiently.

Table 1. Optimal objective values under uncertainty sets with different approaches.

minimum cost ($)	Box	Budget	DPMM	PCA & KDE	SVC
No Clustering	461,020	441,155	445,741	455,089	444,191
K-Means	451,178	440,041	441,277	447,271	440,508
DBSCAN	447,712	440,041	440,092	446,121	440,217

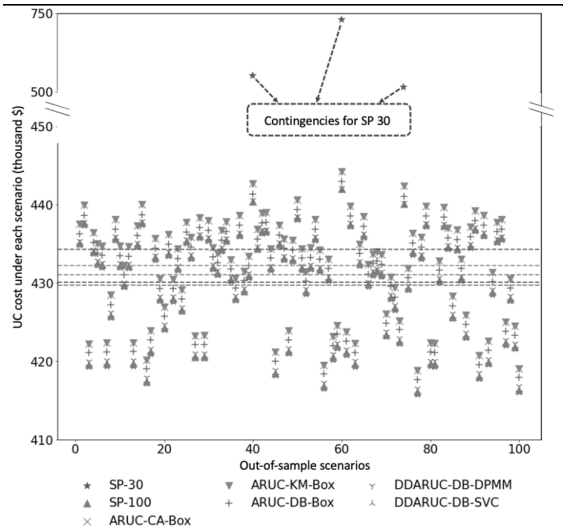


Figure 3. Simulation results under the optimal solutions from ARO and SP approaches.

To benchmark the performance of the proposed approach, we simulate the UC costs under the optimal solutions from both the ARO models and the conventional two-stage stochastic UC models. We obtain the optimal stochastic UC solutions through 30 and 100 random scenarios, denoted as SP-30 and SP-100 solutions, respectively. The stochastic UC problems tend to be more computationally demanding, as obtaining SP-30 and SP-100 solutions take 342 CPUs and 1,194 CPUs, which are more than three times longer than the solution time using the proposed framework with disjunctive uncertainty sets. Figure 3 presents the simulated UC costs following ARO and SP solutions throughout 100 randomly generated out-of-sample scenarios, where the horizontal lines indicate the average out-of-sample costs. The SP-30 solution cannot handle the systems contingencies effectively, as shown by two scenarios with UC costs over $500,000. As for the proposed approach, the solution corresponding to the K-means-based disjunctive box uncertainty sets shows a noticeably lower average UC cost than the solution using the conventional box uncertainty set. The differences of average simulated UC costs between the SP-100 solution and the robust UC solutions using the

proposed data-driven disjunctive uncertainty sets are below 0.001%, while the solution time of SP-100 is 13 times longer than the proposed approach.

4. Conclusions

This paper proposed a novel robust UC framework with data-driven disjunctive uncertainty sets for the wind power forecast errors with disjunctive structures. The uncertainty data were grouped using K-means and DBSCAN, and the proposed disjunctive uncertainty sets were constructed accordingly as the union of multiple basic uncertainty sets, including conventional uncertainty sets and data-driven uncertainty sets using DPMM, PCA coupled with KDE, and SVC. The resulting problem was formulated as a two-stage robust UC model with disjunctive uncertainty sets. To facilitate the solution process, a tailored decomposition-based optimization algorithm was developed. A case study based on the IEEE 39-bus system was performed. The results presented that the price of robustness reduced by 8-48% with the proposed framework, compared to the traditional "one-set-fits-all" approach.

References

D. Bertsimas, E. Litvinov, X. A. Sun, et al., 2013, Adaptive Robust Optimization for the Security Constrained Unit Commitment Problem. IEEE Transactions on Power Systems, 28, 1, 52-63.

H. Bludszuweit, J. A. Dominguez-Navarro, A. Llombart, 2008, Statistical Analysis of Wind Power Forecast Error. IEEE Transactions on Power Systems, 23, 3, 983-991.

C. Ning, F. You, 2017, Data-driven adaptive nested robust optimization: General modeling framework and efficient computational algorithm for decision making under uncertainty. AIChE Journal, 63, 9, 3790-3817.

C. Ning, F. You, 2017, A data-driven multistage adaptive robust optimization framework for planning and scheduling under uncertainty. AIChE Journal, 63, 4343-4369.

C. Ning, F. You, 2018, Data-driven decision making under uncertainty integrating robust optimization with principal component analysis and kernel smoothing methods. Computers & Chemical Engineering, 112, 190-210.s

C. Ning, F. You, 2018, Data-driven stochastic robust optimization: General computational framework and algorithm leveraging machine learning for optimization under uncertainty in the big data era. Computers & Chemical Engineering, 111, 115-133.

C. Ning, F. You, 2019, Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming. Computers & Chemical Engineering, 125, 434-448.

C. Ning, F. You, 2019, Data-Driven Adaptive Robust Unit Commitment Under Wind Power Uncertainty: A Bayesian Nonparametric Approach. IEEE Transactions on Power Systems, 34, 2409-2418.

C. Ning, F. You, 2022, Deep Learning Based Distributionally Robust Joint Chance Constrained Economic Dispatch Under Wind Power Uncertainty. IEEE Transactions on Power Systems, 37, 191-203.

N.P. Padhy, 2004, Unit commitment-a bibliographical survey. IEEE Transactions on Power Systems, 19, 1196-1205.

H. Qiu, F. You, 2020, Decentralized-distributed robust electric power scheduling for multi-microgrid systems. Applied Energy, 269, 115146.

C. Shang, X. Huang, F. You, 2017, Data-driven robust optimization based on kernel learning. Computers & Chemical Engineering, 106, 464-479.

C. Shang, F. You, 2019, A data-driven robust optimization approach to scenario-based stochastic model predictive control. Journal of Process Control, 75, 24-39.

Y. Wang, Q. Hu, D. Meng, P. Zhu, 2017, Deterministic and probabilistic wind power forecasting using a variational Bayesian-based adaptive robust multi-kernel regression model. Applied Energy, 208, 1097-1112.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50195-0

Heat integration for superstructure models: A MILP formulation for easy implementation and fast computing

Philipp Kenkel [a,b*], Timo Wassermann [a,b], Edwin Zondervan [c]

[a] *Advanced Energy Systems Institute, University of Bremen, 28359 Bremen, Germany*
[b] *artec Sustainability Research Center, University of Bremen, 28359 Bremen, Germany*
[c] *SPT-PSE,University of Twente, 7522NB Enschede, Netherlands*
kenkel@uni-bremen.de

Abstract

This work presents a mixed-integer linear programming (MILP) heat integration model tailored to superstructure optimization. It includes energy targeting based on a transshipment model combined with linearized heat exchanger capital costs. A variable number of heat utilities at different temperature levels can be implemented. In addition, the model facilitates the utilization of high temperature heat pumps to benefit from low exergy waste heat. The heat integration model is part of the **O**pen s**U**perstruc**T**ure mo**D**eling and **O**ptimizati**O**n f**R**amework (OUTDOOR) and thus can be accessed easily using its intuitive excel-based interface. A model evaluation of the MILP shows low cost deviations of 1–14 % compared to more complex models, with fast solution times. Additionally, a practical superstructure case study is presented, where internal heat recovery reduces the external heat consumption of a power-to-methanol process by 40 %, thus underlining the relevance for adequate consideration.

Keywords: Superstructure optimization, Heat integration, MILP, Open-source

1. Introduction

Heat integration and the optimization of heat exchanger networks (HEN) is an important domain in chemical engineering and process synthesis. It includes three major steps: 1) minimization of external utility demand, 2) minimization of heat exchanger (HEX) area and 3) minimization of the HEX matches (Yee and Grossmann, 1990). Different approaches, from simple pinch analysis to complete simultaneous optimization of all three tasks exist, the latter resulting in complex MINLP's (**M**ixed-**I**nteger **N**on-**L**inear **P**rogramming models). Renowned MINLP approaches were developed by Yee and Grossmann as well as by Ciric and Floudas (Ciric and Floudas, 1991; Yee and Grossmann, 1990). However, the integration of MINLP's in tools for preliminary process design, like superstructure models, tends to increase their complexity. A superstructure model is a simplified representation of many possible flowsheets that is used for process synthesis (Kenkel et al., 2021a). Solving MINLP superstructures is generally much slower if not even impossible for large models. Thus, heat integration is often omitted, which can lead to suboptimal solutions if flowsheets include considerable heat integration potential. This work presents a novel MILP heat integration model tailored to superstructure models. It is implemented in the **O**pen s**U**perstruc**T**ure mo**D**eling and **O**ptimizati**O**n f**R**amework (OUTDOOR) and provides fast solution times, easy code implementation and data input together with reasonable accuracy (Kenkel et al., 2021b).

2. Methodology

2.1. Energy targeting model

The heat integration model is part of OUTDOOR's MILP superstructure model for process synthesis which is described in detail by Kenkel et al. (Kenkel et al., 2021a). OUTDOOR can be used to derive optimal flowsheets based on interaction of different unit-operations using mass and energy balances as well as cost and emission functions (Kenkel et al., 2021b, 2021a). The underlying energy targeting model is based on a transshipment model and depicted in Figure 1. Prior to optimization, the inlet and outlet temperatures of unit-operations, external heating and cooling utilities as well as heat pump (HP) inlet and outlet temperatures are defined. Those temperatures form a temperature grid. To account for a minimum temperature difference of $\Delta T_{min} = 10$ K within the temperature grid, the inlet and outlet temperatures of the hot/cold units are reduced/increased by 5 K. The required heating and cooling demands of the individual unit-operations are defined by the calculated (case specific) mass flow rates (in t/h) and a predefined specific energy demand (in MWh/t_{input}). These are partitioned into the different temperature intervals on the temperature grid. Heat can be exchanged in the different temperature intervals. Residual heat at high temperatures can be cascaded down to lower intervals. If a heat deficit persists, it has to be satisfied by external heat. Residual heat in the lowest temperature interval is cooled down by the cooling water utility. For every heat interval that exchanges heat, a virtual HEX is calculated, which is used for the cost calculation of the virtual HEN. The HP can raise low temperature heat to higher temperature intervals based on predefined HP inlet and outlet temperature and coefficient of performance (COP).

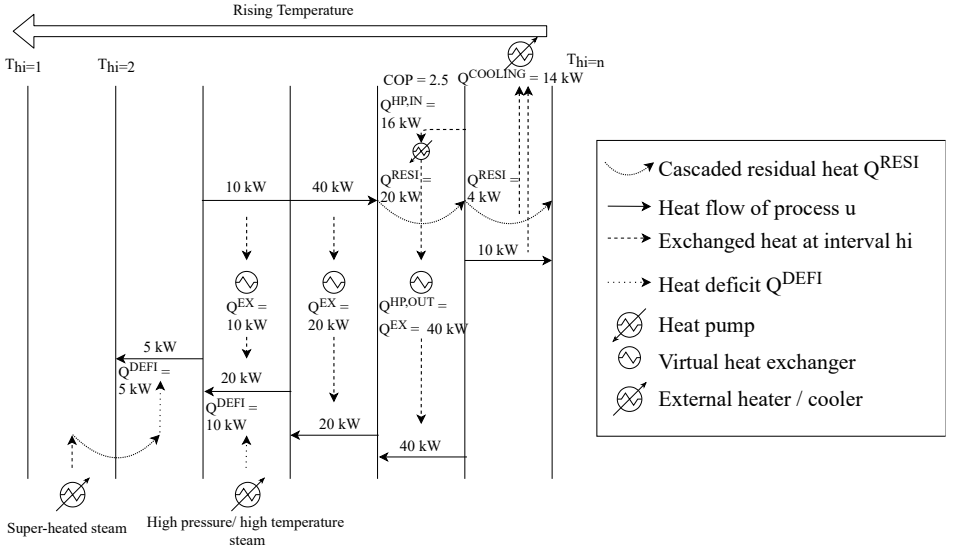


Figure 1: Representation of OUTDOOR's energy targeting model.

2.2. Heat integration costs

The heat integration costs are calculated on an annual basis using the total external utility demand and capital costs for a virtual HEX depicting the HEN. Additional costs are added for the HP if it is utilized. Heating costs are calculated based on derived heat deficits in the different temperature intervals; cooling costs are derived from the residual heat in the

lowest temperature interval. Annual capital costs of virtual HEX are determined by a linear cost function presented in Eqs. (1) – (3). Here, Q_{hi}^{EX} depicts the exchanged heat flow in temperature interval *hi*, M_{hi} represents a large Big-M parameter and Y_{hi} a binary variable that defines if heat is exchanged in temperature interval *hi*. *m* and *b* are linear coefficients which are derived from several pre-simulated, sized and costed HEX using Aspen Plus. Their default values are m = 13.459 k€ MW^{-1} yr^{-1} and b = 3.3893 k€ yr^{-1}. The costs of the HP are derived from its electricity demand, which depends on the defined COP and utilized heat, plus the capital costs, which result from a linear correlation with the heat output.

$$C\_HEX_{hi} \leq m \cdot Q_{hi}^{EX} + b + M_{hi} \cdot (1 - Y_{hi}) \quad (1)$$

$$C\_HEX_{hi} \geq m \cdot Q_{hi}^{EX} + b - M_{hi} \cdot (1 - Y_{hi}) \quad (2)$$

$$C\_HEX_{hi} \leq M_{hi} \cdot Y_{hi} \quad (3)$$

3. Model evaluation

Two performance tests are performed for the MILP using well-known cases provided by Linnhoff et al. and Floudas and Grossmann (Floudas et al., 1986; Linnhoff et al., 1982).

3.1. Performance test cases

The first performance test is a simple HEN optimization given by Linnhoff et al. including two hot/cold streams together with external utilities. (Linnhoff et al., 1982). It was tested by Linnhoff et al., by usage of the MAGNETS software and the rigorous MINLP by Yee and Grossmann. (Floudas et al., 1986; Linnhoff et al., 1982; Yee and Grossmann, 1990). The second test is more complex, considering five hot streams and one cold stream. This case is an example from the MAGNETS user manual also tested by Yee and Grossmann as well as Ciric and Floudas, among others (Ciric and Floudas, 1991; Floudas et al., 1986; Yee and Grossmann, 1990). Initial data for the tests is given in Table 1 and 2.

Table 1: Initial data for performance test 1 taken from Linnhoff et al. (Linnhoff et al., 1982)

Stream name	T_{IN} (°C)	T_{OUT} (°C)	F_{cp} (kW K^{-1})	Costs ($ kW^{-1} yr^{-1})
H_1	169.85	59.85	30	–
H_2	149.85	29.85	15	–
C_1	19.85	134.15	20	–
C_2	79.85	139.85	40	–
Steam	176.85	176.85	–	80
Cooling water	19.85	39.85	–	20

Table 2: Initial data for performance test 2 taken from MAGNETS user manual (Floudas et al., 1986)

Stream name	T_{IN} (°C)	T_{OUT} (°C)	F_{cp} (kW K^{-1})	Costs ($ kW^{-1} yr^{-1})
H_1	226.85	46.85	6	–
H_2	206.85	106.85	4	–
H_3	186.85	86.85	6	–
H_4	106.85	86.85	20	–
H_5	106.85	46.85	12	–
C_1	16.85	386.85	18	–
Steam	426.85	426.85	–	140
Cooling water	26.85	46.85	–	10

3.2. Results

The implementation of the stream data in OUTDOOR only takes ca. 5 minutes due to its ready-made excel-templates. To enable a fair cost comparison, the virtual HEX capital costs are first adjusted to the according reference year by applying the chemical engineering cost plant index (CECPI) and further transformed from € to $.
The total run time of the OUTDOOR program for performance test 1 is 1.33 s of which 0.33 s are solver time by Gurobi. The calculated total annual HEN costs are 102,123 $ which is about 14 % higher than the results from Linnhoff et al., the MAGNETS solution and one case of Yee's and Grossmann's MINLP (Yee and Grossmann, 1990) (ref. Figure 2). Annual external utility costs are equal to Linnhoff's and the MAGNETS solution, whilst annual capital costs are about 20–25 % higher (Yee and Grossmann, 1990).
Total run time of the OUTDOOR program for performance test 2 is 1.96 s of which 0.72 s are solver time by Gurobi. For this case the annual HEN costs are 569,651 $ which is only about 1 % smaller than the reported solutions in literature (Ciric and Floudas, 1991; Yee and Grossmann, 1990). The total heat exchange of 3040 kW is in the range of solutions recorded by Yee (2984 kW) and Ciric (3068 kW) and depicts the energy target for a minimum approach temperature of $\Delta T_{min} = 10$ K (Ciric and Floudas, 1991; Yee and Grossmann, 1990). The two examples show that the computing time of the proposed approach is very fast, while the energy targeting along with the HEN costs provide acceptable accuracy. It is further expected, that the total cost deviation would be lower in a real process synthesis case, where HEN costs only make a minor share of the overall costs.

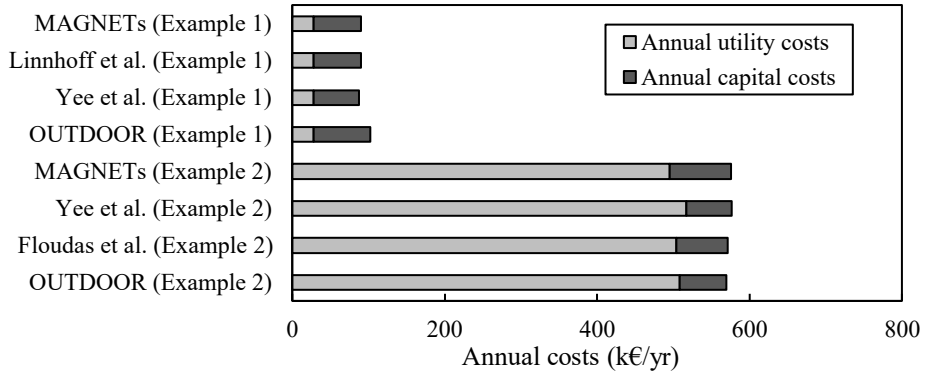


Figure 2: Cost results from performance tests.

4. Applied case study

To demonstrate the capability of the MILP to a real-world problem a case study is developed. The selected case study describes a power-to-methanol (PtM) process which was introduced in Kenkel et al. (Kenkel et al., 2021a).

4.1. Power-to-Methanol application

The superstructure optimization aims to synthesize the flowsheet of a methanol production by direct hydrogenation of CO_2. As depicted in Figure 3, different water electrolysis technologies are considered for H_2 production: Low- and high-pressure alkaline and proton exchange membrane electrolysis as well solid oxide electrolysis. Available CO_2 sources are ambient air and flue gases from an oil refinery or oxyfuel fired

cement factory. Methanol synthesis is included as a tripartite process (reaction, purge separation, purification) based on data from Wassermann et al. (Wassermann et al., 2020). The purge stream is combusted to produce either steam or electricity, which can be used internally or sold. For a detailed description of the superstructure, as well as bi-criteria flowsheet optimization we refer to Kenkel et al. (Kenkel et al., 2021a). In total 18 unit-operations are implemented in the superstructure. 10 units are potentially exothermic and 3 processes are potentially endothermic. The complete superstructure model consists of 30,145 constraints, 23,709 variables of which 5,437 are binary. It was solved on a MacBook Pro with a 2 GHz-Dual Core Intel Core i5 processor and 8 GB RAM.

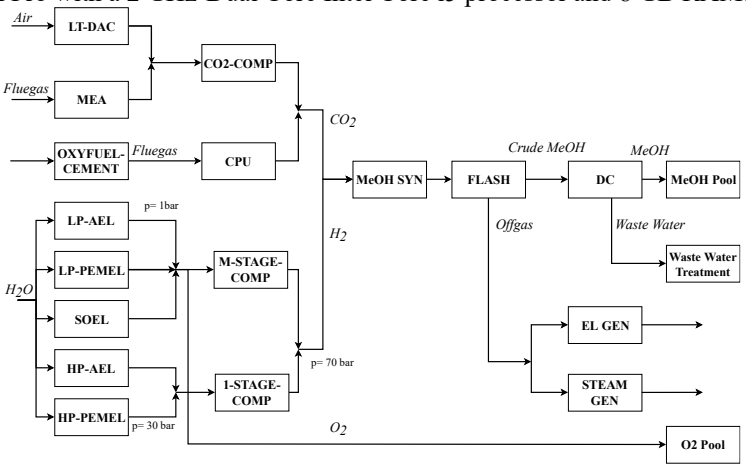


Figure 3: Simplified PtM superstructure adopted from Kenkel et al. (Kenkel et al., 2021a)

4.2. Flowsheet synthesis and heat integration results

An economic optimization of the PtM superstructure takes about 9 s on the defined computer system utilizing Gurobi (Kenkel et al., 2021a). The proposed flowsheet captures CO_2 from a refinery by absorption and provides H_2 by low pressure alkaline electrolysis with subsequent multi-stage compression. The purge stream as well as waste heat from methanol synthesis and H_2 compression is utilized for heat integration. The net production costs of the methanol are 892 € t^{-1}, where the utility supply (external heat, cooling water and HEN) makes up only 3 % of the total costs. The emerging excess heat is employed in the desorption process of the CO_2 capture, which requires heat at temperatures of about 120°C. Waste heat from methanol synthesis and H_2 compression provide ca. 28 % of the required heat, while purge stream combustion supplies further 12 %, leading to a total utility reduction of ca. 40 %. Only slightly lower shares were calculated in a detailed process analysis by Wassermann et al. (Wassermann et al., 2020).

4.3. Model limitations

The model evaluation and the applied case study show that the MILP heat integration model presents a possibility for easy implementation and fast calculation in superstructure models. Nonetheless, this formulation also comes with certain drawbacks. The main disadvantage is that only a virtual HEN is calculated, which gives good indication on the total costs, but neglects practical stream matching. A second drawback originates from the pre-calculated virtual HEX capital costs function. It is based on average gas/liquid/phase changing exchangers. If a superstructure with low pressure gas-gas HEX

above all is considered the required exchange area will probably be higher, which leads to underestimation of HEN costs.

5. Conclusion

This work presents a MILP model for heat integration in superstructure models, which is part of the automated **O**pen s**U**perstruc**T**ure mo**D**eling and **O**ptimizati**O**n f**R**amework (OUTDOOR). Model evaluation determines fast solution times, intuitive data input and acceptable accuracy of total HEN costs. A practical study on a PtM case with 30,145 constraints and 23,709 variables, which is solved in 9 s, emphasizes the benefits of OUTDOORs heat integration model. The results indicate that 40 % of required heat can be supplied by heat recovery, hence highlighting the relevance of heat integration in superstructure optimization. Model limitations, concerning simplified capital costs calculation and virtual HEN construction are outweighed by benefits of considering heat recovery with first indication of additional HEN costs. This is especially true for superstructure optimization, where many different flowsheets are investigated as part of preliminary design and HEN costs only make up small shares of total costs. Additionally, heat integration adds value to process synthesis by minimizing the environmental burden from external utilities, which is particularly relevant for the design of sustainable processes. Therefore, this work can provide the basis for selecting specific process alternatives for detailed process simulation and analysis.

6. Funding

Funding of this research by the German Federal Ministry for Economic Affairs and Energy within the KEROSyN100 project (funding code 03EIV051A) is gratefully acknowledged.

References

Ciric, A.R., Floudas, C.A., 1991. Heat exchanger network synthesis without decomposition. Comput. Chem. Eng. 15, 385–396. https://doi.org/10.1016/0098-1354(91)87017-4

Floudas, C., Ciric, A., Grossmann, I., 1986. Automatic synthesis of optimum heat exchanger network configurations. AIChE J. 32, 276--290.

Kenkel, P., Wassermann, T., Rose, C., Zondervan, E., 2021a. A generic superstructure modeling and optimization framework on the example of bi-criteria Power-to-Methanol process design. Comput. Chem. Eng. 150, 107327. https://doi.org/10.1016/j.compchemeng.2021.107327

Kenkel, P., Wassermann, T., Rose, C., Zondervan, E., 2021b. OUTDOOR–An open-source superstructure construction and optimization tool, in: Computer Aided Chemical Engineering. pp. 413–418.

Linnhoff, B., Townsend, D.W., Boland, D., Hewitt, G.F., Guy, A.R., Marsland, R., 1982. A user guide on process integration for the efficient use of energy, Revised 1s. ed. IChemE, Rugby, UK.

Wassermann, T., Schnuelle, C., Kenkel, P., Zondervan, E., 2020. Power-to-Methanol at Refineries as a Precursor to Green Jet Fuel Production: a Simulation and Assessment Study, in: Computer Aided Chemical Engineering. Elsevier B.V., pp. 1453–1458. https://doi.org/10.1016/B978-0-12-823377-1.50243-3

Yee, T.F., Grossmann, I.E., 1990. Simultaneous optimization models for heat integration-II. Heat exchanger network synthesis. Comput. Chem. Eng. 14, 1165–1184. https://doi.org/10.1016/0098-1354(90)85010-8

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50196-2

A Software Framework for Optimal Multiperiod Carbon-Constrained Energy Planning

Purusothmn Nair S Bhasker Nair[a], Dominic C. Y. Foo[a*], Raymond R. Tan[b], Michael Short[c]

[a]*Centre of Excellence for Green Technologies/Department of Chemical and Environmental Engineering, University of Nottingham Malaysia, Broga Road, 43500 Semenyih, Selangor, Malaysia*
[b]*Chemical Engineering Department/Centre for Engineering and Sustainable Development Research, De La Salle University, 2401 Taft Avenue, Manila, Philippines*
[c]*Department of Chemical and Process Engineering, University of Surrey, Guildford, Surrey GU2 7XH, United Kingdom*
dominic.foo@nottingham.edu.my

Abstract

The delay in action to mitigate climate change has resulted in a greater dependence on carbon dioxide removal (CDR) to achieve net-zero carbon targets by 2050. This work reports a newly developed optimal decarbonisation software framework that is based on a superstructure targeting approach to energy planning. The novel mathematical optimisation tool, which is formulated as a mixed-integer linear program (MILP), determines the optimum deployment of renewable energy sources, negative emission technologies (NETs), and CO_2 capture and storage (CCS) for long-term regional energy planning, subject to budget and emissions constraints. The software can be used by policymakers to determine long-term energy decarbonisation strategy including when to decommission which plants, what technologies to employ when, and which fuels could be replaced by lower-carbon alternatives. The application of the software framework is demonstrated with a case study containing seven power plants. In this multiperiod work, CCS deployment is favoured for coal-based power plants due to their high CO_2 intensity, while energy-producing NETs is deployed for all periods.

Keywords: Multiperiod Energy Planning; Negative Emission Technologies; Process Integration; Policymaking; Decarbonisation Software

1. Introduction

In 2015, 196 countries adopted the Paris Agreement, which aimed to limit global warming below 2 °C and preferably 1.5 °C above pre-industrial levels. Given the delay of major mitigation actions, there is an increasing urgency for countries to cut their emissions by 45 % by year 2030 from 2010 baseline levels (United Nations, 2021). The long-term target should be to achieve net-zero CO_2 emissions by 2050. At the current trajectory, achieving net-zero emissions would require the deployment of CDR, typically achieved via NETs. Examples of CDR methods are bioenergy with CCS and direct air capture. The former is categorised as energy-producing NETs (EP-NETs), while the latter is categorised as energy-consuming NETs (EC-NETs).

In the seminal work of carbon emissions pinch analysis (CEPA), Tan and Foo (2007) proposed the use of a graphical targeting tool to determine the minimum renewable

energy sources in satisfying the emission limit in a geographical region. This graphical CEPA approach was later extended for the incorporation of CCS in energy planning (Tan et al., 2009). Lee et al. (2009) developed the automated targeting model (ATM) to complement the limitation of CEPA, which was then extended by Ooi et al. (2013) for targeting CCS retrofit during energy planning. Most recently, the deployment of NETs during energy planning was proposed using both graphical (Nair et al., 2020) and algebraic targeting approaches (Nair et al., 2021). In the past, several software had been developed to aid in energy planning, covering both technical and economic approaches. The novelty of this decision-making software framework lies in the possibility of allowing potential retrofit of various CCS technologies at point sources of CO_2, and deployment of NETs, via a multiperiod planning formulation. The framework is built on mathematical programming models, where optimal energy planning scenarios may be generated. The interpretable results can act as guides for policymakers on the optimal decarbonisation strategies to be employed at a national level.

2. Problem Statement

For period $k \in K$, the demand and CO_2 emission limits of the energy planning system are specified as D_k and L_k respectively. Power plant $i \in I$ with a lower bound energy output $F_{i,LB}$, upper bound energy output $F_{i,UB}$, CO_2 intensity CS_i and cost CT_i make up the energy planning system for period k. The removal of the CO_2 emissions in period k is aided with the deployment of CCS technology $n \in N$, EP-NETs technology $p \in P$ and EC-NETs technology $q \in Q$. In this work, the superstructural mathematical formulation is optimised to synthesise an optimal decarbonisation plan.

3. Mathematical Optimisation Formulation

For period k, the summation of the energy output from power plants $i \in I$ must be equivalent to the demand of a specified geographical region (D_k), as shown in Eq.(1). Also, the energy output from power plant i in period k ($FS_{i,k}$) should be in the range of lower ($F_{i,LB}$) and upper bound of energy output ($F_{i,UB}$).

$$\sum_i FS_{i,k} = D_k \qquad \forall k \tag{1}$$

Next, the carbon intensity of power plant i with CCS technology n in period k ($CR_{i,n}$) is determined from Eq.(2) (Ooi et al., 2013).

$$CR_{i,n} = \frac{CS_i \times (1 - RR_n)}{1 - X_n} \qquad \forall i \; \forall n \tag{2}$$

where RR_n and X_n is the removal ratio and parasitic power loss of CCS technology n.

The net energy output from power plant i with CCS technology n in period k ($FNR_{i,k,n}$) is calculated from Eq.(3). Note that the reduced energy output from power plant i is due to the parasitic power losses during CCS. Also, $FNR_{i,k,n}$ should not exceed its upper bound of energy output in period k, as shown in Eq.(4).

$$FR_{i,k,n} \times (1 - X_n) = FNR_{i,k,n} \qquad \forall i \; \forall k \; \forall n \tag{3}$$

$$FR_{i,k,n} \leq F_{i,UB} \times B_{i,k,n} \qquad \forall i \; \forall k \; \forall n \tag{4}$$

where $FR_{i,k,n}$ is the extent of CCS retrofit technology *n* on power plant *i* in period *k* while $B_{i,n}$ is a binary variable for selection of power plant *i* with CCS technology *n* in period *k*.

The summation of the extent of CCS retrofit of power plant *i* with all CCS technologies in period *k* ($FR_{i,k}$) is calculated from Eq.(5). Also, the total extent of CCS retrofit of power plant *i* should not exceed the energy output from power plant *i* in period *k,* as shown in Eq.(6).

$$\sum_n FR_{i,k,n} = FR_{i,k} \qquad \forall i \; \forall k \tag{5}$$

$$FR_{i,k} \leq FS_{i,k} \qquad \forall i \; \forall k \tag{6}$$

For a given period *k*, The summation of the net energy output by power plant *i* without CCS retrofit ($FNS_{i,k}$) and the extent of CCS retrofit of power plant *i* with CCS technology *n* ($FR_{i,k,n}$) should equate to the energy output from power plant *i*; shown in Eq.(7).

$$FNS_{i,k} + \sum_n FR_{i,k,n} = FS_{i,k} \qquad \forall i \; \forall k \tag{7}$$

For a given period *k*, the summation of the energy output from all energy sources e.g., compensatory energy (FC_k), EP-NETs ($FEP_{k,p}$) etc. must fulfil the total demand of the energy system; the latter includes the total power requirement (D_k) and that required by EC-NETs ($FEC_{k,q}$) etc. as demonstrated in Eq.(8). Equally, the total CO_2 load contribution from all energy sources is equivalent to the total CO_2 emissions at the end of energy planning for period *k* (TE_k), shown in Eq.(9).

$$\sum_i \sum_n \left(FNS_{i,k} + FNR_{i,k,n}\right) + FC_k + \sum_p FEP_{k,p} = \sum_q FEC_{k,q} + D_k \; \forall k \tag{8}$$

$$\sum_i \sum_n \left(FNS_{i,k} CS_i + \left(FNR_{i,k,n} \, CR_{i,n}\right)\right) + FC_k \, CIC_k + \sum_p FEP_{k,p} \, CIEP_k + \sum_q FEC_{k,q} \, CIEC_k = TE_k \qquad \forall k \tag{9}$$

where CIC_k, $CIEP_k$ and $CIEC_k$ represent the carbon intensities of compensatory energy, EP-NETs technology *p* and EC-NETs technology *q* in period *k* respectively.

Meanwhile, the total energy costs at the end of energy planning in period *k* (TC_k) are calculated from Eq.(10).

$$\sum_i \sum_n \left(FNS_{i,k} CT_{i,k} + \left(FNR_{i,k,n} \, CTR_{i,k,n}\right) + \left(CFX_{i,k,n} B_{i,k,n}\right)\right) + FC_k \, CTC_k + \sum_p FEP_{k,p} \, CTEP_k + \sum_q FEC_{k,q} \, CTEC_k = TC_k \qquad \forall k \tag{10}$$

where $CTR_{i,k,n}$ and $CT_{i,k}$ represent the cost of energy output by power plant i with and without CCS technology n in period k respectively, while CTC_k, $CTEP_k$ and $CTEC_k$ represent the cost of compensatory energy, EP-NETs technology p, and EC-NETs technology q in period k respectively. Meanwhile, $CFX_{i,k,n}$ represents the fixed cost of power plant i with CCS technology n in period k.

The extent of CCS retrofit on power plant i at a later period is at least equal to that in its previous period, as shown in Eq.(11). This ensures that a decision taken to CCS retrofit power plant i in period k would not be reversed in the subsequent periods.

$$(FR_i)_{k+1} \geq (FR_i)_k \qquad k = 1,2,\dots,n-1 \tag{11}$$

The constraints regarding the total CO_2 emissions and total energy costs in period k are presented in Eq.(12) and Eq.(13).

$$TE_k = L_k \qquad \forall k \tag{12}$$

$$TC_k \leq BD_k \qquad \forall k \tag{13}$$

where BD_k is the budget allocation in period k.

The mathematical formulation may be optimised according to either Eq.(14) or Eq.(15). For Eq.(14), the total energy cost is minimised subject to constraints in Eq.(1) to Eq.(12). In other words, the minimisation of the total energy costs would ensure that the CO_2 emission limit in a geographical region in period k is satisfied. Meanwhile, for Eq.(15), the total CO_2 emissions are minimised subject to the constraints in Eq.(1) to Eq.(11) and Eq.(13). In other words, the minimisation of the total CO_2 emissions is conducted subject to the budgetary constraint for period k. Therefore, the CO_2 emission limit in a geographical for period k may or may not be satisfied.

$$min\, TC_k \qquad \forall k \tag{14}$$

$$min\, TE_k \qquad \forall k \tag{15}$$

The resulting mathematical formulation is a mixed-integer linear programming (MILP) model. The model is implemented in Pyomo, with an easy-to-use input spreadsheet to formulate one's problems. A user guide is available, including installation instructions for the Octeract solver, and the code is available at https://github.com/mchlshort/DECO2.

4. Case Study

The application of the optimal decarbonisation software framework is demonstrated with a hypothetical case study. Seven power plants with various energy sources generate electricity to satisfy the demand of a geographical region. The data relating to these power plants were approximately based on the data in the work by Tan et al. (2009). The energy planning is conducted across three periods, each with a specified demand, emission limit and budget allocation. The CO_2 load reduction is achievable with the potential deployment of three types of EP-NETs and EC-NETs, alongside two choices of CCS technologies. Also, there is a choice of incorporating additional renewable energy (compensatory energy) for satisfying the CO_2 emission limit. The superstructure optimisation of the MILP model provides an overview regarding the optimal deployment

of CCS, EP-NETs, EC-NETs, and compensatory energy for each period. Table 1, Table 2, and Table 3 present the data related to the power plants, NETs, and energy planning respectively, which was approximated based on the work by Tan et al. (2009).

Table 1: Power plant data

Power Plant	1	2 - 4	5	6 - 8
Fuel	Renewable	Natural gas	Oil	Coal
$F_{i,LB}$ / TWh y^{-1}	10	0	0	0
$F_{i,UB}$ / TWh y^{-1}	40	40	10	20
CS_i / Mt (TWh)$^{-1}$	0	0.5	0.8	1.0
$CT_{i,k}$ / mil USD y^{-1}	38	25	49	12
$CTR_{i,k,1}$ / mil USD y^{-1}	0	34	68	20
$CTR_{i,k,2}$ / / mil USD y^{-1}	0	29	59	16

Table 2: NETs data

Period k	$CIEP_k$ / Mt (TWh)$^{-1}$	$CIEC_k$ / Mt (TWh)$^{-1}$	$CTEP_k$ / mil USD y^{-1}	$CTEC_k$ / mil USD y^{-1}
1	-0.8	-0.6	43	49
2	-0.6	-0.4	40	37
3	-0.4	-0.2	37	24

Table 3: Energy planning data

Period k	D_k / TWh y^{-1}	L_k / Mt y^{-1}	BD_k / mil USD y^{-1}	CIC_k / Mt (TWh)$^{-1}$	CTC_k / mil USD y^{-1}
1	60	15	1,500	0	38
2	75	8	2,000	0	38
3	90	3	2,500	0	38

In addition to the power plant data in Table 1, the value of $CFX_{i,k,n}$ is assumed as 200 mil USD y^{-1}. The case study was solved for both objective functions i.e., minimum budget (Eq.(14)) and minimum emissions (Eq.(15)). For the minimum budget objective (Eq.(14)), the total costs for periods 1, 2 and 3 were determined as 1,790, 2,463 and 3,116 mil USD y^{-1} respectively, while emission limits for all periods are satisfied. Note that CCS technology 1 was solely deployed for the coal-based power plants for all periods. In other words, the optimisation model favoured the costlier CCS technology 1 due to its highest removal ratio (0.85) and lower parasitic power loss (0.15). All three periods involved the deployment of EP-NETs, specifically technology 2 in periods 1 and 2, and technology 1 in period 3. The energy requirement of EC-NETs made it less favourable for its deployment in any period. Due to the decreased CO_2 emission limit, the deployment of EP-NETs increased from 5.59 TWh y^{-1} in period 1 to 9 TWh y^{-1} in period

3. As for the minimum emissions objective (Eq.(15)), the optimisation is conducted subject to the budget constraints for each period. The total CO_2 emissions in periods 1, 2 and 3 are 27.1, 29.8 and 29.6 Mt y^{-1} respectively. Note that the total CO_2 emissions for all periods are higher than their limits, due to insufficient budget for the deployment of CCS and/or NETs. For this case, EP-NETs technology 1 was deployed, while CCS technology 1 was deployed for coal-based power plants in all periods.

5. Conclusions

The optimal decarbonisation software framework developed in this work is based on mathematical optimisation models for energy planning. The decision-making software framework is expected to aid policymakers in drafting suitable decarbonisation plans for a national/sectoral level. The MILP optimisation model provides rigorous solutions for optimal deployment of EP-NETs, EC-NETs, CCS and compensatory energy for the satisfaction of the CO_2 in each period. The case study results demonstrated that CCS technology 1 was favoured due to its highest removal ratio and lower parasitic power loss despite it being more expensive than technology 2. On the other hand, EC-NETs was less preferable due to its energy-consuming characteristic. Therefore, it is evident that this software framework could play a role in the achievement of the net-zero carbon target by 2050. Regardless, uncertainties that often arises with energy planning data and forecasts should be considered in future work for more realistic energy planning scenarios and projections.

References

Lee, S.C., Ng, D.K.S., Foo, D.C.Y., Tan, R.R., 2009. Extended pinch targeting techniques for carbon-constrained energy sector planning. Appl. Energy 86, 60–67. https://doi.org/10.1016/j.apenergy.2008.04.002

Nair, P.N.S.B., Tan, R.R., Foo, D.C.Y., 2021. A Generic Algebraic Targeting Approach for Integration of Renewable Energy Sources, CO2 Capture and Storage and Negative Emission Technologies in Carbon-Constrained Energy Planning. Energy 235, 121280. https://doi.org/10.1016/j.energy.2021.121280

Nair, P.N.S.B., Tan, R.R., Foo, D.C.Y., 2020. Extended Graphical Approach for the Deployment of Negative Emission Technologies. Ind. Eng. Chem. Res. 59, 18977–18990. https://doi.org/10.1021/acs.iecr.0c03817

Ooi, R.E.H., Foo, D.C.Y., Tan, R.R., Ng, D.K.S., Smith, R., 2013. Carbon constrained energy planning (CCEP) for sustainable power generation sector with automated targeting model. Ind. Eng. Chem. Res. 52, 9889–9896. https://doi.org/10.1021/ie4005018

Tan, R.R., Foo, D.C.Y., 2007. Pinch analysis approach to carbon-constrained energy sector planning. Energy 32, 1422–1429. https://doi.org/10.1016/j.energy.2006.09.018

Tan, R.R., Ng, D.K.S., Foo, D.C.Y., 2009. Pinch analysis approach to carbon-constrained planning for sustainable power generation. J. Clean. Prod. 17, 940–944. https://doi.org/10.1016/j.jclepro.2009.02.007

United Nations, 2021. Goal 13 : Take Urgent Action to Combat Climate Change and its Impacts [WWW Document]. Dep. Econ. Soc. Aff. URL https://sdgs.un.org/goals/goal13 (accessed 9.1.21).

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50197-4

Superstructure optimisation in various carbon capture and utilisation supply chains

Ikhlas Ghiat, Ahmed AlNouss, Tareq Al-Ansari

College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar.
talansari@hbku.edu.qa

Abstract

Carbon capture and utilisation (CCU) appeared as an attractive complementary scenario to carbon capture and storage for CO_2 sequestration and climate change mitigation. CCU covers various routes that use CO_2 as a feedstock for the generation of value-added commodities, which are classified under four main categories: chemical conversion, mineral carbonation, enhanced oil recovery and biological conversion. CCU networks encompass a variety of material, process, and technology options at every node of the supply chain. This forms a superstructure network of multiple CO_2 sources, carbon capture and separation methods, transportation links, and CO_2 utilisation routes (sinks). The purpose of this work is to investigate the economic performance of different CCU pathways and to determine the most optimal CCU integrative configuration. A mixed linear integer programing (MILP) model is developed that aims at optimising the CCU supply chain. A case study for the State of Qatar is presented in this work to illustrate the application of this optimisation framework, with the use of real data for existing CO_2 sources and sinks, and transportation networks along with potential CO_2 sinks. Findings indicate that for high CO_2 concentration and gas flowrates such as the ones studied in the case study, adsorption provides the most economical option for carbon capture. Moreover, the chemical conversion to GTL route presented the best economic performance amongst the other CO_2 utilisation pathways despite the low readiness level. However, biological conversion in agricultural greenhouses could be a more environmentally viable utilisation route with the potential of a carbon-neutral network.

Keywords: Carbon capture and utilisation, CO_2 sequestration, optimisation, MILP.

1. Introduction

Economic and population growth are considered as one of the main drivers of the continued increase of greenhouse gas emissions (GHG) in the atmosphere which lead to further warming and irreversible changes in the climate system (IPCC, 2007). Global greenhouse gas emissions from anthropogenic activities have significantly increased with emissions from fossil-fuel burning activities representing a large share. Carbon dioxide is considered as a major contributor of GHG emissions from burning fossil-fuels. The share of CO_2 emissions from the energy sector accounts for more than 80% of the total CO_2 emissions (Ghiat et al., 2021a). With the COVID19 pandemic and the associated decline in oil and coal demands, CO_2 emissions have declined by 5.8% in 2020 which represents the largest ever decline. Despite the pandemic, annual CO_2 emissions in the atmosphere originating from the energy sector are still at their peak with amounts as high as 31.5 Gt, representing a concentration of around 412.5 ppm during 2020 (IEA, 2021). The undeniable evidence of the anthropogenic influence on the climate is perceived in the growing vulnerability of the natural and human systems. This suggests that it is a

collective action problem that requires collaborative adaptation and mitigation measures between the different stakeholders and agents in the energy, water, and food (EWF) sectors. Co and tri-benefits can be exploited, and trade-offs can be avoided by having aligned objectives that tackle climate change challenges and deliver holistic sustainable solutions to all EWF sub-systems (IPCC, 2014). The intergovernmental panel on climate change (IPCC) recommends, as one of the Representative Concentration Pathways (RCP2.6), stabilizing the surface temperature increase below 2°C to limit CO_2-eq emissions to 450 ppm in the atmosphere by 2100 (IPCC, 2014). Carbon capture and storage or utilisation (CCUS) is considered as one of the most important CO_2 cutting schemes that can help achieve these global warming targets. CCU alone can reach a CO_2 sequestration potential of 878 Mt/year by 2060 in the case where CO_2 storage is limited. Moreover, CO_2 utilisation can also sequester CO_2 in different time-scale periods depending on the utilisation of the final product. CO_2 source-sink matching in CCU networks is a crucial and complex work that requires careful analysis and optimisation of the CCU supply chain (Ghiat and Al-Ansari, 2021b). The complexities of CCU supply chains lay in the multi-scale components of the network each having different options and being assessed separately, which renders the economic feasibility of the hybrid network. For example, the CO_2 source node can entail different CO_2 sources with different gas flowrates, CO_2 compositions and moisture contents (Hasan et al., 2015). Thus, there is a need to optimally design CCU supply chains using an inclusive and multi-scale approach that will ensure the best economic and environmental performance.

Several studies attempted to tackle the supply chain related complexities of CCU networks to reduce the associated costs and environmental impact and achieve decarbonised pathways. Kalyanarengan Ravi et al. (2017) proposed an economic optimisation model to minimise the overall costs related to a carbon capture and storage (CCS) network and implemented it for a nationwide CO_2 reduction scenario. Hasan et al. (2015) developed a multi-scale framework for the economic optimisation of CCU pathways encompassing enhanced oil recovery (EOR) and storage in both saline formation and unmineable cobalt as CO_2 sink opportunities. Similarly, Zhang et al. (2020) proposed an optimisation-based framework for assessing CCUS pathways that include more utilisation routes other than just EOR and proved its application on a network with 15 utilisation candidates. Both CCUS studies concluded that the CO_2 storage option was more economically optimal solution as compared to CO_2 utilisation, although this latter could offer a better environmental performance with a carbon-neutral cycle (Hasan et al., 2015; Zhang et al., 2020). While many studies tackled the economic optimisation of multi-scale CCUS networks, only few considered CO_2 utilisation in chemical conversion such as methanol, GTL, and urea production (Al-Yaeeshi et al., 2020). Moreover, the biological conversion of CO_2 has been neglected as a possible utilisation route within CCUS supply chains. In this work, we introduce three different CO_2 utilisation routes; 1) enhanced oil recovery (EOR), 2) chemical conversion, and 3) biological conversion; and study their integration within a multi-scale CCU network encompassing different CO_2 sources, different carbon capture (CC) technologies and CO_2 transportation. A mixed linear integer programming (MILP) model is developed to maximise the overall profit of the integrated CCU network and identify best economic pathway.

2. System description

In this study, a CCU optimisation model is developed that can encompass different CO_2 large-scale industrial sources, different carbon capture technologies and materials, CO_2 compression and transportation via pipeline, and varying CO_2 utilisation options as shown

in Figure 1. The aim is to optimise the overall profit of the integrated CCU network and determine the optimal source to sink integration. The CO_2 source node can be perceived as a group of CO_2 suppliers to the network which entails different capacity restrictions and exhaust gas characteristics. Each CO_2 source can be linked to a carbon separation method and material. The integration of the CO_2 source and carbon capture nodes is implemented with the assumption of both nodes being in the same location. Moreover, each CO_2 source can only be coupled with one carbon capture technology. CO_2 transportation from source to sink is only considered via pipeline. The choice of pipeline transportation is mainly motivated by its economic feasibility over other means. In addition, CO_2 sources can deliver to as many sinks with no restrictions apart from the availability of CO_2. Three CO_2 utilisation routes are proposed in this work; 1) enhanced oil recovery (EOR), 2) chemical conversion to methanol, GTL, and urea, and 3) biological conversion in agricultural greenhouses. The proposed optimisation model is implemented for a case study in the state of Qatar comprising of three existing large scale industrial CO_2 sources (Qatargas, QAFAC and QAFCO), three different carbon capture technologies (chemical absorption with MEA or PZ, adsorption PSA with AHT or MVY and membrane separation with FSC PVAm or PEO1 or PEO2), CO_2 transportation via pipeline, and three chemical conversion routes including EOR in an existing oil well, chemical conversion to methanol, GTL and urea in existing process plants within QAFAC, QSGTL, Oryx and QAFCO, and chemical conversion in an existing agricultural greenhouse that uses CO_2 enrichment. Different scenarios are studied; scenario 1 is the baseline scenario with no added constraints, scenario 2 considers the optimisation without the CO_2 chemical conversion options given their low readiness level, scenario 3 constraints the optimisation problem to supply greenhouses with CO_2 to meet at least 21% of Qatar's self-sufficiency in vegetable production, and scenario 4 couples the constraints of both scenario 2 and 3.

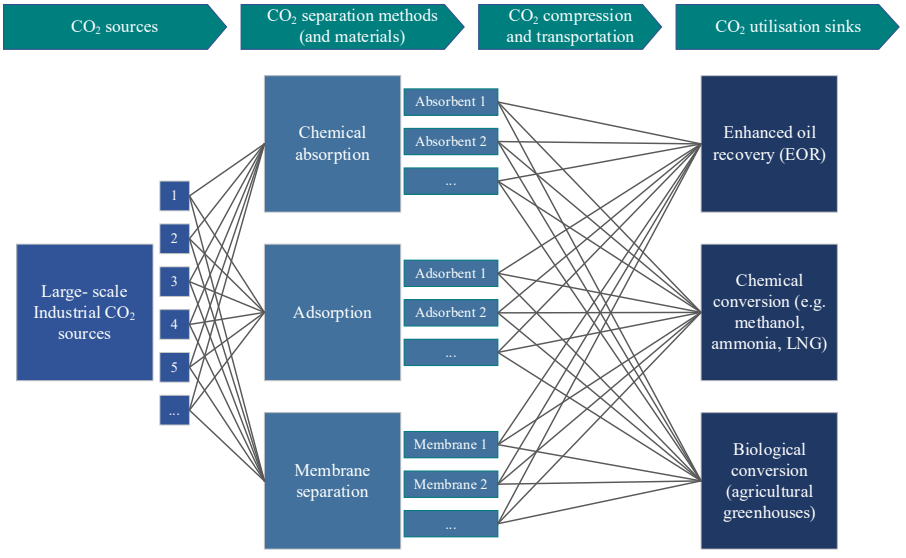


Figure 1: Carbon capture and utilisation network.

3. CCU optimisation model

A mixed integer linear programming (MILP) model is developed for the CCU supply chain network that aims at maximising the profit of the integrated system which is defined as the difference between the total annual cost and total revenue as shown in Eqs.(1&2).

$$Max \sum_{i,j,k}(TR - TC)F \quad (1)$$

$$= \left[\sum_k TR_k - \sum_{i,j} DC_i + ICC_{i,j} + OCC_{i,j} + \sum_{j,k} TTC_{j,k} + \sum_k UC_k\right]F \quad (2)$$

Subject to:

$$\sum_j F_{i,j} \leq 1 \quad (3)$$

The subscripts i,j,k correspond to the CO_2 sources, carbon capture technologies and CO_2 sinks respectively. TR and TC are the total annual revenues and costs respectively and F is the CO_2 flowrate. The total costs include the cost of flue gas dehydration (DC_i), carbon capture ($ICC_{i,j}$ and $OCC_{i,j}$), compression and transportation ($TTC_{j,k}$), and CO_2 utilisation (UC_k). TR_k and UC_k are taken from process-based studies conducted by Ogden et al. (2014) for EOR, Al-Yaeeshi et al. (2020) and AlNouss et al. (2021) for chemical conversion, and Ghiat et al. (2021a) for biological conversion. The presented constraint ensures that each CO_2 source is coupled with only one carbon capture technology.

- Cost of dehydration

Chemical absorption can handle saturated streams; however, adsorption and membrane separation techniques require less than 0.1% of water content. The cost of dehydration to reduce the water content to less than 0.1% is taken as 10.2$/t of CO_2 using tri-ethylene glycol (TEG) absorption (Hasan et al., 2015).

- Cost of CO_2 capture and compression

$ICC_{i,j}$ and $OCC_{i,j}$ are the investment and operational costs of carbon capture respectively depending on the carbon capture technology and material used. These costs also include the cost of CO_2 compression to a supercritical state, approximately 15 MPa. α, β, γ, α', β', γ' are parameters of each carbon capture technology and material based on the optimised economic model presented by Hasan et al. (2015).

$$ICC_{i,j} = \alpha_j + (\beta_j x_{co2_i}^{n_j} + \gamma_j)F^m \quad (4)$$

$$OCC_{i,j} = \alpha_j' + (\beta_j' x_{co2_i}^{n_j'} + \gamma_j')F^{m\prime} \quad (5)$$

- Cost of CO_2 transportation

The investment and operational costs of transportation via pipeline are estimated by the levelised cost in Eq.(6). This cost is estimated based on an inlet CO_2 pressure of 15 MPa and outlet CO_2 pressure of 10 MPa (Ogden, 2004).

$$TTC_{j,k}(\$/y) = (CRF + OM)CP_{base}\left(\frac{Q}{Q_{base}}\right)^{0.48} L_{j,k}\left(\frac{L_{j,k}}{L_{base}}\right)^{0.24} \quad (6)$$

where CRF is the capital recovery factor taken as 15% and OM is the fractional operational and management costs per investment cost and is taken as 4%. CP_{base} is the pipeline base capital cost taken as 700 $/m. Q is the CO_2 flowrate and Q_{base} is the CO_2 base flowrate taken as 16,000 ton/day. $L_{j,k}$ is the length of the pipeline (m) from the carbon capture technology j to the sink k, and L_{base} is the base length of the pipeline taken as 100 km (Ogden, 2004).

4. Results

The results of the optimization problem concerning the maximum profit from competing CO_2 capture and utilisation pathways revealed different results based on the applied constraints. The first scenario where no constraints on the CO_2 utilisation are applied, the results demonstrate a complete domination of GTL as the CO_2 sink and adsorption with PSA-AHT as the carbon capture technology and material with a high economic profit of $38,067. However, when constraints of technology readiness level are applied, results shift to a complete domination of EOR as the CO_2 sink and adsorption with membrane POE2 as the carbon capture technology and material with a much lower objective function estimated at $553. This reflects the importance in the technology investment to increase the readiness level of the chemical conversion as a potential sink for emitted CO_2. Further, with the application of the 21% self-sufficiency constraint in agriculture, the results of the third scenario indicate the utilisation of adsorption with PSA-MYV as the carbon capture technology and material to feed the greenhouse with the required CO_2 to achieve the self-sufficiency objective as illustrated in Figure 2a. In addition to the greenhouse, GTL is presented as the major CO_2 sink as presented in Figure 2a, achieving an overall profit of $29,083. The fourth scenario demonstrates a shift towards EOR with the elimination of GTL due to the readiness level constraint while achieving the self-sufficiency objective and with $435 profit through feeding the greenhouse with 71% of the CO_2 emitted from QAFAC source and captured using adsorption PSA MYV technology as illustrated in Figure 2b. Findings have also indicated that for high CO_2 concentration and gas flowrates such as the ones presented in the case study, adsorption provides the most economical option for carbon capture (Figure 3).

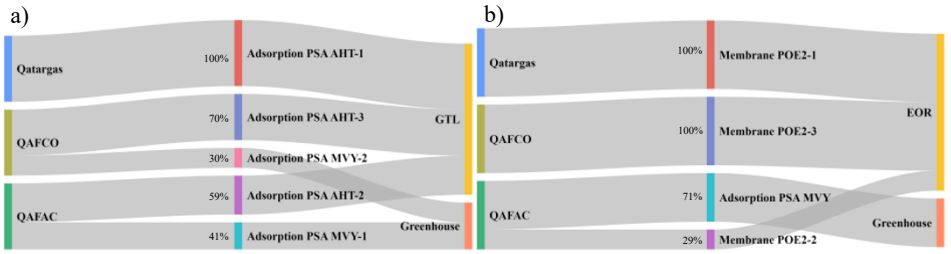


Figure 2: Optimal CO_2 source-sink allocation for a) scenario 3 and b) scenario 4.

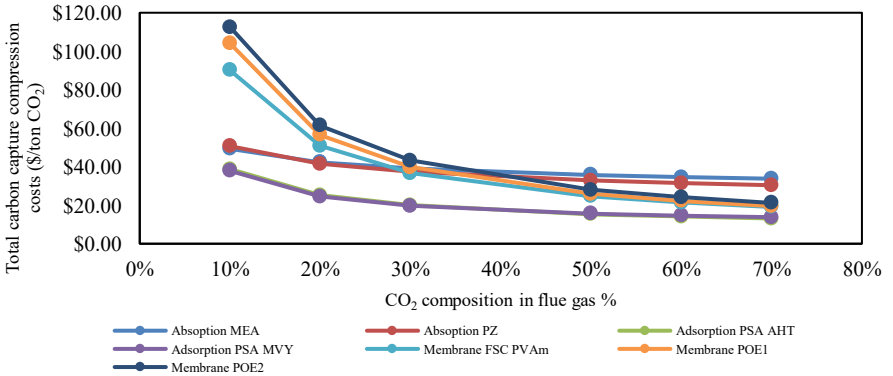


Figure 3: Total CAPEX and OPEX for CC and compression for each technology and material with a gas flow rate of 1000 mol/s.

5. Conclusions

While the principal goal of carbon capture and utilisation technologies is to mitigate CO_2 emissions, it is also important to develop CCU networks that can run on economically feasible supply chains. Hence, this study proposes a multi-scale optimisation model for CCU routes encompassing enhanced oil recovery, chemical conversion, and biological conversion as possible utilisation pathways. The aim of the model is to maximise the overall profits of the integrated CCU supply chain and determine the best optimal CCU integration. The model demonstrates its practicability as it involves different CO_2 sources, different carbon capture technologies and materials, CO_2 transportation via pipeline and finally different possible utilisation sinks. This model can be applied to specific CCU supply chains as it is the case in this study. The results of the optimisation model applied to the case study in Qatar reveal the economic feasibility of carbon capture using adsorption for the selected high CO_2 concentration and high gas flowrate sources coupled with chemical conversion for GTL production. The model can also be tailored to investigate different scenarios which can aid in the decision-making process for the deployment of CCU technologies. Finally, it is important that decision-makers help facilitate the deployment of CCU technologies by building the necessary infrastructure for CO_2 transportation and storage and by creating CO_2 source-sink clusters and networks that will use the same shared infrastructure and benefit from a circular economy.

References

A . AlNouss, G. Mckay, T. Al-Ansari, 2021, Utilisation of Carbon Dioxide and Gasified Biomass for the Generation of Value Added Products, Computer Aided Chemical Engineering, 50, 1567-1572.

A. A. Al-Yaeeshi, R. Govindan, T. Al-Ansari, 2020, Techno-economic-based dynamic network design for optimum large-scale carbon dioxide utilisation in process industries, Journal of Cleaner Production, 275, 122974. https://doi.org/10.1016/j.jclepro.2020.122974

IEA (2021), Global Energy Review 2021, IEA, Paris https://www.iea.org/reports/global energy review-2021

IPCC, 2007, Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.

IPCC, 2014, Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.

I. Ghiat, F. Mahmood, R. Govindan, T. Al-Ansari, 2021a, CO2 utilisation in agricultural greenhouses: A novel 'plant to plant' approach driven by bioenergy with carbon capture systems within the energy, water and food Nexus, Energy Conversion and Management, 228, 113668.

I. Ghiat, T. Al-Ansari, 2021b, A review of carbon capture and utilisation as a CO2 abatement opportunity within the EWF nexus, Journal of CO2 Utililisation, 45, 101432.

J.M. Ogden, 2004, Conceptual Design of Optimized Fossil Energy Systems with Capture and Sequestration of Carbon Dioxide, UC Davis: Institute of Transportation Studies.

M.M.F. Hasan, E.L. First, F. Boukouvala, C.A. Floudas, 2015, A multi-scale framework for CO2 capture, utilization, and sequestration: CCUS and CCU, Computers and Chemical Engineering, 81, 2–21. https://doi.org/10.1016/j.compchemeng.2015.04.034

R.N. Kalyanarengan, M. Van Sint Annaland, J.C. Fransoo, J. Grievink, E. Zondervan, 2017, Development and implementation of supply chain optimization framework for CO2 capture and storage in the Netherlands, Computers and Chemical Engineering, 102, 40–51.

S. Zhang, Y. Zhuang, L. Liu, L. Zhang, J. Du, 2020, Optimization-based approach for CO2 utilization in carbon capture, utilization and storage supply chain, Computer and Chemical Engineering, 139, 106885. https://doi.org/10.1016/j.compchemeng.2020.106885

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50198-6

Efficient Scenario Generation for Stochastic Programs with Extreme Events

David Young[a], Mark Carpenter[b], and Selen Cremaschi[a*]

[a]*Department of Chemical Engineering, Auburn University, Auburn, AL 36849, USA*
[b]*Department of Mathematics and Statistics, Auburn University, Auburn, AL 36849, USA*
selen-cremaschi@auburn.edu

Abstract

Stochastic programming (SP), a popular approach for solving optimization problems under uncertainty, is commonly used to tackle chemical engineering problems, e.g., in production planning or process synthesis. Scenarios, which represent uncertain outcomes, significantly impact the SP solution. This study evaluates seven methods to generate scenarios for a two-stage stochastic program where the decision-maker only sees benefits within the rare-event space of the uncertainty. The methods belong to one of three main categories: Monte Carlo sampling, space-filling sampling, and clustering. We assess the methods using (1) the difference between the optimum objective values and (2) the distance between the decision variable values of the optimum SP solution and the best-known solution for the problem. The results revealed that the SP solutions obtained using the scenarios generated by the clustering-based approaches were close to the best-known solutions and did not change significantly as the number of scenarios increased. The SP solution for the scenarios generated by Latin Hypercube sampling was the closest to the best-known solution for the maximum number of scenarios.

Keywords: Stochastic Programming; Rare-event Sampling; Scenario Generation.

1. Introduction

Stochastic programming (SP) is a common approach to solve optimization problems under uncertainty. Such problems arise when the decision-maker incorporates aspects of a problem that are not fully known or cannot be controlled, e.g., the weather, demand, or estimated process parameters. This uncertain information is modeled as a random distribution based on previous data or expert opinion. Generally, the SP models estimate the distribution using a discrete set of outcomes, or scenarios, generated by sampling from the distribution. Naturally, the better a set of scenarios depicts the underlying distribution, the more accurate the SP solution becomes. However, as the number of the scenarios grows, so does the problem size, resulting in computationally intractable models. An efficient scenario generation (SG) approach would balance the model size and the accuracy of the distribution estimation (Park et al., 2019; Shapiro et al., 2014).

In many stochastic programs, the most likely scenarios are enough to estimate the distribution for a high-quality solution. There are, however, some problems, such as those in safety and reliability engineering (e.g., Moskowitz et al. 2018), supply chain resilience (e.g., Ehrenstein et al. 2019), and health services, in which the least likely events have the most impact on the optimal decisions. These events occur in the tail-end(s) of the distributions. In general, a large number of scenarios would need to be generated to ensure their presence in the SP scenario set. How to properly represent the

events associated with the tails of the distributions in the SP scenario set while maintaining a computationally tractable model is an open challenge.

Two recent studies assess different methods for SP scenario generation. The aim of Park et al. (2019) was to find the best scenario reduction method assuming the full scenario set was known. They concluded that a reduced scenario set generated by a distance-based method yields the closest solution to the optimum. Generating the full scenario set may be infeasible due to the sheer number of possible uncertainty realizations for some problems. Li et al. (2020) performed a comprehensive review of SG methods that have been used in energy systems SP problems. They outlined the characteristics of the methods and assessed them based on the solution stability. However, they have not considered the ability of the SG methods to capture rare events.

This paper evaluates seven SG methods (Section 3) for SP problems where the least likely events significantly impact the optimal decisions, and generating the full scenario set is infeasible. We carry out the evaluation using a healthcare-related problem (Section 2) by comparing the solutions identified by the SP models with scenarios generated using the SG methods to the best-known solution of the problem (Section 4).

2. Optimization Model – Screening Plan for Colorectal Cancer

The problem is to identify the ages at which screening for colorectal cancer (CRC) should occur to maximize the expected gain in Quality-Adjusted Life-Years (QALY). It is modeled as a two-stage stochastic program with fixed recourse. The general model form is given in Eqns. 1-4. The uncertain vector, $\boldsymbol{\omega}_s$, describes the lifetime of individual s through various health states to track the progression of CRC (Fig. 1). The first stage decision variables, $\boldsymbol{x}$, are the age(s) at which screening with a colonoscopy takes place. The second stage variables, $\boldsymbol{y_s}$, represent the modified lifetime of individual s based on the outcome of the recommended screening.

The objective function, Eqn. 1, calculates the expected change in QALY given the implementation of screening over every scenario, considering the loss of years due to the implementation of the screening and the gain in life due to early detection or prevention of the disease. Eqn. 2 places restrictions on the screening decisions, i.e., the maximum number of screens in a lifetime, ordering of the screenings, and only one screening test per year. Eqn. 3 implement the fixed recourse (clinical) actions following the implementation of the screening decisions, $\boldsymbol{x}$. They modify an individual's lifetime.

$$max\ EQALY \approx max \sum_{s\in S} g_2(\boldsymbol{x}, \boldsymbol{y_s}, \boldsymbol{\omega_s}) \tag{1}$$

$$s.t.\ \ A_1\boldsymbol{x} + b_1 \leq 0 \tag{2}$$

$$A_2\boldsymbol{x} + B\boldsymbol{y}_s + C\boldsymbol{\omega}_s + b_2 \leq 0 \qquad \forall s \in S \tag{3}$$

$$\boldsymbol{x} \in \mathbf{X} = \mathbb{R}^{\mathrm{n}}, \boldsymbol{y}_s \in \mathbf{Y} = \mathbb{R}^{\mathrm{m}} \qquad \forall s \in S \tag{4}$$

A defining feature of the problem is the impact of uncertainty on the solution. Only a small percentage of the population, 4-4.5 % (American Cancer Society, 2020), develops CRC within their lifetime, meaning that screening for CRC is a burden for most of the population. When only considering the most likely scenarios, the solution to this problem is to do nothing, as the detriments of screening outweighing the benefits. Many studies have shown that the implementation of screening for CRC is overall a benefit for society (e.g., Knudsen et al., 2016). Therefore, to properly represent the expected benefit of screening, it is essential to portray the lower probability scenarios accurately.

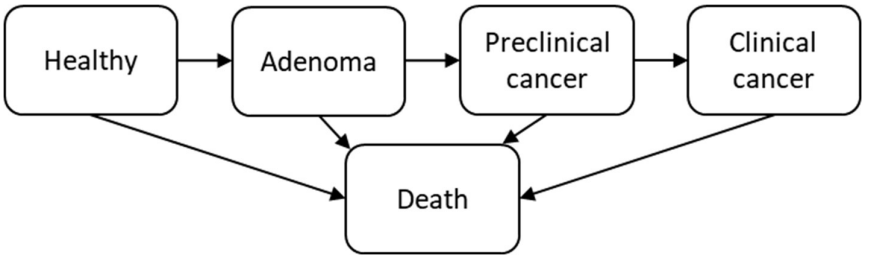


Figure 1 Colorectal cancer progression flow chart (Loeve et al., 1999)

3. Scenario Generation Methods and Their Application to the Model

The problem uncertainty space was constructed using data from a microsimulation model (Young et al., 2021; Young and Cremaschi, 2018) that simulates the progression of CRC within a population. We used the model to simulate 400 replications of a population of 1,000,000 males and recorded the values of a cancer progression random vector for each person. The vector includes the age of the first adenoma, time from the first adenoma to the first case of preclinical cancer, time from the first preclinical cancer to the first clinical cancer, and cancer survival time. This data was used to construct an empirical distribution, where each outcome of the random vector was considered a potential realization. Another empirical distribution was constructed using the US lifetables (Arias, 2002) to represent the age of death from sources other than CRC.

The SG methods fall under three main categories: Monte Carlo, space-filling, and clustering. The Monte Carlo and the space-filling methods utilized the empirical distributions to generate the scenarios. The clustering approaches used the raw data (the vectors) from the microsimulation to construct the scenarios.

3.1. Monte Carlo Methods

We consider the Crude or naïve Monte Carlo (CMC) approach and importance sampling. In the CMC approach, the scenarios are generated by directly sampling from the uncertain distributions. Importance sampling is a variance reduction technique. The random samples are generated from an auxiliary distribution, known as the importance distribution (ID), and reweighted using the likelihood ratio to tie the sample probability back to the original distribution. The construction of the ID is a challenge, as the theoretical optimal distribution requires accurate knowledge of the expectation of the distribution. We construct the ID following the work of Papavasiliou and Oren (2013) by first generating a large number of samples from the distribution and then estimating the ID using the expectation over the samples. We then resample from the estimated ID to generate the scenarios. We name this approach M-N-IS, where M refers to the initial large sample size, N to the number of samples from the ID, and IS to importance sampling.

3.2. Space-filling Methods

In space-filling methods, a quasi-Monte Carlo approach, samples are generated to avoid over-grouping or sparsity. This study considers Sobol sampling, Halton sampling, and Latin hypercube sampling (LHS). Sobol and Halton sampling techniques utilize Sobol (Sobol, 1967) and Halton sequences (Halton, 1960), low-discrepancy sequences, where discrepancy measures the uniformity of a set of samples. Latin hypercube sampling generates N samples by subdividing each distribution of a d-dimensional hypercube

into N bins of equal probability and ensures that no two samples are in the same bin on any axis of the hypercube (McKay et al., 1979).

3.3. Clustering Based SG Methods

Clustering is an unsupervised machine learning approach that groups data into subsets, or clusters, of points with similar characteristics. Clustering can be used as a SG method by grouping a large dataset sampled from the uncertain distribution and generating representative samples, or "cluster centers," for each cluster. The representative samples are then used as the scenarios for the SP problem(B. Li et al., 2020). We employed k-means and x-means algorithms for determining the cluster centers. The k-means algorithm separates the data set into k (an input to the algorithm) different clusters based on the distance between the mean vectors of the data (Yadav and Sharma, 2013). We used the elbow method with the sum of the squared distance from the nearest cluster center as the metric to determine k (Yuan and Yang, 2019). As an extension of k-means, the x-means algorithm also determines the number of clusters, k, using the Bayesian Information Criterion.

4. Results and Discussion

Each sampling method was used to generate 128, 512, and 1024 scenarios, except for the k-means and x-means algorithms, where the algorithms automatically determine the number of scenarios. The scenario cases are powers of two due to Sobol sequence convergence properties. The maximum number of scenarios, 1024, was set to keep the SP solution time below 24 hours. We also studied the impact of the sample size, 100,000, 1,000,000, and 2,000,000, for M-N-IS, k-means, and x-means SG methods.

The SP solutions obtained using scenarios generated by different SG methods are assessed using the best solution located when a derivative-free optimization (DFO) framework was used with the CRC microsimulation (Young et al., 2021; Young and Cremaschi, 2018) for a maximum budget of 1,500 evaluations. The performance metrics are the absolute percent difference in the objective function values and the Euclidian distance between the SP and the best solution.

For each SG method, a summary of the resulting absolute percent difference is plotted in Fig. 2, and the Euclidian distance in Fig. 3. In general, as the number of scenarios increases, the percent difference decreases for each method, except for the Sobol and Halton methods. The SP solution obtained using 128 scenarios generated by the Sobol sampling has the lowest percent difference (Fig. 2) but the greatest distance when compared to the best solution (Fig. 3). The largest reduction in percent difference with the increase in the number of scenarios is observed for LHS. The SP solution obtained using 1024 LHS generated scenarios has the second-best object function value with a 28.8% difference. There are no significant improvements in the SP objective function value with increases in the number of scenarios generated using either the CMC approach or the two clustering approaches. The SP solutions obtained using the CMC-generated scenarios had worse percent difference values than most methods (Fig. 2).

Similar to the percent difference measure, as the number of scenarios increases, the distance from the optimum decreases (Fig. 3), except for the clustering algorithms, where the distance stays relatively constant. At 1024 scenarios, there are small

differences in the distances to the best solution amongst different methods, with CMC, Sobol, and Halton sampling methods yielding slightly larger Euclidian distances. The SP solutions obtained using the scenarios generated by LHS are the closest to the best solution both at the minimum and the maximum number of scenarios. When the distances of the SP solutions obtained using CMC approach are compared to the rest, similar trends to the percent difference are observed (Figs. 2 and 3). At 1024 scenarios, the SP solutions obtained using the CMC-generated scenarios are close to the solutions obtained using other methods with at most an 11 % difference in distance.

The number of samples used by the clustering algorithms changes the number of scenarios, i.e., the number of clusters identified (Figs. 2 and 3). While the number of scenarios identified by the k-means algorithm does not change significantly, the number reaches the upper bound for the x-means algorithm for one and two million samples. This observation suggests that the scenarios identified by the x-means algorithm could be improved if a larger scenario size were allowed. For the M-N-IS method, with more samples to construct the ID, both percent difference and distance from optimal metrics are improved as the number of scenarios increase (Figs. 2 and 3). Both percent difference and distance from optimal solution improve as the samples increase for 1024 scenarios, but the inverse is observed for 124 scenarios. Using more samples to construct the ID yields a better estimate. Conversely, introducing more samples into the ID creates more possible realizations of uncertainty, requiring a greater number of sub-samples, scenarios, to better represent the ID.

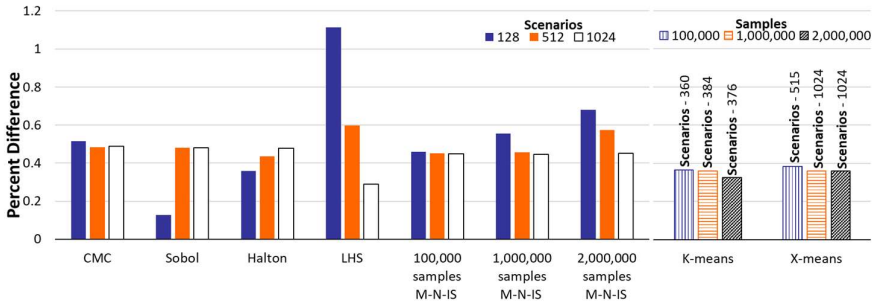


Figure 2 Summary of the absolute percent difference in the objective function values of the SP solutions obtained using scenarios generated by different SG methods and the best solution

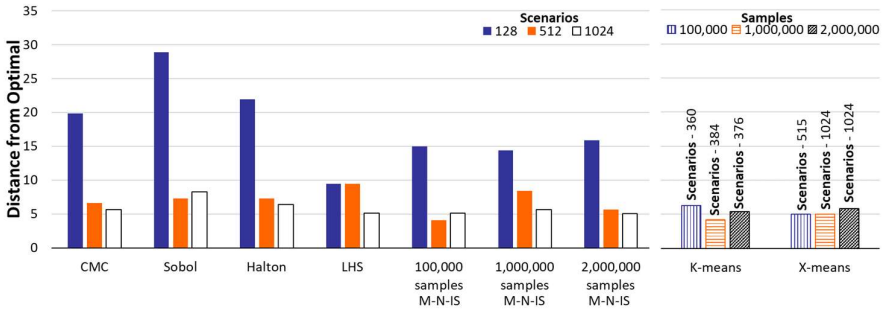


Figure 3 Summary of the Euclidian distance between the SP solutions obtained using scenarios generated by different SG methods and the best solution

5. Conclusions and Future Directions

This paper evaluated seven scenario generation methods for identifying the optimal solution and the objective value for a two-stage stochastic program where rare events significantly impact the objective. Scenarios generated by the Latin Hypercube sampling yielded the closest solution to the optimum at the largest scenario sizes. The clustering algorithms yielded solutions that were consistently close to the optimum. Future work will consider different clustering algorithms.

References

American Cancer Society, 2020. Colorectal Cancer Facts & Figures [WWW Document]. URL https://www.cancer.org/research/cancer-facts-statistics/colorectal-cancer-facts-figures.html (accessed 5.27.21).

Arias, E., 2002. United States life tables, 2000. Natl. Vital Stat. Reports 51, 1–63.

Ehrenstein, M., Wang, C.-H., Guillén-Gosálbez, G., 2019. Strategic planning of supply chains considering extreme events: Novel heuristic and application to the petrochemical industry. Comput. Chem. Eng. 125, 306–323.

Halton, J.H., 1960. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numer. Math. 2, 84–90.

Knudsen, A.B., Zauber, A.G., Rutter, C.M., Naber, S.K., Doria-Rose, V.P., Pabiniak, C., Johanson, C., Fischer, S.E., Lansdorp-Vogelaar, I., Kuntz, K.M., 2016. Estimation of Benefits, Burden, and Harms of Colorectal Cancer Screening Strategies: Modeling Study for the US Preventive Services Task Force. JAMA 315, 2595–2609.

Li, B., Sedzro, K., Fang, X., Hodge, B.M., Zhang, J., 2020. A clustering-based scenario generation framework for power market simulation with wind integration. J. Renew. Sustain. Energy 12.

Li, J., Zhou, J., Chen, B., 2020. Review of wind power scenario generation methods for optimal operation of renewable energy systems. Appl. Energy 280, 115992.

Loeve, F., Boer, R., Van Oortmarssen, G.J., Van Ballegooijen, M., Habbema, J.D.F., 1999. The MISCAN-COLON Simulation Model for the Evaluation of Colorectal Cancer Screening. Comput. Biomed. Res. 32, 13–33.

McKay, M.D., Beckman, R.J., Conover, W.J., 1979. Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics 21, 239–245.

Moskowitz, I.H., Seider, W.D., Patel, A.J., Arbogast, J.E., Oktem, U.G., 2018. Understanding rare safety and reliability events using transition path sampling. Comput. Chem. Eng. 108, 74–88.

Papavasiliou, A., Oren, S.S., 2013. Multiarea stochastic unit commitment for high wind penetration in a transmission constrained network. Oper. Res. 61, 578–592.

Park, S., Xu, Q., Hobbs, B.F., 2019. Comparing scenario reduction methods for stochastic transmission planning. IET Gener. Transm. Distrib. 13, 1005–1013.

Shapiro, A., Dentcheva, D., Ruszczyński, A., 2014. Lectures on Stochastic Programming: Modeling and Theory, Second Edition. Lect. Stoch. Program. Model. Theory, Second Ed.

Sobol', I.M., 1967. On the distribution of points in a cube and the approximate evaluation of integrals. USSR Comput. Math. Math. Phys. 7, 86–112.

Yadav, J., Sharma, M., 2013. A Review of K-mean Algorithm. Int. J. Eng. trends Technol. 4, 2972–2976.

Young, D., Cremaschi, S., 2018. A Simulation-based Optimization Approach to Develop Personalized Colorectal Cancer Screening Strategies, in: Computer Aided Chemical Engineering. Elsevier, pp. 2125–2130.

Young, D., Haney, W., Cremaschi, S., 2021. Derivative-free optimization of combinatorial problems – A case study in colorectal cancer screening. Comput. Chem. Eng. 145, 107193.

Yuan, C., Yang, H., 2019. Research on K-value selection method of K-means clustering algorithm. J 2, 226–235.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50199-8

A Sustainable Framework for Optimal and Flexible Design Under Uncertainty in Separation Processes: Exergy, Energy, Economic, and Environmental Aspects

Jaime D. Ponce-Rocha[a,c], Martín Picón-Núñez[a], Andreia Santos[b], Ana Carvalho[b], Fernando I. Gómez-Castro[a], Ricardo Morales-Rodriguez[a*].

[a]*Departamento de Ingeniería química, Univeridad de Guanajuato, Guanajuato, 36050, México.*
[b]*Centro de Estudos do IST, Universidad de Lisboa, Lisboa, 1049-001, Portugal.*
[c]*Dirección de Ingeniería y Construcción de Plantas, CIATEQ A. C., Querétaro, Querétaro, 76150, México.*
ricardo.morales@ugto.mx

Abstract

The global situation forces to design processes based on sustainable metrics and capable to respond to constant changes in the world. Particularly, the field of chemical processes requires an intensive energetic demand. These requirements force the scientist to develop systematic frameworks focused on mathematical formulations and computer-aided tools. The aim of this work is to develop and implement a systematic framework for multi-objective optimization under uncertainty based on four steps, first step: a mathematical formulation is defined, second step: the uncertainty identification and sampling are implemented, third step: a computational platform is generated based on commercial computer-aided tools (aspen plus & matlab) interconnected using a COM interface, fourth step: making decision criteria and data analysis. The results showed that the framework and employed criteria (The 80 % of cumulative distribution function value) for making decision are adequate for sustainable process design under uncertainty scenarios.

Keywords: Sustainable process design, 4E analysis, MOGA, Optimization under uncertainty.

1. Introduction

Currently, the design of chemical processes demands precise guidelines related with sustainable development targets. Nevertheless, these goals have not been properly and completely associated with rigorous mathematical models, forcing the scientists to reach metrics and making-decision reliable criteria for sustainable targets (Al *et al*, 2020). The chemical processes and mainly bioprocesses are characterized for having, variable yields and unexpected behaviours due to inherent and external effects, which are directly associated with uncertainty sources; therefore, a process that is optimal under a nominal condition may be suboptimal, or infeasible, once the process conditions change. Thus, to ensure the capability to maintain feasible process operation over a range of uncertainty factors, it is necessary to introduce a flexibility approach in process design, employing mathematical formulations focused on representing these phenomena

to obtain reliable and feasible designs (Adams *et al*, 2018). To tackle the concerns mentioned above, disciplines as process systems engineering promote the problem solution through systematic frameworks, based on optimal process design, integrating a collaborative and adaptable synergy between computer-aided tools (*v. g.* Chemical process simulators, programming environments) and mathematical strategies (*v. g.* multi-objective optimization, uncertainty optimization) (Avraamidou *et al*, 2019). Therefore, based on the state of the art, the objective of this work is developed and implement a systematic framework for multi-objective optimization under uncertainty (MOOUU), through a mathematical formulation using a platform based on commercial computer-aided tools, as well as the evaluation of statistical indicators as making-decision criteria for process design under uncertainty.

2. Case study

Nowadays, the use of conventional distillation continues like the principal separation and purification technique, which process large amounts of chemical products, but it is still a high energy consumption stage in the chemical industry. The preliminary structural and operational design of tray distillation columns aims to obtain their optimal specifications to meet product specifications (Gozálvez-Zafrilla *et al*, 2021). The production of Acetone-Butanol-Ethanol (ABE) using fermentation consortiums continues being an attractive pathway, for obtaining diverse intermediate chemical compounds to synthesize biofuels, bioplastics, etc. But the low and variable production, high diluted products, and the intensive energy requirements is still a challenge to tackle at industrial level. Thereby, the ABE separation and purification scheme (SPS) represents an interesting challenge, and it is employed as a case study to implement and evaluate the performance of the proposed sustainable framework for optimal and flexible design under uncertainty based on the analysis of exergy, energy, economic and environmental aspects (4E).

3. Framework description

The sustainable framework for optimal and flexible design under uncertainty in separation processes based on the 4E aspects is illustrated in the Figure 1. The framework employs the principles of stochastic programming using wait-and-see strategy, and simulation-optimization approach. The systematic steps of the framework can be summarized as follows: 1) Process simulation: rigorous thermodynamic description, process design, and state of the art update. 2) Stochastic modelling: mathematical strategies and formulation, uncertainty identification. 3) Optimizer: generation of a computational platform and algorithm selection to solve the problem, mathematical model integration. 4) Optimal design: Data analysis using graphical representations and statistical metrics as making-decision criteria for process design under uncertainty.

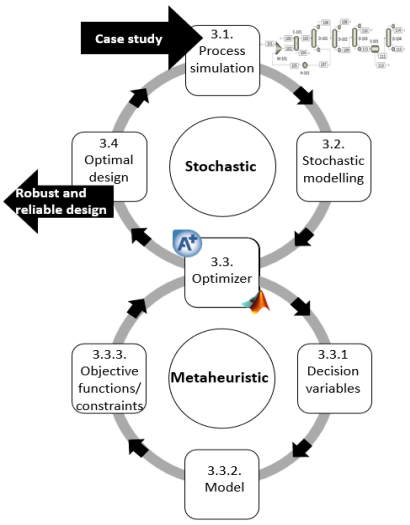


Figure 1. Visual representation of the sustainable framework.

3.1. Process simulation

Aspen plus was employed for the design of the separation process considering heuristic rules, which consists in hybrid arrangement, composed by a liquid-liquid column (C-101), followed by four conventional distillation columns (D-101, D-102, D-103, and D-104) focused on the recovery and purification of the extracting agent, acetone, ethanol, and butanol, respectively. The thermodynamic model NRTL and the equation of state Hayden-O'Connell were selected to describe the phase equilibrium (Ponce-Rocha *et al*, 2021). This work does not used surrogate models, the generated simulation was used directly in the optimization section to calculate the objective functions values.

3.2. Stochastic modelling: mathematical formulation

The MOOUU was formulated as follow [Eq. (1)]:

$$\underset{x}{\text{Min}}\ \underline{Z} = \left[Z_k\left(\underline{x}, \theta^*\right)\right], \text{s.t. } h\left(\underline{x}, \underline{\theta}^*\right) = 0;\ g\left(\underline{x}, \underline{\theta}^*\right) \leq 0;\ \underline{x} \in X, \underline{\theta}^* \in \Theta^* \tag{1}$$

Where $\underline{Z}$, describes the vector of k objective functions under uncertainty (Exergy, energy, economy, and environment), $\underline{x}$ represents the vector of n-dimensional states linked with structural and operational conditions (FA: extracting agent mass flow, NS: number of stages, FS: feed stage, RR: reflux ratio, DF: distillate flow, or BF: bottom flow), $\underline{\theta}^*$ is the vector of s-dimensional uncertainties present in the selected variables. **h** & **g** represent the vectors of equality and inequality constraints associated with operational and structural limitations.

3.2.1 Uncertainty representation & sampling

Latin hypercube sampling method was used for the generation of seventy-five random scenarios, all variations in the composition vectors for the feed stream (100) were generated simultaneously obtaining the following sample space: acetone (A) 2,275±112 kg/h, butanol (B) 5,931±276 kg/h, ethanol (E) 1,592±135 kg/h, water (W) 10,479±498 kg/h, and carbon dioxide (CO_2) 82±5 kg/h. The boundaries of the sampling in each flow value were obtained and adapted from previous experimental results (Valdez-Vazquez *et al.*, 2015).

3.2.2 4E aspects: Exergy, energy, economic, and environmental analysis

The 4E analysis was reported as a tool to analyse and optimize processes, which allowed improving indicators of life cycle impact assessment (LCIA) (Ponce-Rocha *et al*, 2021). 4E analysis was integrated as a tool and each component was added in the objective function employing the following mathematical equations.

3.3. *Exergy*

The exergy losses are a thermodynamic metric to detect the inefficient energy use in any process due to system irreversibility. The exergy losses balance was used on the process design considering heating and cooling processes (*Q*), separation (*Sep*) and concentration changes (*Conc*), over all process streams and equipment [Eq. (2)].

$$\sigma T_0 = \Delta \dot{E}x_{In} - \Delta \dot{E}x_{Out}$$

$$\sigma T_0 = \sum_{j=1}^{n}\sum_{i=1}^{m}\left[\Delta \dot{E}x_{Q,j} + \Delta \dot{E}x_{Sep,i} + \Delta \dot{E}x_{Conc,i}\right]\ [kW] \tag{2}$$

3.4. *Energy*

The ratio between total heat requirements and the total amount of purified products (ER) is a common metric to evaluate the performance of the process separation, mainly in fuel recovery [Eq. (3)].

$$ER = Energy\ requeriments\ \ [MJ]/Products\ [kg] \tag{3}$$

3.5. *Economy*

Total annual cost (TAC) is a typical economic indicator that allows to relate the capital cost, payback period (set at five years), and annualized cost of utilities, as described in Eq. (4).

$$TAC = Capital\ cost/Payback\ period + Utilities \tag{4}$$

3.6. *Environment*

The E-factor (E_{factor}) was used to evaluate the environmental impact, due to includes the waste and products ratio [Eq. (5)].

$$E_{factor} = Waste\ \ [kg]/Products\ [kg] \tag{5}$$

3.7. Optimizer: Multi-Objective Genetic Algorithm

To solve each optimization scenario, the matlab suite: gamultiob (a variant of NSGA-II) was used in the multi-objective optimization. Each optimization considered 50 generations, 200 individuals, and a crossover fraction of 0.8 and an adaptative feasible mutation function. In addition, each design specification vector generated in the multi-objective genetic algorithm is evaluated into a defined simulation scheme, to validate the thermodynamic feasibility and estimate the values of the objective functions through a computational interface. The mass purity (P) and mass recovery (R) constraints are specified for Eq. (6) & (7).

$$[P_{HEP}, P_A, P_B, P_E] \geq [0.999, 0.950, 0.995, 0.900] \tag{6}$$
$$[R_{HEP}, R_A, R_B, R_E] \geq [0.999, 0.995, 0.000, 0.992] \tag{7}$$

3.8. Optimal design under uncertainty: Making decision criteria

The results obtained by each scenario generate a distribution of optimal designs, therefore the use of statistical indicators was employed as a tool for the selection and making decisions to obtain a single design derived from MOOUU for each evaluated scenario. In this sense, three criteria were evaluated: mean, mode and a value of 80 % for cumulative distribution function (CDF), due to provides the basis for a probabilistic interpretation of the results.

4. Results

Each optimal design specification for each scenario under uncertainty was evaluated in the process configuration, the optimal results were represented using histograms plots, and a cumulative distribution function (see Figure 2 a and b, respectively). The statistical metrics were remarked in CDF plot, identified as follow: mean (blue dots),

mode (red dots), and 80 % of CDF (green dots). The structural designs of the process are reported in Table 1, where previous results without considering uncertainties are also illustrated (nominal design) (Ponce-Rocha *et al*, 2021).

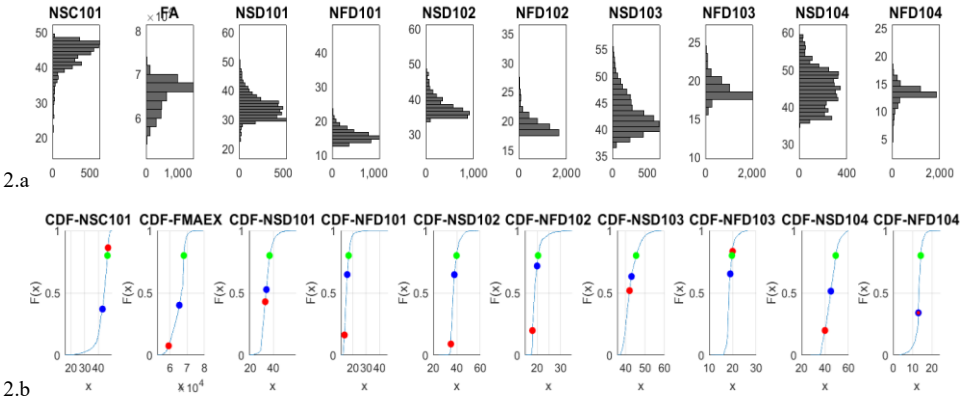


Figure 2. a) Histogram plots for structural specifications; b) Cumulative distribution function for structural specifications, mean (blue dot), mode (red dot), and 80 % of CDF (green dot).

Table 1. Comparative structural designs for each design and operating variables.

Variable	N_S C_{101}	F_A kg/h	N_S D_{101}	N_F D_{101}	N_S D_{102}	N_F D_{102}	N_S D_{103}	N_F D_{103}	N_S D_{104}	N_F D_{104}
Nominal	41	64,332	33	18	38	21	41	18	41	7
Mean	47	59,272	33	14	35	18	44	20	40	13
Mode	43	65,451	34	16	38	20	45	19	45	13
80 % CDF	47	68,080	37	17	40	20	46	20	49	14

To identify the most robust and reliable specifications of each process configuration, the obtained designs reported in Table 1 were evaluated using three flowrate vectors of the samples: minimum (Min), average (Avg), and maximum (Max) values. The results of the objective functions are reported in Table 2. These results show that the use of nominal and mean designs did not satisfy the requirements when the maximum flowrate vector is employed, in contrast to the designs using the mode and 80 % CDF value, which are robust enough to be able to fulfil the separation and purification requirements for the most extreme uncertain conditions.

Table 2. Evaluation of feed flow to SPS specifications, and objective functions values.

Design parameter	**Flow**	**TAC (USD/year)**	**Exergy (MW)**	**Energy (MJ/kg)**	E_{factor}
Nominal	Min	7.88	1.91	12.36	0.24
	Avg	8.05	1.93	11.83	0.25
	Max	-	-	-	-
Mean	Min	**7.63**	**1.85**	12.01	0.24
	Avg	7.80	1.87	11.51	0.25
	Max	-	-	-	-

Design parameter	Flow	TAC (USD/year)	Exergy (MW)	Energy (MJ/kg)	E_{factor}
Mode	Min	7.94	1.92	12.44	0.24
	Avg	8.12	1.95	11.91	0.25
	Max	8.42	2.03	**11.68**	**0.24**
80 % CDF	Min	8.11	1.95	12.65	0.24
	Avg	8.29	1.98	12.08	0.25
	Max	8.59	2.06	11.84	0.24

5. Conclusion

The proposed systematic framework allowed to identify that the use of statistical indicators as mode and 80 % CDF value, guarantee the convergence of each evaluated flow vector. Both indicators could be employed as metric to making-decision criteria to design separation process under uncertainty, however, the use of 80 % CDF is considered as a better criterion, due to using probabilistic fundamentals instead of a major value frequency in a set of data values, where the propagated uncertainty is presented. Additionally, it is necessary to evaluate each feed flow vector generated on the 80 % CDF design with the aim to validate the flexibility of the selected design specifications versus the specified feed flow variation.

Acknowledgments

Jaime D. Ponce-Rocha acknowledges the National Council of Science and Technology (CONACyT) for the scholarship received to develop this project (731399). Dr. Morales-Rodriguez acknowledges the partial financial support by the Mexican Bioenergy Innovation Centre, Bioalcohols Cluster (249564) and the Universidad de Guanajuato for the development of this project.

References

T. A. Adams, T. Thatho, M. C. Le Feuvre, & C. L. E. Swartz, 2018, The Optimal Design of a Distillation System for the Flexible Polygeneration of Dimethyl Ether and Methanol Under Uncertainty, Frontiers in Energy Research, 6, 41, 1-19.

S. Avraamidou, S. G. Baratsas, Y. Tian, & E. N. Pistikopoulos, 2019, Circular Economy - a challenge and an opportunity for Process Systems Engineering, Computers & Chemical Engineering, 133, 2, 1-48.

J. M. Gozálvez-Zafrilla, J. C. Gracía-Díaz, & A. Santafé-Moros, 2021, Uncertainty quantification and global sensitivity analysis of continuous distillation considering the interaction of parameter uncertainty with feed variability, Chemical Engineering Science, 235, 116509.

J. D Ponce-Rocha, M. Picón-Núñez, A. Carvalho, A. Santos, F. I. Gómez-Castro, & R. Morales, 2021, A design and optimization framework for (bio-) chemical process based on exergo-economic and environmental aspects, Computer Aided Chemical Engineering, 50, 1897-1902.

I. Valdez-Velazquez, M. Pérez-Rangel, A. Tapia, G. Boitrón, C. Molina, G Hérnandez, & L. Amaya-Delgado, 2015, Hydrogen and butanol production from native wheat straw by synthetic microbial consortia integrated by species of Enterococcus and Clostridium, Fuel, 159, 214–222.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50200-1

Application of nonlinear surrogate models on optimization of carbon capture and utilization network

Wonsuk Chung and Jay H. Lee*

[a]*Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, REPUBLIC OF KOREA*
jayhlee@kaist.ac.kr

Abstract

Carbon capture and utilization (CCU) can solve recent global climate change, allowing large-scale mitigation of carbon emission. Estimation of cost and carbon reduction for CCU systems is an essential step as most of them are energy-intensive and costly, hindering further adaptation. Nonlinear surrogate models can help identification of sustainability for each CCU pathway without rigorous simulations. The surrogate model as a simple equation allows being incorporated into multiscale decision-making. This work applies nonlinear surrogate models on decision making of CCU through the construction of a CCU superstructure systemization of the best CCU pathway selection problem as mathematical programming. The surrogate models that predict the capture energy and cost from the flue gas condition and CO_2 removal rate are incorporated into the network. A case study as a small CCU superstructure is performed to test ability and complexity of the CCU superstructure optimization problem as mixed-integer-nonlinear-programming (MINLP).
Keywords: Carbon capture and utilization; Superstructure; Network optimization; Surrogate model; Mined-integer nonlinear programming.

1. Introduction

Carbon capture and utilization (CCU) is a promising option for global climate change by capturing a large amount of CO_2 from plantwide flue gases and utilizes it to useful products. CCU attracts industries for its potential economic benefits in addition to reduction of carbon emission. A number of researches have been studied in carbon capture as well as carbon utilization in both technology level and process level to reduce the energy consumption and cost (Boot-Handford et al, 2014; Cuellar-Franca and Azapagic, 2015). Despite the efforts, many CCU pathways from flue gas to capture, conversion, and product are not sustainable in terms of economics and carbon reduction. The major concern in industries is on which technologies should they invest in to achieve best expected carbon reduction and profit.

A systemic approach that analyzes multiple CCU technologies as mathematical programming is extensively studied to solve the concern. CCU superstructure where the CCU pathway is identified and evaluated is constructed so that mathematical solver can optimize the superstructure and find out the best CCU pathway (Kim et al, 2013; Bertran et al, 2017; Roh et al, 2019). Those works formulated superstructure optimization

problems as mixed-integer linear programming (MILP) and a pathway is represented as a set of binary integers.

The MILP approach has a limitation on flexibility in process level. The CCU processes are simulated before superstructure construction and the key performance indicators of the processes such as CO_2 capture rate and capture energy are treated as constant parameters. As a chemical process, a CCU process is nonlinear and the performance indicators change not linearly as the process input changes. To reflect this nonlinearity, the superstructure optimization problem must be formulated in mixed-integer nonlinear programming (MINLP), the most difficult type of problem. As far as our knowledge, there is little attempt to solve the problem in MINLP. Hasan et al (2015) incorporated nonlinear cost models of CO_2 capture processes on a CCU supply chain network and solved them as MILP with piecewise linearization. As direct incorporation of full process models in the systemic framework is not realistic, it is essential to apply the nonlinear cost models or surrogate models instead, and it is worthy to investigate the problem complexity of the superstructure optimization problem as MINLP.

In this work, a CCU superstructure is formulated with known surrogate models for capture processes from literature. Then the MINLP problem is formulated and solved to identify the feasible pathways. Several industrial sectors with a different number of CO_2 sources are selected and optimized in several scenarios to test the complexity of the optimization problem.

2. Optimization problem formulation

A superstructure model is defined as a network model and the embedded process interval model. The network model formulates the topology of the superstructure as a graph, treating the CO_2 source, processing interval, and products as nodes and their connections as directed arcs. The processing interval model is a set of equations where the key performances of the CCU processes are captured as interval parameters and the process cost can be calculated from the interval flowrates. The decision variable for the superstructure optimization problem is binary variable z which represents the selection of the nodes and their connections. Inside N, the user can set constraints on z using binary parameters y to prevent the selection of undesired connections as Eq (1, 2)

$$z_v \le y_v \quad \forall v \in \mathbb{v} \tag{1}$$

$$z_a \le y_a \quad \forall a \in \mathbb{v} \times \mathbb{v} \tag{2}$$

where $\mathbb{v}$ is the set that contains the feeds, intervals, and products, v is the node, and a is the arc. The feed-feed connection and product-product connection are prevented by y (e.g. $y_{f,f_2} = 0 \; \forall f, f_2 \in \mathbb{f}$, where $\mathbb{f}$ is the set of feeds). The logical constraints that $z = 1$ if and only if the flow is larger than zero is described by the big M method as Eq (3 and 4).

$$F_v \le z_v M \quad \forall v \in \mathbb{v} \cup \mathbb{v} \times \mathbb{v} \tag{3}$$

$$z_v \le F_v M \quad \forall v \in \mathbb{v} \cup \mathbb{v} \times \mathbb{v} \tag{4}$$

where F_v is the corresponded flow to z_v.

The interval models consist of raw material mixing, reaction, and separation steps (e.g. Bertran et al, 2017). The previous MILP approach only allows linear equations for all three steps with constant parameters. Instead, nonlinear surrogate models can be incorporated.

Appropriate labeling on each arc can handle the assignment of outlets after separation (Quaglia et al, 2012). The connections are labeled so that an outlet can only be assigned to one of the connections with the same label. These allocations are modeled by Eq (5 and 6)

$$\sum_{v_2 \in \mathbb{v}} z_{v,v_2} y_{v,v_2,n} \leq z_v \quad \forall v \in \mathbb{v}, n \in \mathbb{m} \tag{5}$$

$$\sum_{v_2 \in \mathbb{v}} y_{v,v_2,n} \leq z_v \quad \forall v \in \mathbb{v}, n \in \mathbb{m} \tag{6}$$

where $\mathbb{m}$ is the set of labels for arc and the labeling is determined by binary parameter $y_{v,v_2,n}$. For example, in Figure 1, Interval A has two outlets after separation. The first outlet is labeled as 1 and the second outlet is labeled as 2. The interval is connected with four intervals B to E, where the first outlet can only flow through connection A-B or A-C. Likewise, the second outlet can only flow through connections A-D and A-E.

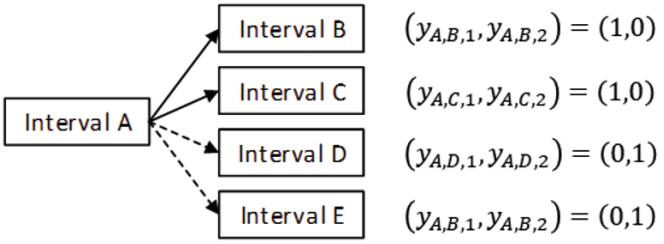


Figure 1. Labeling of arcs in the superstructure. The connections between A-B and A-C are labeled 1 (solid line) and the connections between A-D and A-E are labeled 2 (dashed line).

The objective function is to maximize the gross profit (GP) with nonnegative CO_2 reduction (E^{Red})

$$\max_z GP(z,F) \text{ s.t. } E^{\mathrm{Red}}(z,F) \geq 0, I(F,\lambda) = 0 \text{ and } N(z,F,y) \geq 0 \tag{7}$$

with process interval model (I) being equality constraints with the interval parameters (λ) and network model (N) being inequality constraints so that GP and E^{Red} can be calculated from the interval flowrates (F).

The multiple pathways can be identified using integer cuts (Kim et al, 2013). The optimization problem finds an optimal pathway in the superstructure. The next optimal pathway is identified by adding additional logical constraints as Eq (8)

$$\sum_v y_{p,v} z_v \leq \sum_v y_{p,v} - 1 \quad \forall p \in \mathbb{p} \tag{8}$$

where $\boldsymbol{y_{p,v}}$ is a binary parameter as the previous selection result of interval $\boldsymbol{v}$ for pathway $\boldsymbol{p}$ in pathway set $\mathbb{p}$, and $\boldsymbol{z_v}$ is a binary variable as a selection of interval $\boldsymbol{v}$ for the newly

optimal pathway. After the optimization, the newly identified pathway is stored by expanding $\mathbb{p}$. This integer cut can be repeated until no feasible pathway is identified.

3. Case study

A case study is performed for the refinery (Figure 2). The superstructure has four flue gases with different flowrates and CO_2 mole fractions from 0.04~0.5% mol. (furnace off-gas, utility plant off-gas, fluidized catalytic cracker (FCC) off-gas, and hydrogen processing unit (HPU) off-gas, Table 1). Three capture processes are considered: monoethanol (MEA) based amine scrubbing, piperazine (PZ) based amine scrubbing, and membrane separation. Three utilization intervals are considered: combined reforming (CR) process that produces syngas (H_2:CO=2) from CO_2 CH_4, and H_2O, CO_2 hydrogenation (CH) process that produces methanol from CO_2 and H_2, and methanol synthesis (MeOHsyn) process from syngas from CR. One dummy interval (Dummy) is added that collects waste gas from intervals. Three products are considered: it is assumed that CO_2 can be directly sold with \$100/ton, methanol with \$293/ton, and waste gas with zero price. The superstructure has eight compounds (CO_2, N_2, O_2, H_2O, H_2, CO, MeOH, and CH_4) and three compounds (H_2, CH_4, H_2O) are treated as raw materials in utilization intervals.

Table 1. Summarization of flue gases in the case study.

	Mass flowrate (ton/sec)	Mass fraction CO_2	N_2	O_2	H_2O
Furnace off-gases	0.160	12.3%	73.5%	7.8%	6.3%
Utility plant off-gas	0.237	6.2%	74.4%	13.6%	5.7%
FCC off-gas	0.035	28.2%	69.1%	2.1%	0.6%
HPU off-gas	0.035	61.1%	38.9%	0.0%	0.0%

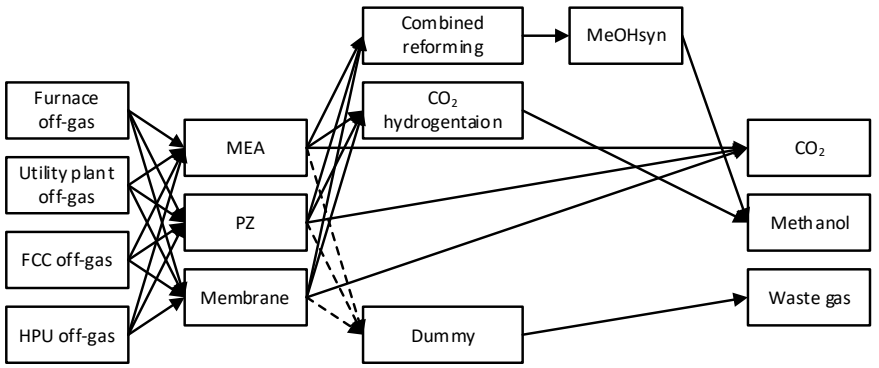


Figure 2. A CCU superstructure with four flue gases, three capture processes, three conversion processes, and two products with dummy interval and waste gas. The solid line and the dashed line indicate labeling of arc as 1 and 2, respectively.

Chung et al (2020) suggests model formula as Eq (9) and Eq (10)

$$U_{\mathrm{k,steam}} = \alpha\left(\ln x_{CO2} + \beta \ln\left(1 - r_{\mathrm{k,CAP}}\right)\right) + f_{\mathrm{k,steam,0}}, k \in \{\mathrm{MEA, PZ}\} \quad (9)$$

$$\ln EPC_{\mathrm{k}} = \alpha\left(\ln x_{CO2} + \beta \ln\left(1 - r_{\mathrm{MEA,CAP}}\right) - 1/r_{\mathrm{k,CAP}}\right) + \gamma \ln F_{CO2,k} + f_{k,EPC,0}, k \in \{\mathrm{MEA, PZ}\} \quad (10)$$

for steam consumption ($U_{\mathrm{k,steam}}$) and equipment purchase cost (EPC_{k}) from inlet CO_2 mole fraction (x_{CO2}), flowrate of captured CO_2 ($F_{CO2,k}$), and CO_2 capture rate ($r_{\mathrm{k,CAP}}$) with model parameters ($\alpha, \beta, \gamma, f_0$). The model is valid for $0.03 \leq x_{CO2} \leq 0.5$. Hasan et al (2015) suggests model formula as Eq (11)

$$IC_{\mathrm{k}} \text{ or } OC_{\mathrm{k}} = \alpha + (\beta x_{CO2}^{n} + \gamma) F_{in}^{m} \quad (11)$$

for operating cost and capital cost from inlet CO_2 mole fraction (x_{CO2}) and feed gas flowrate ($F_{in,k}$) with model parameters ($\alpha, \beta, \gamma, m, n$). The model is valid for $0.01 \leq x_{CO2} \leq 0.7$. The interval data for utilization processes are assumed to be constant, referred from Roh et al (2019). The case study is formulated as MINLP in Pyomo and solved by BARON (Kilinc and Sahinidis, 2018) using AMD Ryzen 9 5950X.

4. Results and discussion

There are 30 feasible pathways in the superstructure. The pathways from HPU to PZ and MEA are infeasible as the CO_2 mole fraction of HPU exceeds 0.5. A test is performed to test whether the framework can identify all the feasible pathways. Without the constraint for nonnegative CO_2 reduction, all the 30 pathways are identified within 346.262 seconds. Then, with the nonnegative CO_2 reduction constraint, the six sustainable pathways are identified (Table 2). Pathway No. 7 is at the nonnegative CO_2 reduction constraint with small production rate of methanol and a small annual profit ($ 6.0M/yr). This pathway is originally the most profitable but non-CO_2-reducing, and the capture rate is adjusted so that the constraint is satisfied. A dummy pathway is identified (pathway No. 8) with a non-integer solution of z. This is due to the integer relaxation algorithm of BARON.

Table 2. The identified sustainable CCU pathways in the case study in order of identification by the solver. Furn means furnace off-gas. Pathway No. 7 and 8 are dummy pathways.

No.	Feed	Capture	Conversion-product	Profit (M$/yr)	CO_2 reduction ($MtCO_2$/yr)	Computational time (sec)
1	HPU	Membrane	CR-MeOH	785.5	3.361	15.663
2	FCC	PZ	CR-MeOH	391.1	1.078	1.419
3	FCC	MEA	CR-MeOH	389.2	1.058	16.457
4	FCC	Membrane	CR-MeOH	355.2	0.994	1.160
5	HPU	Membrane	CH-MeOH	162.8	0.019	1.346
6	HPU	Membrane	CO_2	45.0	0.671	1.817
7	Furn	PZ	CR-MeOH	6.0	0.000	30.383
8*	FCC	MEA	CR-MeOH	-4.0	0.000	18.229
No more feasible pathway is identified						17.165

Despite superstructure with small size (210 integer variables, 2710 continuous variables, and 2094 constraints), it takes several minutes to identify all the pathways in the case study. If more CO_2 sources, capture intervals, utilization technologies, and products are added, the size of the superstructure can be very large and the computational time may take hours or days. MILP approach takes only minutes to solve such a large superstructure (>100,000 variables) (Bertran et al, 2017). Other algorithms instead of solving MINLP directly (e.g. piecewise linearization) will reduce the computational time.

5. Conclusions

An optimization problem of a CCU system for the purpose of identifying sustainable pathways is formulated in MINLP. The nonlinear surrogate models for capture processes allow more accurate identifications. A case study is performed with four flue gases, three capture processes, and three utilization options to test the complexity of the problem. The results suggest that the MINLP approach can successfully identify the pathways; however, the computational times can be issues for large CCU superstructure.

References

M. Boot-Handford et al, 2014, Carbon Capture and storage update, *Energy and Environmental Science* 7, 130-189.

R. Cuellar-Franca and A. Azapagic, 2015, Carbon capture, storage and utilisation technologies: A critical analysis and comparison of their life cycle enrironmental impacts, *Journal of CO_2 Utilization* 9, 82-102.

J. Kim, M. Sen, and C. Maravelias, 2013, An optimization-based assessment framework for biomass-to-fuel conversion strategies, *Energy and Environmental Scienc*, 6, 1093-1104.

M. Bertran, R. Frauzem, A. Sanchez-Arcilla, L. Zhang, J. Woodley, and R. Gani, 2017, A generic methodology for processing route synthesis and design based on superstructure optimization. *Computers and Chemical Engineering* 106, 892-910.

K. Roh, A. Al-Hunaidy, H. Imran, and J. H. Lee, 2019, Optimization-based identification of CO_2 capture and utilization processing paths for life cycle greenhouse gas reduction and economic benefits, *AIChE Journal* e16580.

F. Hasan, E. First, F. Boukouvala, and C. Floudas, 2015, A multi-scale framework for CO_2 capture, utilization, and sequestration: CCUS and CCU, *Computers and Chemical Engineering* 81, 2-21.

A. Quaglia, B. Sarup, G. Sin, and R. Gani, 2012, Integrated business and engineering framework for synthesis and design of enterprise-wide processing networks, *Computers and Chemical Engineering* 38, 213-223.

W. Chung and Lee J. H., 2020, Input-output Surrogate Models for Efficient Economic Evaluation of Amine Scrubbing CO_2 Capture Processes, *Industrial and Engineering Chemistry Research* 59, 1895-18964.

M. Kilinc and N. Sahinidis, 2018, Exploiting integrality in the global optimization of mixed-teger nonlinear programming problems in BARON, *Optimization Methods and Software*, 33, 540-562

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50201-3

Systematic process energy optimization via multi-level heat integration: A case study on low-temperature reforming for methanol synthesis

Alexander Guzman-Urbina [a] *, Haruka Tanaka [a], Hajime Ohno [a], Yasuhiro Fukushima [a]
[a]*Department of Chemical Engineering, Graduate School of Engineering, Tohoku University, 6-6-07 Aramaki-Aza Aoba, Sendai, Japan*
guzman.urbina.alexander.a6@tohoku.ac.jp

Abstract

The development of new technologies for chemical process systems is essential to achieve the sustainable development goals. To ensure the path forward is viable, technology development includes a rigorous evaluation stage that contrasts the benefits of the proposed technology with existing or alternative systems. Energy efficiency and the extent of heat integration are key indicators that directly influence the evaluation and implementation of process system technologies. Methodologies such as the pinch analysis together with mathematical optimization have been extensively applied to determine the degree of heat integration potentially achievable by chemical processes. The aim of these methodologies is to find optimal heat exchange networks (HENs) within a plant or process by minimizing the use of external utilities for heating or cooling. However, since the mathematical formulation of the conventional optimization procedure excludes factors considered important in decision-making processes (operability of complex HEN, plant configurations, safety, etc.), process evaluations fail to fully reflect the consequences of technologies that alter the process conditions. In order to close this gap, an alternative methodology for process energy optimization is proposed in this study which ranks near-optimal HENs solutions, and also considers external factors for decision-making. The methodology proposed consists of heat integration over multiple levels (e.g., sub-process level, plant level) using two major methods of mathematical optimization: the Stage-Wise Superstructure (SWS) and the Integer-Cuts Constraint (ICC). Results from the application of the proposed methodology to a case study of a methanol production process show the possibility of reducing the external utilities consumption by more than 55% compared to the non-integration scenario.

Keywords: Heat Exchange Network, Energy Optimization, Heat Integration, Superstructure, Methanol Synthesis.

1. Introduction

Making a comprehensive evaluation of the impacts of a new technology over the whole process is one of the key factors in the success of technologies for chemical systems in real-life applications. Among the most important aspects to consider when evaluating these systems are the investment cost of the technology, operation costs, environmental impacts, safety, and energy efficiency. Heating requirements, cooling requirements, and heat exchange networks are determinants for all aspects of the evaluation, especially when assessing the sustainability of a new process technology.

Heat exchanger networks have the role of transferring energy among high-temperature streams that require cooling (hot process streams) and low-temperature streams that require heating (cold process streams) to reduce the consumption of external utilities for heating or cooling. Improving energy efficiency through effective heat utilization in process systems is known as "heat integration". The variations in the extent of heat integration influence significantly the evaluation of new technologies because utility reduction leads to reductions in cost and/or environmental impact. Methodologies like the “pinch analysis” serve as a target-based design tool for revealing the minimal utility consumption (MUC) potentially achieved. Actual process systems are rarely developed considering the MUC from initial stages. Moreover, conventional optimization procedures exclude factors considered important in the decision-making processes (operability of complex HEN, plant configurations, safety, etc.). Besides, given the complexity of actual process systems, HENs determination is commonly limited to the interaction among streams within a level of sub-processes and does not extend to the whole plant. While process evaluations based on designs without any HENs may deliver some insight, such evaluations typically fail to fully reflect the consequences of process technologies that alter process conditions. Therefore, this study introduces a scheme that systematically presents multiple near-optimal HENs solutions for interaction among multiple levels (at the sub-process level and plant level).

The framework proposed by this study consists of three major steps: segmentation of the process plant in functional sub-processes, an independent heat integration at the sub-process level, and a global heat integration at the plant level. The heat integration objective at both levels is to find a rank of combinations among process streams that minimize operational costs in terms of utility consumption. For that purpose, this study employs a novel methodology that consists of a combination of two major methods of mathematical optimization. One method commonly used for HEN synthesis, known as the Stage-Wise Superstructure (SWS) (Yee & Grossmann, 1990), considers heat exchange over multiple stages, and another method used to rank the synthesized HENs, proposed by Maronese S. et al. (2015), is called the Integer-Cuts Constraint (ICC). Heat integration is performed as a bottom-up hierarchical sequence, from the sub-process level to the plant level. Here, one of the near-optimal solutions is selected for each sub-process according to a certain design criterion (safety, heat exchanger size, network configuration, etc.), and then taken for further integration at the plant level.

2. Background on heat integration and optimization

2.1 Pinch analysis

The pinch analysis is a method developed to reveal the minimal utility consumptions potentially achieved (Linnhoff & Hindmarsh, 1983). It reveals the utility consumptions when heat exchange in a process system is performed to the maximum. In this method, the heat loads of all existing streams over any given temperature range are added together in what is referred to as a "composite curve". When performing heat exchange, some temperature difference is required as a driving force. The minimum value of the temperature difference, the minimum approach temperature ($\Delta T_{\min}$), is often set empirically based on the industrial field or the combination of streams (Tatsumi & Matsuda, 2002). The maximum amount of heat exchange between process streams can be known graphically by superposing the hot composite curve and the cold composite curve separated by $\Delta T_{\min}$. However, the pinch analysis provides only the maximum value of the heat exchangeable in the processing system and the minimum value of the utility

amounts achieved. The pinch analysis does not determine how hot and cold streams should be paired or to what extent heat should be exchanged. Thus, it is necessary to implement optimization methods that explicitly synthesize the heat exchanger networks based on operational and investment cost criteria.

2.2 Mathematical optimization

The mathematical optimization methods employed for HEN design usually adopt the concept of superstructure optimization (Fig. 1). Superstructure is defined as a structure that contains all possible structures as its substructures. By performing optimization, a structure containing some substructures which has the best objective function value can be obtained. By using superstructure in heat exchanger network synthesis, it is possible to know the combination of streams to exchange the heat and the amount of each heat load explicitly. There are basically two approaches to the heat exchanger network synthesis using mathematical optimizations (Escobar, M. & Trierweiler, 2013). One is to perform optimization multiple times (Papoulias & Grossmann, 1983; Floudas et al. 1986), and the other is to perform all optimizations at once (Ciric et al., 1991; Yee & Grossmann, 1990). In the former approach, the utility consumptions are first determined by minimizing the utility cost. Next, the heat exchange area is determined using the objective function of minimizing the investment cost or the sum of heat exchanger area. In the latter approach, however, the quantity of utilities, the combination of streams to be heat-exchanged, each heat load, and each heat exchanger area are determined by a single optimization. In this case, the objective function is to minimize the total annual cost of the plant.

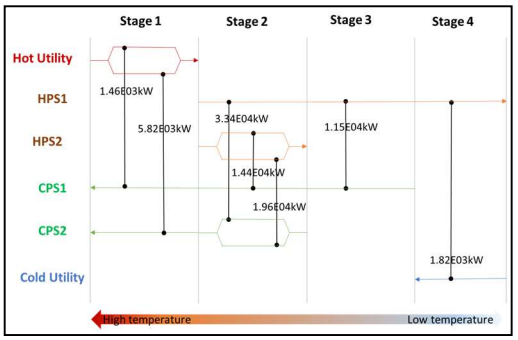


Figure 1. HEN Superstructure (4 Stages)

By employing mathematical optimization, only a single optimal solution can be obtained. However, whether the solution obtained can be implemented, or even to determine if it is actually optimal, remains unknown. This is because the introduction of heat exchanger networks is often subject to factors externalized from the mathematical formulation of the design problem, such as operability, plant configuration, and safety. These factors need to be considered when deciding on whether to introduce the heat exchanger network obtained by optimization.

3. Method for process energy optimization via multi-level heat integration

The purpose of the present study was to develop a tool that systematically presents multiple heat exchanger networks for a process system. Among the methods reported which show multiple solutions in mathematical optimization, none are applied to synthesize multiple HENs. Therefore, we developed a model to synthesize multiple HENs with the intention of providing decision-makers with multiple options (near-optimal solutions) to choose from when selecting a heat exchanger. Also, in the event that there are common characteristics in the near-optimal solutions, the common characteristics highlight what should be achieved with priority when implementing the heat exchanger network. The procedure of this study is illustrated schematically in Fig. 2:

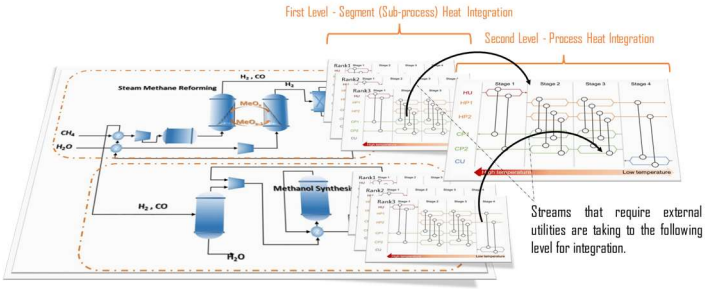


Figure 2. Optimization procedure via multi-level heat integration

The method proposed employs mathematical optimization for multiple heat exchanger networks. We adopted a method proposed by Yee & Grossmann (1990) that considers heat exchange over multiple stages. This method is called the "Stage-Wise Superstructure" (SWS). A conceptual diagram of the SWS over four stages is shown in Fig. 1. This diagram presents a structure with two hot process streams (HPs), two cold process streams (CPs), one hot utility (HU), and one cold utility (CU). The solution of this SWS is a combination of heat exchangers that minimizes the total costs and satisfies the thermodynamical constraints of heat transfer. The objective function is the minimization of the total annual cost related to heat exchange (C_{total}). The first and second terms in Eq. (1) are the operation costs, which represent the cold utility costs and hot utility costs, respectively. The third and fourth terms are the investment and operational costs.

$$\min\ C_{\text{total}} = \left(\sum_{i\in HP}\sum_{j\in CU}\sum_{k\in K} C_{\text{U},j}\cdot q_{i,j,k} + \sum_{i\in HU}\sum_{j\in CP}\sum_{k\in K} C_{\text{U},i}\cdot q_{i,j,k}\right) + F_{\text{A}}\left(\sum_{i\in H}\sum_{j\in C}\sum_{k\in K} C_{\text{F},ij}\cdot y_{i,j,k} + \sum_{i\in H}\sum_{j\in C}\sum_{k\in K} C_{\text{A},ij}\cdot a_{i,j,k}\right) \tag{1}$$

Where $C_{\text{A},ij}$ represents the area cost coefficient for heat exchangers, $C_{\text{F},ij}$ is the fixed cost for heat exchangers, and C_{U} is the annual cost per unit of utility, respectively. F_{A} is the annualization factor for investment, $a_{i,j,k}$, $q_{i,j,k}$, and $y_{i,j,k}$ represent the variables for area, heat load, and exchanger existence respectively. The subscript *I, j,* and *k* indicates the hot streams, the cold streams, and number of stages, respectively.

3.1 Optimization using the Integer-Cut Constraint method

The Integer-Cut Constraint (ICC) is a suitable method for obtaining multiple solutions in optimization problems of system synthesis (Maronese et al. 2015). For this study, the ICC method was adopted to perform system synthesis of the energy system with rank order. In the ICC method, after obtaining the optimal solution, a constraint is added to the original problem, then it prohibits the identical combination of variables as the already obtained solution (Eq. 2).

$$\sum_{i\in H}\sum_{j\in C}(2y^{n}_{i,j,k}-1)y_{i,j,k} \le \left(\sum_{i\in H}\sum_{j\in C} y^{n}_{i,j,k}\right) - 1 \qquad \forall\, n = 1,\dots N \tag{2}$$

Where $y^{n}_{i,j,k}$ is the value of $y_{i,j,k}$ of the n-th solution, N is the number of obtained solutions, *i* represents the heat transfer operation, y_s is binary variable representing the existence of

the unit for heat transfer ($y_s = 1$: exist, $y_s = 0$: not exist), and k is the number of optimizations. After this procedure, the HEN optimization is performed again to obtain the next best solution, which has a different combination of units from the previous solutions. In summary, the overall procedure proposed by this study consists of three major steps: an initial segmentation of the process plant in functional sub-processes, an independent heat integration at the sub-process level applying the combination of SWS-ICC, and, a global heat integration at the plant level. The heat integration is then performed as a bottom-up hierarchical sequence, from the sub-process level to the plant level. Here, one of the near-optimal solutions is selected for each sub-process according to design criteria and then taken for further integration at the plant level.

4. Implementation for Methanol Synthesis

To demonstrate the potential relevance of the scheme developed, the method was applied to the task of assisting in the development of a large-scale methanol production plant (50 kt/day) that employs a sub-process of low-temperature methane reforming. The plant was segmented primarily into six major subprocesses: a looping system, hydrogen production, syngas compression, methanol synthesis, combustion and steam generation, and methanol purification. Table 1 provides a sample of the input data for the looping system. Here, the minimum approach temperature ($\Delta T_{\min}$) was 5 K and the number of stages was set to four. The cost parameters were the area cost coefficient for heat exchangers ($C_{\mathrm{A},i,j}$) and fixed cost for heat exchangers ($C_{\mathrm{F},i,j}$), set at 350 $\$/\mathrm{m}^2$ and \$10000, respectively.

Table 1. Sample of input data for the looping system

Stream	Type	Supply Temp. [K]	Target Temp. [K]	Flowrate $[\mathbf{kW/K}]$	$\boldsymbol{h}$ $[\mathbf{kW/(m^2K)}]$
CH_4F- CH_4I	Cold	298.15	573.15	99.80	0.8
Water-Steam	Cold	417.15	573.15	377.68	0.8
FR1-FR2	Hot	571.75	427.15	324.11	0.8
SR1-SR2	Hot	572.85	465.55	317.21	0.8

4.1 Results and Discussion

Figures 3a and 3b show a summary of the results obtained for the methanol case. As shown in Fig. 3a, the amount of total utility consumption achieves a reduction of 55% compared to the case when heat integration was not performed. In addition, although the empirical integration achieves a significant reduction of utilities consumption in comparison to a non-integration case, further reduction is viable by applying superstructure modeling. Fig. 3b presents a ranking of solutions for the heat integration of the looping system in terms of the transition of the objective function value. For this sub-process, the total annual cost of the first rank was 3.287 M\$/year, and that of the 10^{th} rank was 3.298 M\$/year. In the optimal solution. In the optimization using SWS, it is possible to explicitly provide the configuration of the heat exchanger network and the values of each variable. It is possible to determine whether the solution can be actually implemented considering the factors externalized from the mathematical formulation.

5. Conclusions

In this study, a tool that presents multiple heat exchanger networks systematically was developed. By combining a method to obtain multiple solutions with different combinations of binary variables with rank order (ICC method) and a method to obtain

the range of variables keeping the objective function value within a specified range and modifying them, heat exchanger networks were systematically synthesized for multiple system levels. Using this model, even if the single optimal heat exchanger network determined by optimization cannot be implemented due to factors not included in the mathematical formulation of the design problem, a heat exchanger network can be selected from a range of choices reflecting the intentions of the decision makers. Additionally, the common characteristics of the higher ranked solutions will help in the decision-making process by highlighting what should be achieved with priority in the actual implementation of heat exchanger networks.

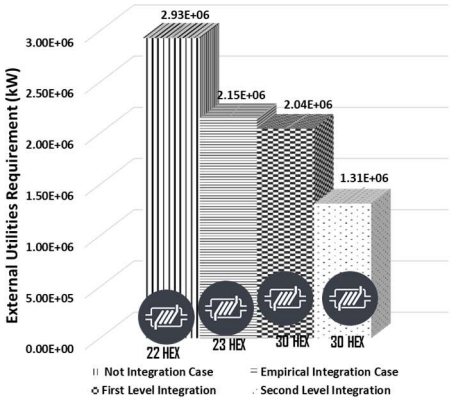


Figure 3a. Total Utilities Requirement

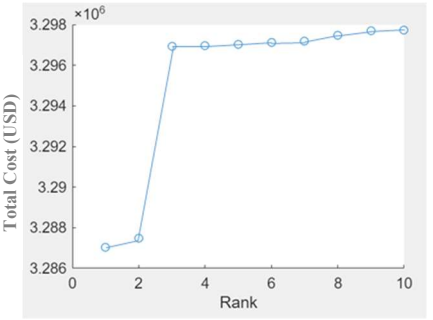


Figure 3b. Ranking for Looping sys.

Acknowledgements

This work was supported by The Japan Science and Technology Agency (JST) [Grant: JPMJMI17E4], and New Energy and Industrial Technology Development Org. of Japan NEDO.

References

Chen, Y., Eslick, J.C., Grossmann, I.E. and Miller, D.C., 2015. Simultaneous process optimization and heat integration based on rigorous simulations. Comput. Chem. Eng., 81, pp.180-199.

Ciric, A. R. & Floudas, C. A. Heat exchanger network synthesis without decomposition. Comput. Chem. Eng. 15, 385–396 (1991).

Escobar, M. & Trierweiler, J. O, 2013. Optimal heat exchanger network synthesis: A case study comparison. Appl. Therm. Eng. 51, 801–826 (2013).

Floudas, C. A., Ciric, A. R. & Grossmann, I. E. Automatic synthesis of optimum heat exchanger network configurations. AIChE J. 32, 276–290 (1986).

Linnhoff, B., Hindmarsh, E., 1983. The pinch design method. Chem. Eng. Sci. 38, 745–763.

Maronese, S., Ensinas, A. V., Mian, A., Lazzaretto, A. & Maréchal, F., 2015. Optimum Biorefinery Pathways Selection Using ICC. Ind. Eng. Chem. Res. 54, 7038–7046.

Papoulias, S. A. & Grossmann, I. E. A structural optimization approach in process synthesis—II: Heat recovery networks. Comput. Chem. Eng. 7, 707–721 (1983).

Tatsumi H., Matsuda K., 2002. Pinch Technology Energy Conservation Analysis Method and Practice. Energy Conservation Center.

Yee, T.F. and Grossmann, I.E., 1990. Simultaneous optimization models for heat integration—II. Heat exchanger network synthesis. Comput. Chem. Eng., 14(10), pp.1165-1184.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50202-5

Bayesian Optimization for Automobile Catalyst Development

Sanha Lim, Hwangho Lee, Shinyoung Bae, Jun Seop Shin, Do Heui Kim and Jong Min Lee*

School of Chemical and Biological Engineering, Institute of Chemical Processes, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea
jongmin@snu.ac.kr

Abstract

In this study, we propose an efficient computational methodology for developing Selective catalytic reduction (SCR) with high NOx conversion and resistance to hydrothermal aging, using Bayesian optimization (BO). In order to focus on the catalytic performance at low temperature, Cu-Fe bimetallic catalyst supported by SSZ-13 (Si/Al = 12) is targeted. An initial surrogate model is constructed by referring experimental data from previously published papers. The next sampling points are determined from the Bayesian optimization algorithm. NOx conversion is observed under fresh condition and hydrothermally aged condition after manufacturing a catalyst sample consisting of suggested metal compositions. We also consider the catalytic activity after hydrothermal aging in the air of 900 ℃ containing 10 % water for 16 hours. In this way, the optimal composition for bimetallic SCR catalyst is discovered, maximizing NOx conversion and hydrothermal resistance in only a few steps of experimentation. The proposed SCR catalyst can reduce 95.86 % of nitrogen oxides at 250 ℃. After hydrothermal aging, it can eliminate 88.83 % of nitrogen oxides at the same temperature.

Keywords: Bayesian optimization (BO); Selective catalytic reduction (SCR); Activity; Hydrothermal aging

1. Introduction

Diesel engines typically show higher fuel efficiency than gasoline engines, but they emit more air pollutants, such as nitrogen oxides (NOx). The emission standards for nitrogen oxides become stringent recently to decrease air pollutants emitted from automobile exhaust gas. Euro 6 regulations have already been implemented since 2014 and more strict restriction, Euro 7, is to be introduced in the near future. Selective catalytic reduction (SCR) with urea injector is a standard after-treatment system which reduces the nitrogen oxides in the exhaust gas emitted from the diesel engines. The performance of the SCR catalyst is excellent at high temperature, removing more than 90 % of nitrogen oxides in the exhaust gas. However, catalytic activity in low temperature range is a major challenge in the automobile aftertreatment catalyst field. Low NOx conversion at low temperature causes air pollution when the exhaust gas is not hot enough, for example, during cold start and on travelling short distances (Shan and Song, 2015). Furthermore, for the actual diesel engine exhaust system, the SCR catalyst suffers from the deactivation caused by hydrothermal aging due to H_2O content with high temperature exhaust gas. Therefore, achieving hydrothermal stability is also crucial for commercial SCR catalyst.

For commercial SCR for diesel vehicle, urea solution is usually applied to generate ammonia. The hot exhaust gas stream from diesel engine evaporates water droplet and thermally decomposes urea to ammonia.

$$(NH_2)_2CO \longrightarrow NH_3 + HNCO \quad (1)$$

$$HNCO + H_2O \longrightarrow NH_3 + CO_2 \quad (2)$$

The generated ammonia is adsorbed on the active sites of the catalyst to remove nitrogen oxides from the exhaust gas. The SCR reactions are usually described in three types of reactions: the standard, fast, and slow SCR reactions. (Pant and Schmieg, 2011)

$$4NH_3 + 4NO + O_2 \longrightarrow 4N_2 + 6H_2O \quad (3)$$

$$2NH_3 + NO + NO_2 \longrightarrow 2N_2 + 3H_2O \quad (4)$$

$$4NH_3 + 3NO_2 \longrightarrow 3.5N_2 + 6H_2O \quad (5)$$

Nitrogen monoxide is oxidized to nitrogen dioxide at high temperature (above 250℃). Since the extent of NO oxidation to NO_2 under low temperature is small, NO_2 concentration at the SCR inlet gas can be negligible (Olsson et al., 2008).

Many SCR catalysts have been explored to improve NOx removal efficiency. It is reported that SCR catalysts containing transitional metal, particularly Cu and Fe, show good activity at low temperature (Li et al, 2011). However, in the case of catalytic system with bimetallic components, finding the optimal composition with both excellent activity and hydrothermal stability is too inefficient and time-consuming to be proceeded only experimentally.

In this study, we propose an efficient computational methodology for developing SCR catalyst with high NOx conversion and resistance to hydrothermal aging, using Bayesian optimization (BO).

2. Methodology

2.1. Experimental condition

In this study, we focus on Cu-Fe bimetallic catalyst for SCR. In terms of catalytic structure for SCR, various types of zeolite structures have been studied: Beta, ZSM-5, and SSZ-13. Among them, Kwak et al. (2010) have reported that SSZ-13 demonstrates superior activity in comparison with beta and ZSM-5. Therefore, we prepare Cu-Fe SCR catalyst supported by SSZ-13 (Si/Al = 12) to maximize NOx conversion at low temperature.

The catalytic performance is measured using lab-scale reactor system. The feed gas contained 500 ppm NO, 500 ppm NH_3, 10 % O_2, 5 % H_2O, and balance N_2. The total gas flow rate is 200ml/min, and the gas hourly space velocity (GHSV) is 200,000/h. Most importantly, the temperature of the feed gas is maintained at 250 ℃. The resistant ability to hydrothermal aging is also measured after exposure to 900 ℃ gas containing 10% water for 16 hours.

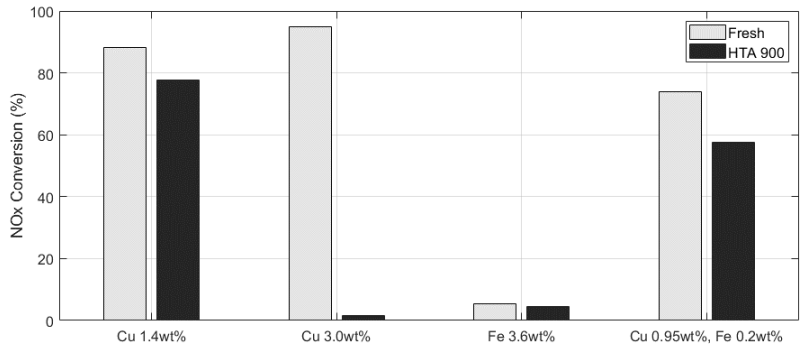


Figure 1. Experimental results for nitrogen oxides conversion of Cu/Fe SCR. Gray bar for fresh condition and black bar for hydrothermally aged (HTA) condition.

The experimental results of initial samples for Bayesian optimization are shown in Figure 1. As the copper content of SCR catalyst increases, the number of active sites on the surface of the catalyst also increases, which leads to higher NOx conversion. However, high metal content reduces catalytic stability simultaneously, making SCR catalyst susceptible to catalytic deactivation by the high temperature and water in the exhaust gas. Thus, it is required to optimize performance of the catalyst with respect to the content of active sites. The effect of two metals coexisting is also difficult to grasp intuitively. The performance of a catalyst with a specific metal content is not known until experimentally observed. A large number of trials and errors are required to develop SCR catalyst with high activity and hydrothermal resistance.

2.2. Bayesian optimization

Bayesian optimization is a useful algorithm to optimize a black-box function. It is usually employed to optimize hyperparameters of neural network. This method balances exploitation and exploration utilizing information obtained from a series of experiments (Snoek et al., 2012). Table 1 shows Bayesian optimization procedure, described by Brochu et al. (2010). $u(x)$ means the acquisition function, $f(x)$ stands for the function to be maximized, and $\mathcal{D}_{1:t}$ is the observed dataset from iteration 1 to t.

Table 1. Bayesian optimization procedure adopted from Brochu et al. (2010)

Algorithm 1 Bayesian optimization
1: for $t = 1,2,\ldots$ do
2: Find x_n that maximizes the acquisition function over the GP (Gaussian process): $x_t = argmax_x u(x\|\mathcal{D}_{1:t-1})$.
3: Sample the objective function: $y_t = f(x_t) + \varepsilon_t$.
4: Augment the data $\mathcal{D}_{1:t} = \{\mathcal{D}_{1:t-1}, (x_t, y_t)\}$ and update the GP.
5: end for

As the acquisition function, the expected improvement (EI) is chosen, which is the widely used one. The improvement I at the point x can be expressed as follows (Jones et al., 1998):

$$I(x) = \max(f_{min} - \mathrm{Y}, 0) \tag{6}$$

where f_{min} is the current best function value and Y is the random variable $\sim\mathcal{N}(\hat{y}, s^2)$ that corresponds to the function value at x. Since Y follows a Gaussian distribution, by taking the expected valued, we can obtain the expected improvement, $u_{EI}(x)$.

$$u_{EI}(x) \equiv E\big(I(x)\big) = \mathrm{E}[\max(f_{min} - \mathrm{Y}, 0)] \tag{7}$$

$$u_{EI}(x) = (f_{min} - \hat{y})\Phi\left(\frac{f_{min}-\hat{y}}{s}\right) + s\phi\left(\frac{f_{min}-\hat{y}}{s}\right) \tag{8}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density function and distribution function, respectively.

In this research, the optimization variable x is weight percent of copper and iron (Cu wt% and Fe wt%). The objective of the suggested algorithm is to maximize the catalytic activity in fresh condition and hydrothermally aged condition. For minimization problem, the objective function can be expressed as follows:

$$f(Cu, Fe) = -(\mathrm{DeNOx}_{Fresh} + \mathrm{DeNOx}_{HTA}) \tag{9}$$

where DeNOx stands for NOx conversion efficiency,

$$\mathrm{DeNOx} = \frac{\mathrm{NOx}_{inlet} - \mathrm{NOx}_{outlet}}{\mathrm{NOx}_{inlet}} * 100 \tag{10}$$

3. Results

We conduct Bayesian optimization for the optimal metal composition that shows the best NOx conversion at low temperature (250 ℃) and resistance to hydrothermal aging. Initial dataset for the surrogate model is based on the experimental data from previously published papers (Shishkin et al., 2014, Yin et al, 2016, Wang et al., 2019). As shown in Figure 2, Bayesian optimization algorithm proposes the next experimental point that maximizes the acquisition function, u_{EI}. The Cu/Fe-SSZ-13 catalyst with the suggested metal composition is evaluated for its catalytic performance though the experimental process mentioned in section 2.1. The objective function is calculated from NOx concentration of outlet gas flow. The surrogate model of BO is updated with the newly added data. These processes are repeated until the stopping criterion is met.

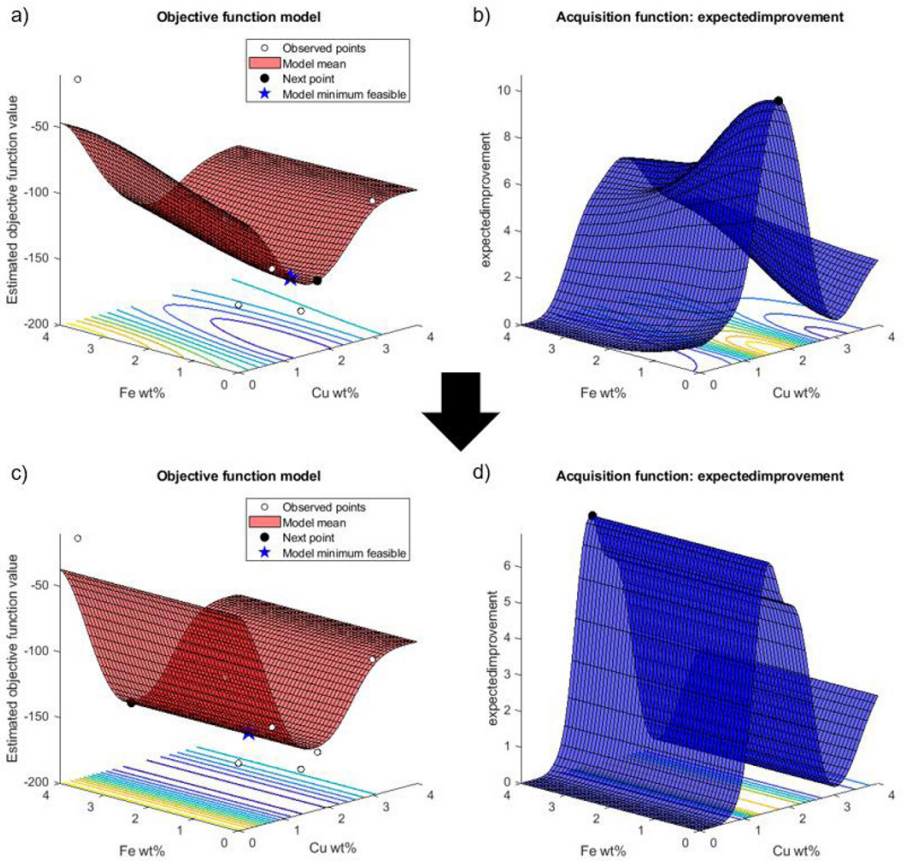


Figure 2. Bayesian optimization procedure for Cu-Fe SCR. a) objective function model of 5th step, b) acquisition function of 5th step, c) objective function of 6th step, d) Acquisition function of 6th step

After only seven steps of experiments, we found a high-performance catalyst for SCR. The optimal catalyst proposed by Bayesian optimization consists of 2.0 wt% Cu and 2.0 wt% Fe. As a result of a total of five experiments, the NOx conversions under fresh condition are 94.75 %, 94.54 %, 93.52 %, 93.71 % and 93.80 %, respectively. After hydrothermal aging of this Cu-Fe catalyst at 900 ℃ for 16 hours, the catalyst removes 87.75 %, 89.92 %, 89.75 %, 88.92 % and 87.81 % of nitrogen oxides. These results show as good catalytic performance as the SSZ-13 zeolite SCR catalyst presented in the recent paper. By using Bayesian optimization algorithm, automobile catalyst with high-performance can be found efficiently. The methodology suggested in this research can reduce time and cost of catalyst development.

4. Conclusions

In this work, we show that Bayesian optimization can efficiently support the experimental design to discover the optimal catalytic composition for the aftertreatment system of diesel vehicle. The target catalyst is Cu-Fe bimetallic catalyst for selective

catalytic reduction (SCR), supported by SSZ-13 zeolite structure. The bimetallic catalyst proposed by Bayesian optimization has high performance to reduce nitrogen oxide, showing 95.86 % conversion. The catalyst also has resistance to the hydrothermal aging. After 16 hours of exposure to 900 ℃ of air containing 10 % water, the catalyst converts 88.83 % of NOx to N2. The optimal catalyst is discovered in only seven experiments. The suggested method reduces time and cost of catalyst development. We expect that our methodology can be applied to other automobile catalyst system, such as three-way catalytic converter for gasoline vehicle.

Acknowledgements

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) [NRF-2016R1A5A1009592]

References

W. Shan and H. Song, 2015, Catalysts for the selective catalytic reduction of NO x with NH 3 at low temperature, Catalysis Science & Technology, 5(9), 4280-4288.

A. Pant and S.J. Schmieg, 2011, Kinetic Model of NOx SCR Using Urea on Commercial Cu− Zeolite Catalyst, Industrial & Engineering Chemistry Research, 50(9), 5490-5498.

L. Olsson, H. Sjövall, and R.J., Blint, 2008, A kinetic model for ammonia selective catalytic reduction over Cu-ZSM-5, Applied Catalysis B: Environmental, 81(3-4), 203-217.

J. Li, H. Chang, L. Ma, J. Hao and R. T. Yang, 2011, Low-temperature selective catalytic reduction of NOx with NH3 over metal oxide and zeolite catalysts—A review, Catalysis today, 175(1), 147-156.

J.H. Kwak, R.G. Tonkyn, D.H. Kim, J. Szanyi and C.H. Peden, 2010, Excellent activity and selectivity of Cu-SSZ-13 in the selective catalytic reduction of NOx with NH3, Journal of Catalysis, 275(2), 187-190.

J. Snoek, H. Larochelle and R.P. Adams, 2012, Practical bayesian optimization of machine learning algorithms, Advances in neural information processing systems, 25.

E. Brochu, V.M. Cora and N. De Freitas, 2010, A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning, arXiv preprint arXiv: 1012.2599.

D.R. Jones, M. Schonlau and W.J., Welch, 1998, Efficient global optimization of expensive black-box functions, Journal of Global optimization, 13(4), 455-492.

A. Shishjin, H. Kannisto, P.A., Carlsson, H. Härelind and M. Skoglundh, 2014, Synthesis and functionalization of SSZ-13 as an NH 3-SCR catalyst, Catalysis Science & Technology, 4(11), 3917-3926.

C. Yin, P. Cheng, X. Li and R.T. Yang, 2016, Selective catalytic reduction of nitric oxide with ammonia over high-activity Fe/SSZ-13 and Fe/one-pot-synthesized Cu-SSZ-13 catalysts, Catalysis Science & Technology, 6(20), 7561-7568.

A. Wang, Y. Wang, E.D. Walter, N.M. Washton, Y. Guo, G. Lu, C.H.F. Peden and F. Gao, 2019, NH3-SCR on Cu, Fe and Cu+ Fe exchanged beta and SSZ-13 catalysts: Hydrothermal aging and propylene poisoning effects, Catalysis Today, 320, 91-99.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50203-7

Capacity Planning for Sustainable Process Systems with Uncertain Endogenous Technology Learning

Tushar Rathi, Qi Zhang*

Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN 55455, USA

qizh@umn.edu

Abstract

The development and deployment of renewable technologies are key to achieving decarbonization. Optimal capacity expansion requires complex decision making that accounts for future cost reduction with increased deployment, which is also termed technology learning. Having a perfect foresight over the technology cost reduction, however, is highly unlikely. This has motivated us to develop a capacity planning model that incorporates such uncertainty. To this end, we apply a multistage stochastic programming approach with endogenous uncertainty, which results in a mixed-integer linear programming (MILP) formulation. The proposed model is applied to a case study on power capacity expansion planning, highlighting the differences in expansion decisions for low- and high-learning scenarios, which indicates the importance of stochastic optimization.

Keywords: stochastic optimization, endogenous uncertainty, technology learning

1. Introduction

Over the past few decades, the unfavorable shift in global climatic conditions has driven us to focus on renewable technology development to lower carbon emissions. The increasing energy demand has further aggravated the need for alternatives to traditional fossil energy sources. However, in addition to developing new technologies, making them economical as fast as possible remains a challenging task. In general, the cost of a technology is a function of several interrelated factors, including pricing and the number of competitors, government regulations and policies, the scale of production, and demand. The reduction in the cost of a new technology due to these factors is often termed technology learning.

Of all the stated, the scale of production constitutes a major driving force for cost reduction in new technologies. The reduction in cost as a function of installed capacity is often expressed using learning curves. Learning curves have often been used as a tool to estimate the time for a new technology to become cost-competitive. For example, Rubin et al. (2007) utilize learning curves for cost projection of power plants equipped with carbon capture and storage technology.

A less considered aspect is utilizing learning curves to make optimal capacity expansion decisions for driving down the cost of a plant or a technology in the least possible time. Most of the literature on optimization concerning learning curves assumes that they can be constructed deterministically. For example, Heuberger et al. (2017) present a power capacity expansion formulation assuming fixed learning curves for various power generation

and storage technologies. However, the lack of reliable historical data, the dependence of learning on the decisions made in real time, and the influence of other external factors make it very difficult to predict the learning curves. Therefore, decisions obtained based on deterministic learning curves may be severely sub-optimal.

To increase the practical relevance of capacity expansion models, our work incorporates uncertainty in technology learning curves. Uncertainty in learning rates has been accounted for, if at all, using methods such as sensitivity analysis and Monte Carlo simulation (Kim et al., 2012). Even though such methods provide valuable insights, their inability to account for non-anticipativity constraints demands a more rigorous optimization framework. For this reason, we explore the feasibility of stochastic programming in incorporating uncertain learning curves for multiperiod capacity expansion problems.

Uncertainty is generally classified as either exogenous or endogenous. The uncertainty not affected by decisions is termed exogenous, whereas decision-dependent uncertainty is termed endogenous. Endogenous uncertainty is further classified as type-1 and type-2. Type-1 uncertainty arises when decisions alter the probability distribution of the uncertain parameters (Peeta et al., 2010), whereas type-2 uncertainty affects the timing of the realization of the uncertain parameters (Goel and Grossmann, 2006). In a capacity expansion problem with an uncertain learning curve, the uncertainty in expansion cost resolves only when the capacity is actually increased; thus, the uncertainty here classifies as type-2 endogenous. In this work, we develop a multistage stochastic programming model for capacity planning with uncertain endogenous technology learning and apply it to a power expansion case study.

2. Stochastic programming model

To capture the interconnectivity of technologies, model their simultaneous availability to satisfy product demand, and optimize their selection for capacity expansion and operations, we consider a general process network comprising process and resource nodes as illustrated in Figure 1. Processes and resources are denoted by square and circular nodes, respectively. The arcs in the network denote the directed flow of resources. Process nodes can represent chemical and manufacturing processes or, generally, technologies. Resource $j \in \mathcal{J}$ from a process $k \in \mathcal{K}$ can either serve as an input resource to process $k' \in \mathcal{K}\backslash\{k\}$, be discharged from the process network, or be purchased from outside the network.

The goal is to determine optimal capacity expansion decisions during the planning horizon $\mathcal{T}$, and devise optimal operational decisions in each scheduling horizon $\mathcal{H}_t$ based on each process's installed capacity, demand of resources, and all the involved costs. Uncertainty in technology learning curves is accommodated by considering different possible scenarios (combination of learning curves for multiple uncertain technologies).

2.1. Capacity expansion constraints

Based on the process network in Figure 1, we define binary variable x_{kits} that equals 1 if process k undergoes capacity expansion to (at least) the permissible point $i \in \mathcal{I}_k$ in time period $t \in \mathcal{T}$ of scenario $s \in \mathcal{S}$. We further define the variables C_{kts} and Δ_{kts} such that they represent the cumulative installed capacity and additional capacity installed of

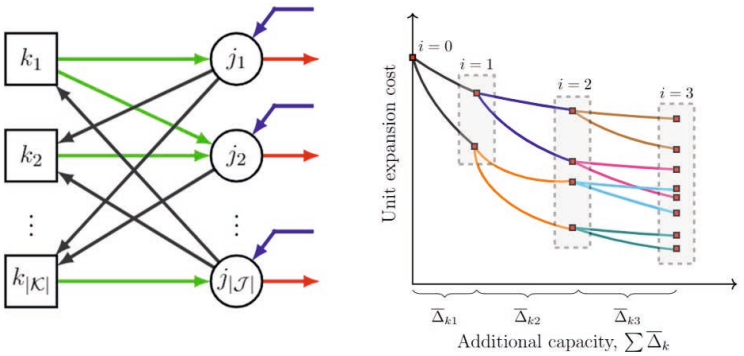


Figure 1: A general process network and an illustrative uncertain learning curve. Each discrete expansion point i acts as a source of uncertainty. In this case, we have two, four, and eight possible unit expansion costs at $i = 1, 2,$ and, 3 respectively.

a process k in time period t of scenario s, respectively. Then, the following constraints control the timing and extent of capacity expansion for each technology:

$$C_{k0s} = \overline{C}_{k0} \qquad \forall k \in \mathcal{K}, \forall s \in \mathcal{S} \quad \text{(1a)}$$

$$C_{kts} = C_{k,t-1,s} + \Delta_{kts} \qquad \forall k \in \mathcal{K}, t \in \mathcal{T}, \forall s \in \mathcal{S} \quad \text{(1b)}$$

$$\Delta_{kts} = \sum_{i \in \mathcal{I}_k} x_{kit} \overline{\Delta}_{ki} \qquad \forall k \in \mathcal{K}, t \in \mathcal{T}, \forall s \in \mathcal{S} \quad \text{(1c)}$$

$$\Delta_{kts} \leq b_{kt} \qquad \forall k \in \mathcal{K}, t \in \mathcal{T}, \forall s \in \mathcal{S} \quad \text{(1d)}$$

$$x_{kits} \leq \sum_{\tau=1}^{t} x_{k,i-1,\tau s} \qquad \forall k \in \mathcal{K}, i \in \mathcal{I}_k \backslash \{1\}, t \in \mathcal{T}, \forall s \in \mathcal{S} \quad \text{(1e)}$$

$$\sum_{\tau=1}^{t} x_{ki\tau s} \leq 1 \qquad \forall k \in \mathcal{K}, i \in \mathcal{I}_k, t \in \mathcal{T}, \forall s \in \mathcal{S} \quad \text{(1f)}$$

$$g(Q_{hts}, C_{kts}) \leq 0 \qquad \forall k \in \mathcal{K}, h \in \mathcal{H}_t, t \in \mathcal{T}, \forall s \in \mathcal{S} \quad \text{(1g)}$$

$$x_{kits} \in \{0, 1\} \qquad \forall k \in \mathcal{K}, i \in \mathcal{I}_k, t \in \mathcal{T}, \forall s \in \mathcal{S} \quad \text{(1h)}$$

$$C_{kts}, \Delta_{kts} \geq 0 \qquad \forall k \in \mathcal{K}, t \in \mathcal{T}, \forall s \in \mathcal{S} \quad \text{(1i)}$$

$$Q_{hts} \in \mathbb{R}^{|\mathcal{J}||\mathcal{K}|} \times \mathbb{Z}^{|\mathcal{K}|} \qquad \forall t \in \mathcal{T}, h \in \mathcal{H}_t, \forall s \in \mathcal{S} \quad \text{(1j)}$$

where $\overline{C}_{k0}$ denotes the initial installed capacity of process k. The incremental capacity for process k from point $i-1$ to i is denoted by $\overline{\Delta}_{ki}$. Constraints (1a)-(1c) together represent the capacity balance. Constraints (1d) limit the capacity expansion of a process k by the available budget b_{kt} in time period t. Constraints (1e) ensure that we move in the positive direction on the learning curve in a sequential fashion, i.e., we can only install additional capacity corresponding to point i if we have already installed the additional capacity corresponding to point $i-1$. Constraints (1f) imply that investment at any point $i \in \mathcal{I}_k$ cannot be made more than once in any time period. Constraints (1g) are a condensed representation of all the operational constraints, including production scheduling, inventory management, scheduling startup/shutdown of units, limiting emissions and storage,

to name a few. Operational decision variables Q_{hts} can be both continuous and discrete and are constrained by the installed capacities of the processes in the network.

2.2. *Non-anticipativity constraints*

Non-anticipativity constraints (NACs) ensure the equality of decisions for all pairs of indistinguishable scenarios at any point in time during the planning horizon. Mathematically, NACs are represented as follows:

$$x_{ki1s} = x_{ki1,s+1} \quad \forall k \in \mathcal{K}, i \in \mathcal{I}_k, s \in \mathcal{S}, s < |\mathcal{S}| \tag{2a}$$

$$\begin{bmatrix} Z_t^{s,s'} \\ x_{ki,t+1,s} = x_{ki,t+1,s'} \quad \forall k \in \mathcal{K}, i \in \mathcal{I}_k \end{bmatrix} \vee \left[\neg Z_t^{s,s'}\right] \quad \forall (s,s') \in \mathcal{P}', t \in \mathcal{T}\backslash\{T\} \tag{2b}$$

$$Z_t^{s,s'} \iff \bigwedge_{(r,i)\in\mathcal{D}(s,s')} \left[\bigwedge_{\tau=1}^{t} (\neg x_{ri\tau s})\right] \quad \forall (s,s') \in \mathcal{P}', t \in \mathcal{T}\backslash\{T\} \tag{2c}$$

$$Z_t^{s,s'} \in \{\texttt{true}, \texttt{false}\} \quad \forall (s,s') \in \mathcal{P}', t \in \mathcal{T}\backslash\{T\} \tag{2d}$$

where $\mathcal{D}(s,s')$ is the set containing sources of endogenous uncertainty (expansion points in our case) that distinguish scenario s from s'. The Boolean variable $Z^t_{s,s'}$ is true if uncertainty has not been realized in any of the uncertain parameters that belong to the set $\mathcal{D}(s,s')$. Further, $\mathcal{P}'$ denotes the minimum or reduced set of scenario pairs that is sufficient to express all the NACs. The details on the disjunction and logic-based formulation of NACs for endogenous uncertainty problems can be found in Goel and Grossmann (2006). Also, we refer the reader to Hooshmand and MirHassani (2016) for redundant NAC removal strategies in case of endogenous uncertainty and an arbitrary scenario set.

2.3. *Objective function*

The objective is to minimize the expected net cost over the entire planning horizon; thus, the overall stochastic optimization problem can be summarized as follows:

$$\min \quad \sum_{s\in\mathcal{S}} p_s \sum_{t\in\mathcal{T}} \alpha_t \left[\sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{I}_k} \left(\int_{\Phi_{k,i-1}}^{\Phi_{ki}} f_{ks}(\Phi_k) \mathrm{d}\Phi_k \right) x_{kits} + \sum_{h\in\mathcal{H}_t} \sum_{k\in\mathcal{K}} \sum_{j\in\mathcal{J}} u_{jkhts}(Q_{hts}, C_{kts}) \right]$$

s.t. Eqs. (1a) - (1j), (2a) - (2d)

where p_s denotes the probability of scenario s and α_t denotes the discount factor for time period t. The learning curve for process k is encoded in the model as $f_k(\Phi_k)$ and $\Phi_{ki} := \sum_{i'=1}^{i} \overline{\Delta}_{ki}$. Note that we make no assumptions on the form of the learning curve since the integral term (expansion cost on increasing capacity from point $i-1$ to i) is a parameter that can be pre-calculated. The cost function u captures all operating costs including the cost of specific modes of operation, utilizing storage, purchasing and discharging resources, tax on emissions, etc.

3. Industrial case study

The proposed framework is applied to a capacity expansion case study for a network of power generation technologies. Specifically, we consider seven technologies and categorize them into one of the following three categories – conventional (no cost reduction), deterministic (known learning curve), and uncertain technology (uncertain learning curve). Nuclear, coal, combined cycle gas turbine (CCGT), and open cycle gas turbine (OCGT) are considered conventional, onshore wind and solar are assumed to be deterministic, and offshore wind is assumed to have an uncertain learning curve. The model and data for this case study are partially adapted from Heuberger et al. (2017). The planning problem was modeled using JuMP v0.21.10 in Julia v1.6.3 and was solved using Gurobi v9.1.2. The model was solved to optimality (0.01% tolerance) in 3,150 s.

The planning horizon spans eight 5-year time periods from 2015 to 2055. The capacity expansion decisions are made at the start of each of these time periods. Figure 2 illustrates the eight possible learning curves for offshore wind technology and the eventual scenario tree based on the expansion decisions made. The scenario tree indicates that the offshore wind capacity increases by 2.5 GW at $t = 1$; however, as expected, we do not see any further expansion for the low-learning case (high-cost scenarios). On the contrary, for the high-learning case (low-cost scenarios), at $t = 2$, the capacity further expands by 5.8 GW, resulting in four scenario tree nodes. Thus, stochastic programming adapts its decisions to the future expansion cost, generating practically viable solutions in the process.

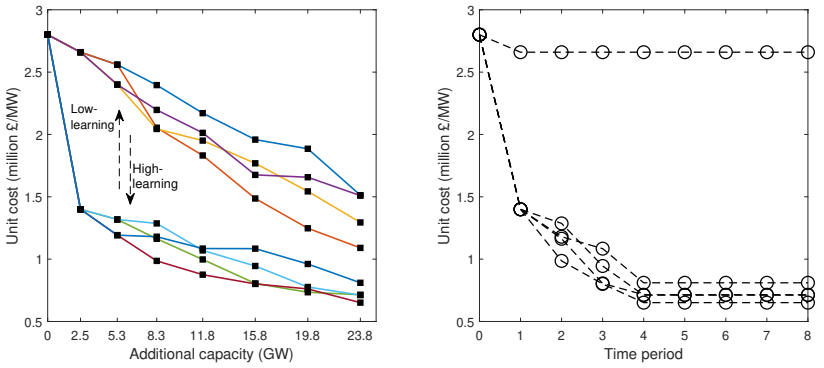


Figure 2: The possible learning curves considered for offshore wind are illustrated on the left. Scenario tree (right) reveals expansion decisions for offshore wind.

Next, Figure 3 illustrates the distribution of capacity for all technologies throughout the planning horizon. Clearly, in comparison to the high-learning scenario, the low-learning scenario does not favor offshore wind expansion. This reduced capacity expansion in offshore wind is compensated by expansions of conventional technologies such as nuclear and OCGT. Note that the expansions are governed not only by the expansion cost but also by the expansion budget, lifetime of each technology, and the time-varying power generation capacity. The proposed stochastic programming model effectively integrates the above factors with the uncertain cost to generate the optimal capacity distribution.

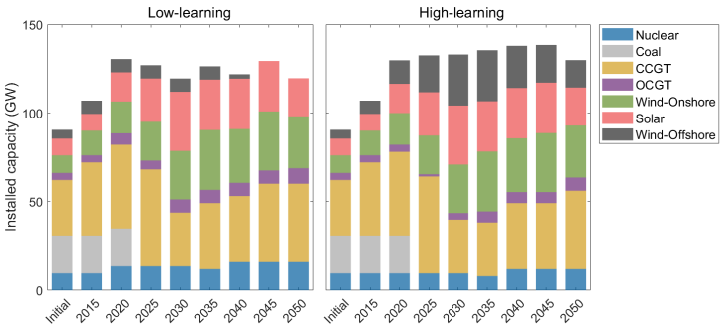


Figure 3: Capacity distribution of power generation technologies under low- and high-learning scenarios.

4. Conclusions

In this work, we proposed a rigorous optimization framework for a general process network that can be utilized to model energy systems containing both renewable and non-renewable technologies. We utilize stochastic programming to account for the long-neglected aspect of uncertainty in technology learning curves. The case study on power capacity expansion showcases the adaptability of stochastic programming in providing decisions optimal to individual scenarios. The difference in decisions also indicates that any solution obtained through a deterministic model, which essentially is a single scenario case, would often be sub-optimal for any perturbation in the assumed deterministic learning curves.

References

V. Goel and I. E. Grossmann, 2006. A class of stochastic programs with decision dependent uncertainty. *Mathematical programming*, 108(2):355–394.

C. F. Heuberger, E. S. Rubin, I. Staffell, N. Shah, and N. Mac Dowell, 2017. Power capacity expansion planning considering endogenous technology cost learning. *Applied Energy*, 204:831–845.

F. Hooshmand and S. MirHassani, 2016. Efficient constraint reduction in multistage stochastic programming problems with endogenous uncertainty. *Optimization Methods and Software*, 31(2):359–376.

S. Kim, J. Koo, C. J. Lee, and E. S. Yoon, 2012. Optimization of Korean energy planning for sustainability considering uncertainties in learning rates and external factors. *Energy*, 44(1):126–134.

S. Peeta, F. S. Salman, D. Gunnec, and K. Viswanath, 2010. Pre-disaster investment decisions for strengthening a highway network. *Computers & Operations Research*, 37 (10):1708–1719.

E. S. Rubin, S. Yeh, M. Antes, M. Berkenpas, and J. Davison, 2007. Use of experience curves to estimate the future cost of power plants with CO_2 capture. *International Journal of Greenhouse Gas Control*, 1(2):188–197.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50204-9

Development of a bi-objective optimisation framework for mixed-integer nonlinear programming problems and application to molecular design

Ye Seol Lee[a], George Jackson[a], Amparo Galindo[a], Claire S. Adjiman[a]*

[a]*Department of Chemical Engineering, Sargent Centre for Process Systems Engineering, Institute for Molecular Science and Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom*

c.adjiman@imperial.ac.uk

Abstract

We present a novel algorithm (SDNBI) to tackle the numerical challenges associated with the solution of bi-objective mixed-integer nonlinear programming problems (BO-MINLPs), with a focus on the exploration of nonconvex regions of the Pareto front. The performance of the algorithm as measured by the accuracy of the resulting approximation of the Pareto front in the disconnected and nonconvex domain of Pareto points is assessed relative to two multi-objective optimisation (MOO) approaches: the sandwich algorithm (SD) and the modified normal boundary intersection (mNBI) method. The features of these MOO algorithms are evaluated using two published benchmark models and a molecular design problem. Initial results indicate that the new algorithm presented outperforms both the SD and the mNBI method in convex, nonconvex-continuous, combinatorial problems, both in terms of computational cost and of the overall quality of the Pareto-optimal set.

Keywords: Multi-objective optimisation, Mixed-integer nonlinear programming, Computer-aided molecular design

1. Introduction

Multi-objective optimisation (MOO) techniques have been applied to problems across a wide range of engineering fields to identify trade-offs between conflicting decision criteria that cannot easily be placed on the same quantitative footing. Some of the most widely used approaches to solving MOO problems are based on scalarisation methods and include the weighted sum method (Marler and Arora, 2004), the normal boundary intersection (NBI) method (Das and Dennis, 1998) and the sandwich (SD) algorithm (Rennen et al., 2011). However, these methods suffer from limitations that prevent them from being used to reliably producing optimal solutions along the nonconvex or discrete regions of a Pareto front. As a result, the performance of these methods during the solution of many practical problems is limited when nonconvexities are arise due to the presence of discrete decision variables and/or nonlinear model equations.

In this work, we present a robust bi-objective optimisation approach, SDNBI algorithm that combines features of the SD and NBI methods in order to overcome difficulties in converging to the true Pareto front and maintaining a well-distributed set of solutions. The main improvements introduced are the identification of regions where no further optimal solution exists, i.e., disconnected parts of the Pareto front, and the exploration of nonconvex parts of the Pareto front. The proposed approach is evaluated using two published benchmark problems with different levels of complexity in terms of problem size and numerical difficulty. The performance of the algorithm is compared with that of the SD method and of the mNBI algorithm, a modified version of the NBI method (Shukla, 2007). The comparison is based on the accuracy of the approximation of Pareto fronts generated.

The efficiency of the proposed algorithm is further investigated through a computer-aided molecular and process design (CAMPD) problem. This provides an opportunity to examine the applicability and reliability of the proposed algorithm in a mixed-integer nonlinear setting.

2. Background and motivation

The generic mathematical formulation of the MOO problem is defined as:

$$\begin{array}{ll} \min & (f_1(\boldsymbol{x}), \ldots, f_m(\boldsymbol{x}))^\top \\ \text{subject to} & \boldsymbol{x} \in X := \{\boldsymbol{x} \in \mathbb{R}^{n_1} \times \mathbb{N}^{n_2} \mid \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}, \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}, \end{array} \tag{MOP}$$

where objective functions $f_j : \mathbb{R}^n \to \mathbb{R},\ j = 1, 2, ..., m$, $\boldsymbol{x}$ is an n-dimensional vector consisting of n_1 continuous variables and n_2 discrete variables ($K = n_1 + n_2$), $X \neq \emptyset$ is an nonempty feasible set, $\boldsymbol{g}(\boldsymbol{x})$ is a p-dimensional vector of inequality constraints and $\boldsymbol{h}(\boldsymbol{x})$ is a q-dimensional vector of equality constraints, $q \leq K$. In this paper, we are specifically interested in bi-objective problems, i.e., $m = 2$.

2.1. The Sandwich Algorithm

The sandwich (SD) algorithm proposed by Rennen et al. (2011) provides an efficient way to approximate a convex Pareto front based on the successive solution of weighted-sum sub-problems in which the multiple objective functions are scalarised into a single objective function through a weight vector, $\boldsymbol{w}$, as follows:

$$\min_{\boldsymbol{x} \in X} \quad \boldsymbol{w}^\top \boldsymbol{f}(\boldsymbol{x}) = \sum_{i=1}^{m} w_i f_i(\boldsymbol{x}). \tag{WSPw}$$

Within the algorithm, the Pareto front is approximated by recursively improving inner and outer approximations, with the Pareto surface "sandwiched" between them. The inner approximation, which provides an upper bound of the Pareto front in the objective space, is constructed by generating a convex hull from the current set of the Pareto points, while the outer approximation, which supports the Pareto front, is improved at each iteration by adding a hyperplane tangential to any new Pareto point $\boldsymbol{z}$ identified. At each iteration, the parameters of the next sub-problem to be solved, i.e., the weight coefficients $\boldsymbol{w}$ are chosen based on the normal vector of facets derived from the inner approximation.

2.2. The Normal Boundary Intersection Method

The NBI method was proposed by Das and Dennis (1998) to generate uniformly-spread nondominated solutions of a general nonlinear MOO problem. In the NBI method, the individual minima $\boldsymbol{Z^{Ai}}, i = 1, \ldots, m$, of the objective functions are found as a first step. The convex hull of individual minima (CHIM) is then generated as a set of all convex combinations of those extreme points. The CHIM can be expressed as $\boldsymbol{f^{id}} + \{\boldsymbol{\Phi\beta} : \boldsymbol{\beta} \in \mathbb{R}^m_+, \sum_{j=1}^m \beta_j = 1\}$, where $\boldsymbol{f^{id}}$ is a so-called ideal point, $\boldsymbol{\Phi} \in \mathbb{R}^{m \times m}$ is a matrix with i^{th} column $\boldsymbol{z^{Ai}} - \boldsymbol{f^{id}}$. Next, the NBI sub-problem is solved for the set of reference points $\boldsymbol{\Phi\beta}$ uniformly distributed over the CHIM, searching for the maximum distance t^* along the normal vector $\bar{\mathbf{n}}$ to the CHIM at each reference point. Here, we make use of a modified NBI subproblem suggested by Shukla (2007) to prevent convergence to a dominated boundary point of the set of feasible objective functions. The formulation of the mNBI sub-problem is as follows:

$$\begin{array}{ll} \max\limits_{\boldsymbol{x} \in X, t} & t \\ \text{s.t.} & \boldsymbol{\Phi\beta} + t\bar{\boldsymbol{n}} \geq \boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f^{id}} \\ & \boldsymbol{t} \in \mathbb{R},\ \bar{\boldsymbol{n}} \in \mathbb{R}^m \end{array} \tag{mNBIβ}$$

where $\bar{\boldsymbol{n}}$ is the outer normal direction at some point on the CHIM pointing toward $\boldsymbol{f^{id}}$. Note we assume the problem is such that a constraint qualification holds.

3. Proposed Algorithm

We present a novel adaptive MOO approach, the SDNBI algorithm, for the solution of bi-objective problems. It is an improvement on the sandwich algorithm in that it is applicable to nonconvex problems, in particular to those with a disconnected Pareto front and feasible region. A new feature is the use of the mNBI sub-problem such that the mNBI parameters $(\boldsymbol{\beta}, \bar{\boldsymbol{n}})$ are systematically determined by the SD algorithm, in an adaptive fashion. The main aspects of the proposed methodology are explained in this section.

3.1. Inner and outer approximation

In the original SD algorithm, the outer approximation is improved at each iteration k by solving problem (WSPw) whereby a hyperplane $H(\boldsymbol{w^k}, \boldsymbol{b^k})$ tangential to the Pareto front at $\boldsymbol{z^k}$ (where $b^k = \boldsymbol{z^k w^k}$) is added. The inner approximation is obtained by polyhedral approximation of a current set of Pareto points $\boldsymbol{Z_E}$. The polyhedral approximation is defined as the set of all convex combinations of points in $\boldsymbol{Z_E}$ in the space of feasible objective function values (the "objective space") and is denoted as convhull($\boldsymbol{Z_E}$). This approach requires a convexity assumption on the Pareto front. In the nonconvex case, the hyperplane $H(\bar{\boldsymbol{n}}^{\boldsymbol{k}}, b^k)$ generated as a solution of problem (mNBIβ) may not be tangential to $\boldsymbol{z^k}$, leading to an erroneous lower bound. To overcome this, the generation of the inner and outer approximations is modified such that: 1) the tangential hyperplane with a normal vector $\boldsymbol{w}$ is identified as a solution of problem (mNBIβ) using a relationship arising from the Karush-Kuhn-Tucker (KKT) optimality conditions of sub-problem; and 2) a systematic decomposition of the objective space is introduced, such that the supporting hyperplanes can be used to approximate the Pareto front within each subregion.

Given an efficient solution $\boldsymbol{x^*}$, obtained by solving an mNBI subproblem and at which $t = t^*$, there exist $\boldsymbol{\mu}^*$ and $\boldsymbol{\nu}^*$ such that the KKT optimality conditions for optimisation problem (mNBIβ) are satisfied and therefore:

$$\nabla_{\boldsymbol{x}} \mathcal{L} = \boldsymbol{\mu^*}^\top \nabla_{\boldsymbol{x}} \boldsymbol{f}(\boldsymbol{x^*}) + \boldsymbol{\nu^*}^\top \nabla_{\boldsymbol{x}} \hat{\boldsymbol{h}}(\boldsymbol{x^*}) = \boldsymbol{0} \quad (1)$$

$$\nabla_t \mathcal{L} = -1 + \boldsymbol{\mu^*}^\top \bar{\boldsymbol{n}} = 0 \quad (2)$$

where $\mathcal{L}(\boldsymbol{x}, t, \boldsymbol{\mu}, \boldsymbol{\nu}) = -t + \boldsymbol{\mu}^\top \left(\boldsymbol{f(x)} - \boldsymbol{f^{id}} - \boldsymbol{\Phi\beta} - t\bar{\boldsymbol{n}}\right) + \boldsymbol{\nu}^\top \hat{\boldsymbol{h}}(\boldsymbol{x})$, $\boldsymbol{\mu} \in \mathbb{R}^m$ represents the vector of the Lagrange multipliers corresponding to the augmented objective constraints $\boldsymbol{f(x)} - \boldsymbol{f^{id}} - \boldsymbol{\Phi\beta} - t\bar{\boldsymbol{n}} \leq \boldsymbol{0}$, and $\boldsymbol{\nu} \in \mathbb{R}^s$ is the vector of Lagrange multipliers for the s active constraints in the set $\{\boldsymbol{g(x)} \leq \boldsymbol{0}, \boldsymbol{h(x)} = \boldsymbol{0}\}$, represented by the vector $\hat{\boldsymbol{h}}(\boldsymbol{x}) \in \mathbb{R}^s$, $q \leq s \leq p + q$.

From equation (2), it can be deduced that at least one of the augmented objective constraints must be active. This can be shown by contradiction. Let us assume that all constraints of the augmented objective constraints are inactive, i.e., $\boldsymbol{f(x)} - \boldsymbol{f^{id}} - \boldsymbol{\Phi\beta} - t\bar{\boldsymbol{n}} < \boldsymbol{0}$ and $\boldsymbol{\mu^*} = \boldsymbol{0}$. Then, equation (2) is reduced to $\nabla_t \mathcal{L} = -1 \neq 0$, which is the violation of the KKT necessary conditions. Therefore, if we solve problem (mNBIβ) for any choice of parameters $(\bar{\boldsymbol{n}}, \boldsymbol{\beta})$, there exists a corresponding normal vector $\boldsymbol{w} \geq \boldsymbol{0}$, where the inequality is understood component-wise, that defines a tangent to the Pareto front at the nondominated solution, $\boldsymbol{z}$, and is given by:

$$\boldsymbol{w} = \frac{1}{\sum_{i=1}^m \mu_i^*} \boldsymbol{\mu^*}, \ \sum_{i=1}^m w_i = 1. \quad (3)$$

3.2. Decomposition of the search space

The principle behind the decomposition strategy is to divide the objective space into subregions whenever there exist Pareto points that are non-interior points of the intersection of

supporting hyperplanes and their half-spaces. At the initial step of the SDNBI algorithm, it is assumed that the Pareto front is convex in the objective search space C^l. The search space is then investigated whenever a new Pareto point is obtained to determine whether the assumption holds or the current space needs to be decomposed, so that valid approximations of the Pareto front can be provided, i.e., to sandwich the Pareto front. This can be done based on the rule derived from the supporting hyperplane theorem (Luenberger, 1997):

Suppose convex set C is a polyhedron defined by a finite number of half-spaces and hyperplanes $C = \{\boldsymbol{z} \mid \boldsymbol{w^k}^\top \boldsymbol{z} \geq \boldsymbol{w^k}^\top \boldsymbol{z^k},\ k = 1, \ldots, K\}$. If each point $\boldsymbol{z^k}$ is supported by the hyperplane $H(\boldsymbol{w^k}, b^k)$ where $\boldsymbol{b^k} = \boldsymbol{w^k z^k}$, then every point $\boldsymbol{z} \in C$ must satisfy $(\boldsymbol{w^k})^\top \boldsymbol{z} \geq \boldsymbol{b^k}$ for all k.

If there is a Pareto point $\boldsymbol{z} \in C^l$ that violate the condition above, then the current search space C^l is decomposed into two sub-spaces C^l and C^{l+1}. The convexity or concavity of each decomposed space is postulated based on the sign of the inequalities – for example, if all Pareto points in C^{l+1} satisfy $(\boldsymbol{w^k})^\top \boldsymbol{z} \geq \boldsymbol{w^k z^k}$, then the Pareto front in the region C^{l+1} is assumed to be convex, whereas the Pareto front in C^{l+1} is assumed to be concave if $(\boldsymbol{w^k})^\top \boldsymbol{z} \leq \boldsymbol{w^k z^k}$.

3.3. *Modification of the mNBI sub-problem*

Although the use of the mNBI method makes it possible to avoid the unnecessary investigation of some boundary points that are dominated by others, the mNBI method may not be efficient when a Pareto front consists of many disconnected regions. This is because it becomes difficult to identify the next search direction where as-yet unknown nondominated points exist in the absence of a clear criterion. To address this, the following sub-problem (mNBI$\bar{n}$), which is a modification of the original mNBI sub-problem, is introduced to fathom regions where no Pareto optimal solutions exist:

$$\begin{aligned} &\max_{\boldsymbol{x} \in \boldsymbol{X}, t} \ t \\ &\text{s.t. } \boldsymbol{\Phi\beta} + t\bar{\boldsymbol{n}} \geq \boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f^{id}} \\ &f_1(\boldsymbol{x}) \geq z_1^{k1} + \epsilon_z \text{ or } \ f_1(\boldsymbol{x}) \leq z_1^{k1} - \epsilon_z \end{aligned} \tag{mNBI$\bar{n}$}$$

where $\boldsymbol{z^{k1}}, \boldsymbol{z^{k2}} \in \boldsymbol{Z_E}$ are nondominated points obtained at previous iterations and the inequality $z_1^{k1} < z_1^{k2}$ holds. Given a Pareto front, whenever the solution of sub-problem (mNBIβ) for a current facet generates a previously identified solution $\boldsymbol{z^{k1}}$ or $\boldsymbol{z^{k2}}$, the facet is further explored by solving (mNBI$\bar{n}$), thereby excluding the empty part of the subspace from the search space. Note that only one constraint between $f_1(\boldsymbol{x}) \geq z_1^{k1} + \epsilon_z$ and $f_1(\boldsymbol{x}) \leq z_1^{k1} - \epsilon_z$ is imposed in (mNBI$\bar{n}$) and the choice of the constraints depends on the solution of the problem (mNBIβ) i.e., whether it converges to $\boldsymbol{z^{k1}}$ or $\boldsymbol{z^{k2}}$, respectively. If the solution of (mNBI$\bar{n}$) converges to a point that was obtained at a previous iteration, the entire facet is discarded from the search space in subsequent iterations.

4. Performance of the SDNBI algorithm

To assess the performance of our proposed algorithm, we apply it to two well-known benchmark problems, SCH2 and ZDT5 (Natarajan, 2003), and compare solution statistics with those of the SD algorithm and mNBI method. In addition, the applicability of our SDNBI to a CAMPD problem is examined for the simultaneous design of optimal working fluids and organic rankine cycle processes (Lee et al., 2020; Bowskill et al., 2020) (CS1). For a detailed description of the formulation of the test problems, the reader is referred to Natarajan (2003) and to Bowskill et al. (2020) and case study 2 (Table 1) therein.

Five criteria are chosen as quality measures for the comparison of the Pareto points produced by each algorithm: (1) the number of unique non-dominated solutions (N_{unq}); (2)

the distribution metric (DM), which captures the extent of the spread over the Pareto front approximation and the extent of the front Pareto front covered by the nondominated points; (3) the hypervolume (HV), which is the volume of the m-dimensional region in the objective space enclosed by the obtained non-dominated solutions and a reference point; (4) the average CPU time to generate a nondominated point ($t_{\text{CPU,a}}$); and (5) the total CPU time to generate all solutions ($t_{\text{CPU,t}}$).

The MOO algorithms used here are implemented in Matlab 2018a, using common subfunctions. For problems SCH2 and ZDT5, each scalarised sub-problem is solved through GAMS interfaced with the DICOPT solver, while the CAMPD problem is implemented in C++ interfaced with gPROMS ModelBuilder 7.0.7. To increase the likelihood of identifying globally optimal Pareto points, a multi-start approach is adopted for the solution of the sub-problems. The multiple starting points are generated by a Sobol' sequence (Sobol' et al., 2011) to achieve well-distributed coverage in the space of design variables $\boldsymbol{x}$. For the SD and the mNBI methods, the stopping criterion is defined as the fixed number of iterations N_{iter} chosen as the number of iterations taken for SDNBI to reach a pre-defined error tolerance for each test problem.

5. Results and Discussions

The performance of the SD, mNBI, and SDNBI methods for all test problems is summarised in Table 1 and Figure 1. For SCH2, the Pareto-optimal set consists of one point and one continuous region, in which a large nonconvex and a disconnected area are present. As is apparent from Table 1, the highest HV value and the lowest DM value are achieved when using SDNBI. The performance of mNBI is similar to that of SDNBI in terms of HV and DM, but fewer points are identified within the fixed number of iterations. This is mainly because some iterations using mNBI fail to identify new points in the region where Pareto points do not exist. The strong performance of SDNBI is further highlighted in test problem ZDT5, where the true Pareto front is a set of 31 integer Pareto points. The use of SDNBI guarantees the identification of the complete set of solutions for a given fixed number of iterations, while the mNBI and SD methods appear to be relatively ineffective in achieving high accuracy in HV and N_{unq}. After iterating 40 sub-problems, all 31 Pareto points are generated by SDNBI, confirming its reliability in finding a diverse and reliable Pareto approximation, i.e., the highest HV, N_{unq}, and the lowest DM. It is noticeable that all regions where Pareto points do not exist are removed from the search space after 61 iterations, which allows the search procedure within SDNBI to terminate the algorithm, even though the stopping criteria are not satisfied.

The similar trends of the performance for all algorithms are evident for the CS1 where three nonconvex and two disconnected parts are potentially involved. By analysing the results shown in Table 1 and Figure 1, we can conclude that the SDNBI outperforms the SD and mNBI methods in all comparison criteria, with the exception of the CPU time. The SDNBI method is capable of recognising the region where no additional Pareto points are placed (see the two disconnect regions near $0.8 \leq f_2 \leq 0.85$ and $0.58 \leq f_2 \leq 0.62$

Table 1: Performance metrics for problems SCH2, ZDT5 and CS1 using SD, mNBI and SDNBI. Each test problem is evaluated for a fixed number of iterations N_{iter}=27 for SCH2, 40 for ZDT5 and 61 for CS1, respectively.

	SCH2			ZDT5			CS1		
	SD	mNBI	SDNBI	SD	mNBI	SDNBI	SD	mNBI	SDNBI
N_{unq}	26	25	26	23	23	31	59	58	60
HV / 10^{-2}	53.09	62.70	62.94	89.46	89.48	89.57	91.23	91.47	91.44
DM	0.2512	0.0998	0.0910	0.1377	0.1230	0.0944	0.1130	0.0472	0.0387
$t_{\text{CPU,a}}$ / s	9.35	9.46	9.41	73.40	52.86	51.92	12.41	20.26	29.07
$t_{\text{CPU,t}}$ / s	2.43×10^{2}	2.35×10^{2}	2.44×10^{2}	1.69×10^{3}	1.22×10^{3}	1.61×10^{3}	7.21×10^{2}	8.21×10^{2}	9.04×10^{2}

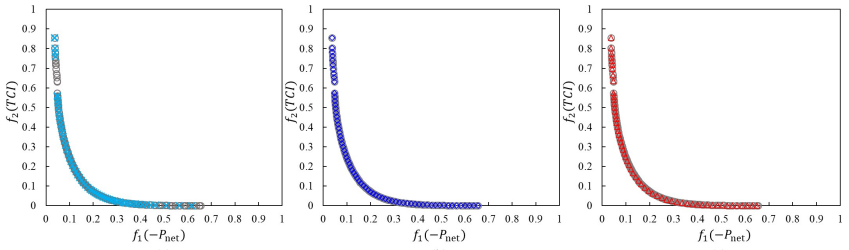


Figure 1: Pareto points in a bi-objective space generated by the (a) SD (×), mNBI (◇), and SDNBI (△) methods for CS1. Grey markers (○) indicate best-known Pareto points.

in Figure 1), making it possible to improve the Pareto approximation efficiently. The higher CPU time with the SDNBI and mNBI can be attributed to the fact that additional constraints are used in their sub-problems and so it is difficult to find feasible regions that satisfy all the constraints.

6. Conclusions

We have proposed a novel algorithm applicable to nonconvex and discrete bi-objective problems. The algorithm is based on the combination of features from the SD algorithm and the mNBI method. It has been tested on two benchmark functions and one CAMPD application to assess its performance in approximating the Pareto front. The comparative results have highlighted the robustness of the SDNBI algorithm in terms of generating a more diverse and better-distributed set of Pareto points. Future work will involve testing the proposed approach on more case studies to derive general conclusions on its performance on MINLP problems, and the extension of the algorithm to solve optimisation problems with more than two objectives.

Acknowledgements

YSL gratefully acknowledges financial support from the British Federation of Women Graduates, and the Department of Chemical Engineering at Imperial College London for a Roger Sargent scholarship.

References

Bowskill, D.H., Tropp, U.E., Gopinath, S., Jackson, G., Galindo, A., Adjiman, C.S., 2020. Beyond a heuristic analysis: integration of process and working-fluid design for organic rankine cycles. Mol. Syst. Des. Eng. 5, 493–510. doi:10.1039/C9ME00089E.

Das, I., Dennis, J.E., 1998. Normal-boundary intersection: A new method for generating the pareto surface in nonlinear multicriteria optimization problems. SIAM J. on Optim. 8, 631–657. doi:10.1137/S1052623496307510.

Lee, Y.S., Graham, E.J., Galindo, A., Jackson, G., Adjiman, C.S., 2020. A comparative study of multi-objective optimization methodologies for molecular and process design. Comput. Chem. Eng. 136, 106802. doi:10.1016/j.compchemeng.2020.106802.

Luenberger, D.G., 1997. Optimization by vector space methods. John Wiley & Sons.

Marler, R.T., Arora, J.S., 2004. Survey of multi-objective optimization methods for engineering. Struct. Multidiscip. O 26, 369–395. doi:10.1007/s00158-003-0368-6.

Natarajan, S.K., 2003. Test Suite for Multiobjective Optimization and Results Using Normal Boundary Intersection (NBI) in Design Explorer. Utah State University.

Rennen, G., Van Dam, E.R., Den Hertog, D., 2011. Enhancement of sandwich algorithms for approximating higher-dimensional convex pareto sets. INFORMS J. Comp. 23, 493–517. doi:10.1287/ijoc.1100.0419.

Shukla, P.K., 2007. On the normal boundary intersection method for generation of efficient front, in: Procedia. Comput. Sci., Springer. pp. 310–317.

Sobol', I.M., Asotsky, D., Kreinin, A., Kucherenko, S., 2011. Construction and Comparison of High-Dimensional Sobol' Generators. Wilmott 2011, 64–79. doi:10.1002/wilm.10056.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50205-0

Data-driven scenario generation for two-stage stochastic programming

Georgios L. Bounitsis[a], Lazaros G. Papageorgiou[a], Vassilis M. Charitopoulos[a*]

[a]*Department of Chemical Engineering, Centre for Process Systems Engineering, University College London, Torrington Place, London WC1E 7JE, UK*
v.charitopoulos@ucl.ac.uk

Abstract

The efficient exploitation of large amount of data for the uncertain parameters constitutes a crucial condition for effectively handling stochastic programming problems. In this work we propose a novel data-driven mixed-integer linear programming (MILP) model for the Distribution Matching Problem (DMP). In cases of multiple uncertain parameters, sampling using copulas is conducted as preliminary step. The integration of clustering methods and DMP in the proposed model is proven to improve the computational efficiency. For the evaluation of the performance of the proposed scenario generation approaches several case studies of a two-stage stochastic programming problem are examined. Compared with state-of-the-art scenario generation (SG) approaches the proposed model is shown to achieve consistently the lowest errors regarding the expected values when compared to full-space stochastic solutions as well as manages to preserve good accuracy in the resulting probabilistic and statistical qualities of the reduced generated sets.

Keywords: Scenario Generation; Stochastic Programming; Distribution Matching; Mixed-Integer Linear Programming (MILP).

1. Motivation

The increasing volatility in modern-day process industries and the access to large amounts of historical data have led to an intensive study of optimisation problems under uncertainty (Li and Grossmann, 2021). Among the various optimisation-based approaches, these problems can be solved as two-stage or multi-stage stochastic programming problems, in which the uncertain parameters are considered though a discrete number of their possible realisations. These realisations are also referred to as scenarios. Although such stochastic programming models can be easily formulated, their ability to capture the uncertainty relies on the number of scenarios considered which in turn tends to grow exponentially with the number of parameters under study. To this end, there has been an increasing interest from the research community, aiming at either the reduction of the uncertainty set or the generation of a representative and smaller in size set of scenarios to be implemented in the problem (Römisch, 2009).

OSCAR (Li and Floudas, 2014) and SCANCODE (Medina-González et al., 2020) are two recently presented methods for scenario aggregation which employ clustering techniques and distance metrics to preserve the quality of the stochastic solution whilst minimizing the number of required scenarios. With respect to scenario generation approaches, a nonlinear programming-based moment and distribution matching method was presented by Calfa et al. (2014). In this work, we employ an initial uncertain

sample and select a reduced set which matches optimally the statistical moments and the cumulative distribution functions of the marginal distributions. In case of multiple uncertain parameters, preliminary copula-based sampling is employed to account for correlations among multiple parameters. The remainder of the article is organised as follows: in Section 2 a brief literature review and problem statement are outlined while key methodological contributions are outlined in Section 3. In Section 4, we employ the proposed model and compare its performance with state-of-the-art models for scenario generation. Finally, conclusions are drawn in Section 5.

2. Literature review

Optimisation problems under uncertainty are dealt in the open literature with a wide range of mathematical techniques including stochastic programming (Sahinidis, 2004). The most common version of the risk-neutral stochastic programming is the two-stage stochastic programming, in which the target is to optimise the objective function of the first-stage costs while optimizing the expected value of the second-stage costs when uncertainty is revealed. The second stage expected value, when there are known probability distributions of the uncertain parameters, i.e., scenarios $\xi_1, \ldots, \xi_k$, with respective probabilities $p_1, \ldots, p_k$, is formulated as:

$$\mathbb{E}[F(x,\xi)] = \sum_{k=1}^{K} p_k \cdot F(x,\xi_k) \tag{1a}$$

and analogously the two-stage stochastic programming problem (TSSP) is modelled as:

$$\begin{aligned} \min_{x,y_1,\ldots,y_k} \quad & c^\top \cdot x + \sum_{k=1}^{K} p_k \cdot q_k^\top \cdot y_k \\ s.t. \quad & T_k \cdot x + W_k \cdot y_k = h_k, \quad y_k \geq 0, \quad k = 1,\ldots,K \\ & A \cdot x = b, \quad x \geq 0 \end{aligned} \tag{1b}$$

In the above formulation, every scenario $\xi_k = (q_k, T_k, W_k, h_k), k = 1,\ldots,K$, results to a two-stage decision vector y_k and by solving the two-stage problem an optimal first-stage solution $\bar{x}$ is obtained (Shapiro et al., 2014). The existence of a very large number of possible realisations could render the problem computationally intractable and motivates the use of scenario generation and/or reduction techniques. These methods aim at the creation of a smaller in size set of scenarios for the uncertain parameters, with certain values and probabilities which are representative of the original uncertainty set (Li and Grossmann, 2021).

Moment Matching Problem (MMP) constitutes a well-known scenario tree generation approach (Høyland et al., 2003). It is based on the minimisation of the errors regarding the statistical moments between the original uncertain set and the final reduced set. In general, MMPs are modelled as nonlinear programming (NLP) problems. Moreover, the parallel matching of the stochastic distribution of the uncertain parameter (DMP), by minimizing the errors regarding the cumulative density function, has been proposed and enhances the effectiveness of the approach (Calfa et al., 2014). Recently an MMP MILP model for scenario selection from an original scenario set was proposed in the literature (Kaut, 2021).

Copula based capture of the multivariate structure of data and dependence between marginal distributions constitutes another concept for the scenario generation methods (Kaut and Wallace, 2011).

3. Methodology and mathematical developments

MMP constitutes an error minimisation problem where the errors regarding the first four statistical moments, e.g. mean, variance, skewness, and kurtosis are considered. In particular, the errors between the values of the moments calculated from the data and the ones computed by the final reduced set are minimised. Finally, the errors can be quantified by various distances such as the Euclidean distance (L^2-norm), Manhattan distance (L^1-norm) or the Chebyshev distance (L^∞-norm) and weights may be considered for the errors of different moments (Calfa et al., 2014). An extended version of the MMP, referred as Distribution Matching Problem (DMP), matches additionally the marginal Empirical Cumulative Distribution Function (ECDF) between the reduced and the original dataset. Calfa et al. (2014) employed a nonlinear approximation of the ECDF using simplified Generalised Logistic Function.

The proposed MILP model considers the original distributions of the uncertain parameters to generate discrete original scenarios as input for the model. Considering multidimensional data sets, copula-based sampling is implemented, generating tuples as original scenarios. Theoretically, an n-dimensional copula expresses the joint cumulative distribution function (CDF) for which the marginal distributions of each variable are uniform on the interval $[0,1]$. Denoting multivariate CDF as F and marginal distributions as $F_1, \ldots, F_n$, the copula C is defined as (Kaut and Wallace, 2011):

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)) \tag{2}$$

The dependence between two variables can be captured by various parametric bivariate copula families, e.g. Gaussian, Archimedean and Student t. For multivariate cases, appropriate pairwise copula families along with the structure of the data set can be considered leading to a vine copula.

For each uncertain parameter $i \in I$ let us consider a set of equiprobable original scenarios $n \in N$ with values denoted as $\tilde{x}_{i,n}$ and ECDF as $ECDF_{i,n}$. In a pre-processing step k-means clustering is utilised to cluster the original set into as many clusters as the desirable size of the reduced set $c \in C$. Hence, we introduce a set $CL_{c,n}$, which maps each scenario $n \in N$ to one $c \in C$. From the original scenarios only one is selected at each prespecified cluster $c \in C$, through binary variables $y_{c,n}$. The corresponding probabilities of occurrence are denoted as $p_{c,n}$ and the cumulative probability sums up to 1. The latter are imposed by Eqs. (4)-(7). For the presented models only L^1-norm or L^∞-norm are used to quantify the errors. The objective function in the following formulation contains the errors regarding the statistical moment using L^1-norm and the summation of the maximum errors of the selected scenarios regarding the ECDF.

$$\min_{y_{c,n}, p_{c,n}} \quad \sum_{i \in I, m \in M} W_m^{sm} \cdot \left(d_{i,m}^+ + d_{i,m}^-\right) + \sum_{i \in I} W_i^{prob} \cdot ed_i \tag{3}$$

s.t.

$$\sum_{n \in CL_{c,n}} y_{c,n} = 1 \qquad \forall\, c \in C \tag{4}$$

$$\sum_{c \in CL_{c,n}} y_{c,n} \leq 1 \qquad \forall n \in N \tag{5}$$

$$P^{min} \cdot y_{c,n} \leq p_{c,n} \leq P^{max} \cdot y_{c,n} \qquad \forall c \in C, n \in N \tag{6}$$

$$\sum_{c,n \in CL_{c,n}} p_{c,n} = 1 \tag{7}$$

$$\sum_{c,n \in CL_{c,n}} \tilde{x}_{i,n} \cdot p_{c,n} + d_{i,m}^{+} - d_{i,m}^{-} = \widetilde{M}_{i,m} \qquad \forall i \in I, m = 1 \tag{8}$$

$$\sum_{c,n \in CL_{c,n}} \left(\tilde{x}_{i,n} - \widetilde{M}_{i,1}\right)^{k} \cdot p_{c,n} + d_{i,m}^{+} - d_{i,m}^{-} = \widetilde{M}_{i,m} \qquad \forall i \in I, m > 1 \tag{9}$$

$$\sum_{c \in C} y_{c,n} \cdot ECDF_{i,n} - \sum_{c,n\prime \in CL_{c,n\prime} \wedge x_{i,n\prime} \leq x_{i,n}} p_{c\prime,n} = \Phi_{i,n} \qquad \forall i \in I, n \in N \tag{10}$$

$$ed_i \geq \Phi_{i,n} - \left(1 - \sum_{c \in CL_{c,n}} y_{c,n}\right) \qquad \forall i \in I, n \in N \tag{11}$$

$$ed_i \geq -\Phi_{i,n} - \left(1 - \sum_{c \in CL_{c,n}} y_{c,n}\right) \qquad \forall i \in I, n \in N \tag{12}$$

$$-1 \leq \Phi_{i,n} \leq 1 \qquad \forall i \in I, n \in N \tag{13}$$

$$ed_i \geq 0 \qquad \forall i \in I \tag{14}$$

$$0 \leq p_{c,n} \leq 1 \qquad \forall c \in C, n \in N \tag{15}$$

$$y_{c,n} \in \{0,1\} \qquad \forall c \in C, n \in N \tag{16}$$

$$d_{i,m}^{+}, d_{i,m}^{-} \geq 0 \qquad \forall i \in I, m \in M \tag{17}$$

Eqs. (8)-(9) calculate the moments of the selected scenarios and the corresponding errors. Eq. (10) defines to variables $\Phi_{i,n}$ the deviations regarding the ECDF curve till the occurrence of each data point $n \in N$ of each uncertain parameter $i \in I$. However, Eqs. (11)-(12) are implemented, along with the minimization objective function Eq. (3), to define the maximum absolute errors regarding each parameter $i \in I$ to variables ed_i.

For the case of one uncertain parameter ($|I| = 1$) and sorted original data points, the computational efficiency is drastically enhanced by substituting $\Phi_{i,n}$ to $\bar{\Phi}_{i,c}$ and Eq. (10)-(13) to the following ones:

$$\sum_{n \in N} y_{c,n} \cdot ECDF_{i,n} - \sum_{n,c\prime \in CL_{c\prime,n}} p_{c\prime,n} = \bar{\Phi}_{i,c} \qquad \forall i \in I, c \in C \tag{18}$$

$$ed_i \geq \bar{\Phi}_{i,c} \qquad \forall i \in I, c \in C \tag{19}$$

$$ed_i \geq -\bar{\Phi}_{i,c} \qquad \forall i \in I, c \in C \tag{20}$$

$$\sum_{n \in CL_{c,n}} y_{c-1,n} \cdot \tilde{x}_{i,n} \leq \sum_{n \in CL_{c,n}} y_{c,n} \cdot \tilde{x}_{i,n} \qquad \forall i \in I, c > 1 \tag{21}$$

$$-1 \geq \bar{\Phi}_{i,c} \geq 1 \qquad \forall i \in I, c \in C \tag{22}$$

Weights of the errors regarding the moments are calculated as $W_{i,m}^{sm} = \bar{w}_{i,m} / |\widetilde{M}_{i,m}|$, where $\bar{w}_{i,m}$ is chosen arbitrarily equal to 1 in this report. W_i^{prob} is also considered equal to 1. The presented formulation will be denoted as DMP MILP using L^1-norm. When

the model is reformulated to minimise the summation of the maximum absolute errors regarding the moments (of all $i \in I, m \in M$) and the maximum absolute ECDF error (of all $i \in I$), we refer to it as DMP MILP using L^{∞}-norm.

4. Case studies – Capacity Planning under uncertainty

In this section, a capacity planning problem under uncertainty is studied. The mathematical model along with the process description, data and different uncertain cases can be found in Li and Floudas (2014). Briefly, 5 products are produced by 5 raw materials using 11 candidate processes. The objective function aims at the maximisation of profit. We consider endogenous uncertainty, regarding production yields of processes. For case study 1 is considered only one uncertain parameter and the scenario set is reduced from 1,000 data points of the original uncertain distribution to 5 selected scenarios. For case studies 2 and 3 we consider 2 and 4 uncertain parameters, generating 20 scenarios by an initial set of 1,000 and 2,000 copula-based generated scenarios respectively. Regarding the computational efficiency of the SG approaches, in the first case study the presented models are solved to optimality after short execution times (<10 seconds) using GUROBI 9.1 in GAMS 30.3. As far as case studies are considered, a time limit of 1,800s and an optimality gap tolerance of 5% are set.

To evaluate the quality of the scenarios generated by the different algorithms the expected value of the two-stage stochastic program is computed using the reduced and the full-space set of scenarios. The results regarding the errors for each case study are summarised in Figure 1.

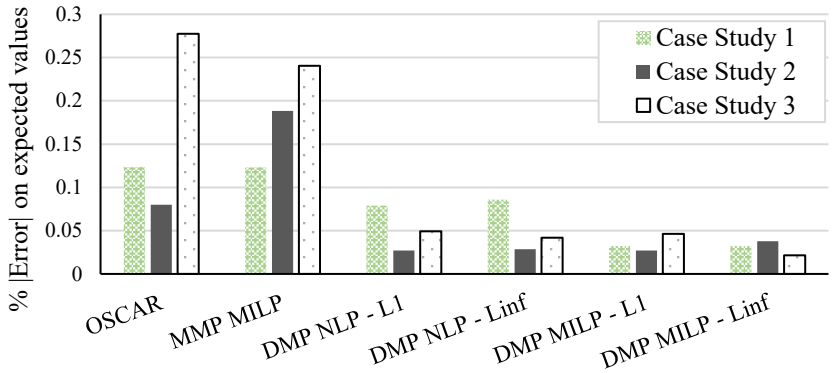


Figure 1: *Errors on expected values of the TSSP problem for each case study.*

The scenarios of the proposed DMP MILP model consistently result to the lowest errors. The second case study considers a reduction from 1000 original sets to 20 and DMP models seem to behave similarly. However, DMP MILP using L^1-norm results to the lowest error. For the third case study with four uncertain parameters DMP MILP using L^{∞}-norm leads to the lowest error.

The proposed approach integrating clustering and distribution matching resolves the issue of under-specificity of original MMP (Calfa et al., 2014). The combined impact prevents the assignment of a value to multiple scenarios and/or zero probabilities for scenarios regardless of the number of the prespecified scenarios/clusters. Overall, although higher total error regarding the moments may be obtained, the ECDF is matched properly and the results indicate enhanced performance of the reduced sets.

The impact of ECDF matching through this approach on the reduced set is visualised in Figure 2. For instance, in case study 2 and the uncertain parameter regarding Process "7", the reduced set of 20 final scenarios obtained by the proposed model matches the ECDF of the original marginal distribution remarkably better than the set obtained by using OSCAR. It is noted that for OSCAR the explicit ECDF marching is not considered.

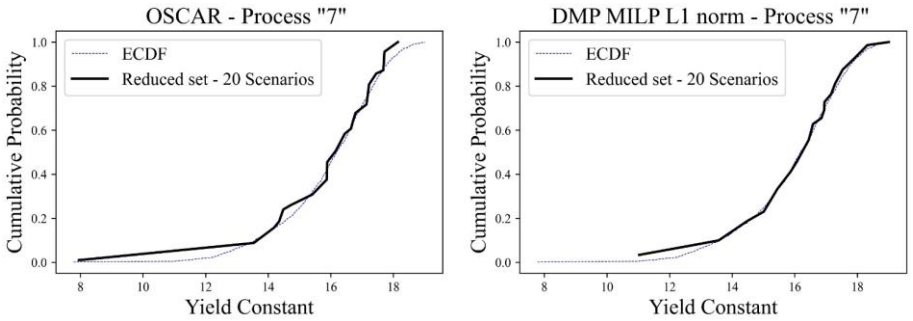


Figure 2: *ECDF matching comparison on case study 2.*

5. Conclusions and future work

The proposed MILP model is proven to be competitive to the existing scenario generation/reduction approaches in terms of errors regarding TSSP problems and statistical properties matching. The integration of copula-based simulation in the framework constitutes a crucial asset as this is not feasible for NLP formulations which also suffer from numerical and under-specificity issues. Future work within our group focuses on the use of decomposition techniques to enhance the computational performance of the proposed model.

References

B. Calfa, A. Agarwal, I. Grossmann & J. Wassick, 2014. Data-driven multi-stage scenario tree generation via statistical property and distribution matching, Comput. Chem. Eng., 68, 7–23.

K. Høyland, M. Kaut & S. Wallace, 2003. A heuristic for moment-matching scenario generation, Comput. Optim. Appl., 24(2–3), 169–185.

M. Kaut, 2021. Scenario generation by selection from historical data, Computational Manag. Sci., 18(3), 411–429

M. Kaut & S. Wallace, 2011. Shape-based scenario generation using copulas, Comput. Manag. Sci., 8(1), 181–199.

C. Li & I. Grossmann, 2021. A Review of Stochastic Programming Methods for Optimization of Process Systems Under Uncertainty, Front. Chem. Eng., 2, 34.

Z. Li & C. Floudas, 2014. Optimal scenario reduction framework based on distance of uncertainty distribution and output performance: I. Single reduction via mixed integer linear optimization, Comput. Chem. Eng., 70, 50–66.

S. Medina-González, I. Gkioulekas, V. Dua & L. G. Papageorgiou, 2020. A graph theory approach for scenario aggregation for stochastic optimisation. Comp. Chem. Eng., 137, 106810.

W. Römisch, 2009. Scenario reduction techniques in stochastic programming. In Stochastic Algorithms: Foundations and Applications, 5, 1–14.

N. Sahinidis, 2004. Optimization under uncertainty: State-of-the-art and opportunities. Comput. Chem. Eng., 28(6–7), 971-983.

A. Shapiro, D. Dentcheva & A. Ruszczyński, 2014. Lectures on stochastic programming: modeling and theory. Philadelphia, Pennsylvania: SIAM.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50206-2

Joint Chance Constrained Process Optimization through Neural Network Approximation

Shu-Bo Yang[a], Jesús Moreira[b], Zukui Li[a*]

[a]*Department of Chemical and Materials Engineering, University of Alberta, 9211 116 St, Edmonton T6G1H9, Canada*
[b]*Imperial Oil, 505 Quarry Park Blvd SE, Calgary, T2C5N1, Canada*
zukui@ualberta.ca

Abstract

A neural network-based approach is proposed in this work for joint chance-constrained optimization (JCCP) problems. In the proposed approach, a joint chance constraint (JCC) is first reformulated as a quantile-based inequality to reduce the complexity in approximation. Then, the quantile function (QF) in the inequality is replaced by an empirical QF through sample average approximation. The empirical QF is further approximated by a ReLU artificial neural network (ANN). Afterwards, the ReLU ANN is incorporated into the optimization model that enables the JCCP to be solved as a deterministic optimization problem. To demonstrate the proposed approach, a case study on ethylene glycol (EG) production process yield maximization is studied. The results show that the proposed approach can efficiently solve a nonlinear JCCP problem with non-conservative constraint satisfaction.

Keywords: Machine Learning and Big Data; Artificial Neural Network; Joint Chance Constrained Optimization; Stochastic Optimization; Sample Average Approximation

1. Introduction

Practical process optimization often faces uncertainties. Chance constrained optimization is a popular technique for addressing uncertainty (Wendt et al., 2002). It enforces that the optimal solution should satisfy the uncertain constraint with a certain probability level. There are two types of chance constraint: the individual chance constraint (ICC) and the joint chance constraint (JCC). The JCC is more general in engineering applications than the ICC since the JCC ensures that all constraints are satisfied simultaneously to a certain confidence level, which is more natural in many applications (You et al., 2021). However, the JCC is generally difficult to solve as it requires dealing with multidimensional distributions. Thus, joint chance-constrained optimization problems (JCCP) are generally solved through approximations. There are two main approximation methods: analytical approximation methods and sampling-based methods (Yuan et al., 2017).

We focus on sampling-based methods in this work because they can avoid overly conservative solutions occurring in analytical approximation methods (van Ackooij et al., 2014). Among all sampling-based methods, the sample average approximation (SAA) is a powerful and widely used method (Pagnoncelli et al., 2009). While adopting the SAA to address a JCCP problem, the empirical joint constraint satisfaction probability is enforced to be greater than or equal to the required value. In this work, we extend the concept of the SAA to construct the empirical quantile function (QF) in the

proposed approach. The key contributions of the proposed approach are as follows: The presented approach involving the empirical QF and ReLU artificial neural network (ANN) approximations can generate a solution without over-conservatism for a JCCP. Also, the presented method can be widely applied to different problems, including those with black-box functions. Finally, the proposed approach can make a stochastic JCCP deterministically solvable.

2. Problem formulation

The general formulation of a JCCP problem is given as:

$$\begin{aligned} &\min_{x} f(x) \\ &\text{s.t.} \quad \Pr(g_i(x,\xi) \le 0,\ i = 1,\dots,p) \ge 1-\varepsilon \end{aligned} \tag{1}$$

where x represents the decision variable. $f(x)$ and $g(x,\xi)$ are the objective and constraint functions. ξ is the uncertain parameter vector. $\Pr(\cdot)$ is the probability measure. The JCCP shown above enforces that all constraints $g_{i=1,\dots,p}(x,\xi)$ are satisfied simultaneously to a certain confidence level $1-\varepsilon$.

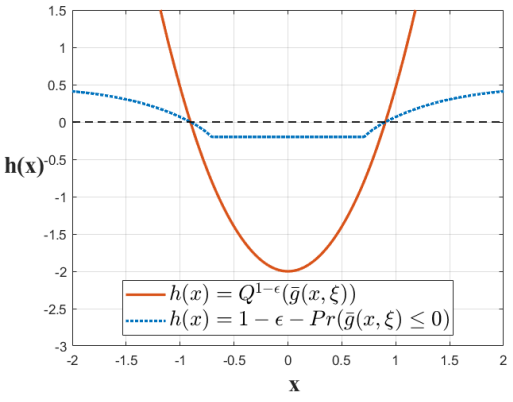
The above JCC can be rewritten as the following individual chance constraint form:

$$\Pr(\bar{g}(x,\xi) \le 0\ ) \ge 1-\varepsilon, \quad \text{with} \quad \bar{g}(x,\xi) = \max_{i=1,\dots,p} g_i(x,\xi) \tag{2}$$

which can be further reformulated as the following equivalent constraint:

$$Q^{1-\varepsilon}\big(\bar{g}(x,\xi)\big) \le 0 \tag{3}$$

where $Q^{1-\varepsilon}(\bar{g}(x,\xi))$ is the $1-\varepsilon$ quantile of $\bar{g}(x,\xi)$. The benefit of rewriting the JCC as the quantile-based form is shown through Figure 1: $Q^{1-\varepsilon}(\bar{g}(x,\xi))$ has better convexity property than $1-\varepsilon-\Pr(\bar{g}(x,\xi) \le 0\ )$. Accordingly, $Q^{1-\varepsilon}(\bar{g}(x,\xi))$ will be used for surrogate modelling in the proposed work.



$$\begin{aligned} &1-\varepsilon = 0.8 \\ &\xi_1 \sim N(0,1),\ \ \xi_2 \sim U(-2,2) \\ &\bar{g}(x,\xi) \\ &= \max\{g_1(x,\xi_1), g_2(x,\xi_2)\} \\ &g_1(x,\xi_1) = 1.5\xi_1 x^2 - 3 \\ &g_2(x,\xi_2) = 2\xi_2 x^2 - 2 \end{aligned}$$

Figure 1. Comparison between $Q^{1-\varepsilon}(\bar{g}(x,\xi))$ and $1-\varepsilon-\Pr(\bar{g}(x,\xi) \le 0)$.

The quantile $Q^{1-\varepsilon}(\bar{g}(x,\xi))$ in Eq.(3) can be approximated by the empirical QF value $\tilde{Q}^{1-\varepsilon}(\bar{g}(x,\xi))$ given as:

$$\tilde{Q}^{1-\varepsilon}\left(\bar{g}(x,\xi)\right) = \inf\left\{\gamma \middle| \frac{1}{N}\sum_{j=1}^{N} \mathbb{I}\left(\bar{g}(x,\xi_j) \le \gamma\right) \ge 1-\varepsilon\right\} = \bar{g}_{\lceil M \rceil}(x) \tag{4}$$

where M equals to $N(1-\varepsilon)$, and N is the number of collected samples of ξ. $\mathbb{I}$ is the indicator function defined as $\mathbb{I}\left(\bar{g}(x,\xi_j) \le \gamma\right) = \begin{cases} 0, & \text{for } \bar{g}(x,\xi_j) > \gamma \\ 1, & \text{for } \bar{g}(x,\xi_j) \le \gamma \end{cases}$. $\bar{g}_{\lceil M \rceil}(x)$ represents the M-th smallest component of $\{\bar{g}(x,\xi_1), \dots, \bar{g}(x,\xi_N)\}$.

3. Neural network approximation-based optimization

The quantile term $\tilde{Q}^{1-\varepsilon}\left(\bar{g}(x,\xi)\right)$ in Eq.(4) can be further approximated by an ANN which is essentially a nonlinear function of x. The max operator for defining $\bar{g}\left(x,\xi_j\right)$ and the indicator function involved in Eq.(4) can be well approximated through ANN.

The ReLU ANN is one type of fully connected feed-forward neural network which is shown in Figure 2. In Figure 2, layer 0 and layer K are the input and output layers, respectively. The rest of the layers are namely hidden layers. $x_1, \dots, x_l$ and $\hat{Y}$ are the inputs and output (predicted QF value). Besides the input layer, all the neurons in each layer are fully connected with all the neurons in the previous layer. The output of one neuron in each layer (except the input layer) is computed using the following equation:

$$o_s^k = \sigma\left(\sum_{r=1}^{R} W_{rs}^k o_r^{k-1} + b_s^k\right), \quad k = 1, \dots, K \tag{5}$$

where k, r, and s are the indices of each layer, each neuron in the previous layer, and each neuron in the current layer, respectively. o, W, b, and R represent the output of the neuron, weights between the current and previous layers, the bias for each neuron in the current layer, and the number of neurons in the previous layer, respectively. σ represents the ReLU activation function given as: $\sigma(y) \coloneqq \max\{0, y\}$.

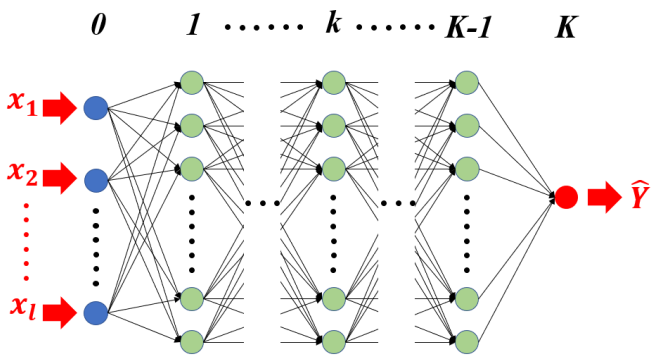


Figure 2. Schematic diagram of fully connected feed-forward neural network

The ReLU ANN for approximating the empirical QF is incorporated into the optimization model in Eq.(1) to predict the quantile value for the quantile-based inequality reformulated from the original JCC. Since the ReLU ANN model can be written as a mixed-integer linear program (MILP), the optimization model involving the ReLU ANN can be written as the formulation shown in Eq.(6).

$$
\begin{aligned}
&\min_{x} f(x) \\
&\text{s.t.} \quad \hat{Y} \le 0 \\
&\quad o_l^0 = x_l, \;\; l = 1, \dots, L \\
&\quad a_s^k = \sum_{r=1}^{R} W_{rs}^k o_r^{k-1} + b_s^k, \qquad s = 1, \dots, S, \qquad k = 1, \dots, K-1 \\
&\quad 0 \le o_s^k \le H(1 - z_s^k) \\
&\quad a_s^k \le o_s^k \le a_s^k + H z_s^k \\
&\quad z_s^k \in \{0,1\} \\
&\quad \hat{Y} = \textstyle\sum_{r=1}^{R} W_r^K o_r^{K-1} + b^K
\end{aligned}
\tag{6}
$$

In the above formulation, $\hat{Y}$ is the quantile value predicted from the embedded ReLU ANN. The embedded ReLU ANN is described by the second to the last constraints. The second and last constraints represent input and output layers, respectively. l is the index of each neuron in the input layer. L is the number of neurons in the input layer. S and R are the numbers of neurons in the current layer and in the previous layer, respectively. a_s^k is the linear combination of outputs from the previous layer. H is a big number. z_s^k is a binary variable. The third to the sixth constraints represent the hidden layer model. The fourth to sixth constraints are for the ReLU activation function. The proposed approach for solving JCCP is based on the solution of the above deterministic problem.

Notably, since $\bar{g}(x, \xi) = \max_{i=1,\dots,p} g_i(x, \xi)$ is used in the quantile-based reformulation, the proposed approach can be applied to the problem with any number of constraints in the JCC. In addition, the proposed method is applicable to different problems, including those with black-box functions.

4. Case Study

This case study is the yield maximization of the EG production process shown in Figure 3. According to Figure 3, ethylene oxide (EO) and water are fed into the CSTR to produce EG. The following reactions occur in the CSTR:

$$
\begin{aligned}
&EO + H_2O \rightarrow EG \\
&EG + EO \rightarrow DEG \\
&DEG + EO \rightarrow TEG
\end{aligned}
\tag{7}
$$

DEG and TEG are diethylene glycol and triethylene glycol, respectively, which are undesirable by-products. The output of the CSTR is connected to the flash separator, and the product (EG) is collected from the bottom of the flash separator. Meanwhile, 97 % of the top flash vapor stream is recycled to mix with the feed flow. The vapor-liquid equilibrium in this process is assumed as ideal phase behaviour. More details about this process can be found in (Kahrs et al., 2007). In this case study, the entire EG production process is simulated as a nonlinear black-box model using Aspen Plus.

In this problem, the objective is to maximize the EG molar flowrate ($\dot{F}_{EG}$, unit: kmol/h) in the bottom of the flash separator. The decision variables $\dot{F}_w$, V_{CSTR}, and T_f are the water feed flowrate (unit: kmol/h), the CSTR volume (unit: m^3), and the temperature in

the flash separator (unit: K), respectively. The pressure inside the flash separator is fixed at 30 kPa. X_{EO} and x_{EG} are the EO conversion rate and the mole fraction of EG in the bottom flow of the flash separator, respectively. $\xi_{X_{EO}}$ and $\xi_{x_{EG}}$ are uncertain parameters following Gaussian distributions $N(0,0.03)$ and $N(0,0.015)$, respectively.

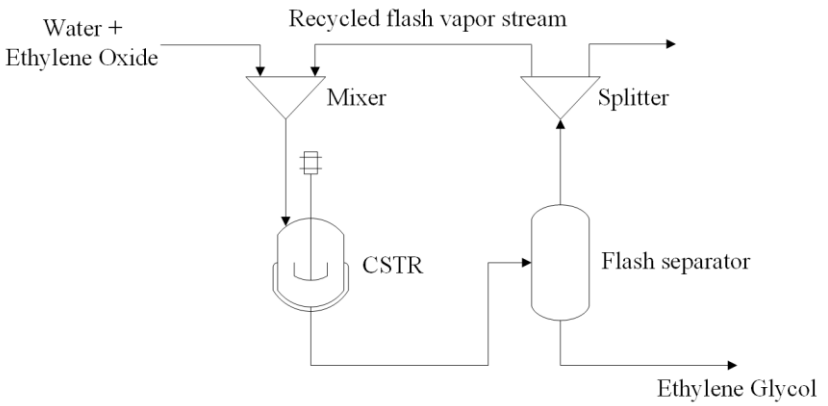


Figure 3. Flowsheet of the studied EG production process

The optimization problem of this case study is given as:

$$\begin{aligned} &\max_{\dot{F}_w, V_{CSTR}, T_f} \ \dot{F}_{EG} \\ &\text{s.t.} \ \ \Pr\begin{pmatrix} -X_{EO} - \xi_{X_{EO}} + 0.5 \le 0 \\ -x_{EG} - \xi_{x_{EG}} + 0.5 \le 0 \end{pmatrix} \ge 1 - \varepsilon \\ &\qquad 700 \le \dot{F}_w \le 5000 \\ &\qquad 221 \le V_{CSTR} \le 321 \\ &\qquad 300 \le T_f \le 390 \end{aligned} \tag{8}$$

We first gather 2000 samples of $\dot{F}_{EG}$, X_{EO}, and x_{EG} corresponding to different sets of decision variables, from the Aspen Plus simulator. Then, we collect 1000 samples of $[\xi_{X_{EO}}, \xi_{x_{EG}}]$from the assumed distributions. The 2000 samples of $\dot{F}_{EG}$ corresponding to different decision variables are used to train the first ReLU ANN to predict $\dot{F}_{EG}$ in the objective function. For addressing the JCC in the above optimization, based on a confidence level $1-\varepsilon$, the 2000 samples of X_{EO} and x_{EG} are combined with 1000 samples of $[\xi_{X_{EO}}, \xi_{x_{EG}}]$ to generate 2000 quantile values based on Eq.(4). Then, we use the 2000 quantile values paired with different decision variables to train the second ReLU ANN to predict the quantile value for the quantile-based inequality reformulated from the JCC. Finally, the two mentioned ReLU ANNs are incorporated into the above optimization to form a MILP problem based on the optimization model shown in Eq.(6). The MILP problem is solved by using CPLEX in GAMS. The attained results are shown in Table 1.

Table 1. Optimization results

$1-\varepsilon$	$\dot{F}_w$	V_{CSTR}	T_f	$\hat{F}_{EG}/\dot{F}_{EG}$	$\Pr(\bar{g} \leq 0)$
0.8	1048.887	321	365.415	427.134 / 426.261	0.8006
0.95	916.229	321	368.191	396.610 / 395.247	0.9508

[a] $\hat{F}_{EG}$ is computed by the first ReLU ANN. $\dot{F}_{EG}$ is from the Aspen Plus simulator.

[b] $\bar{g} = \max\{-X_{EO} - \xi_{X_{EO}} + 0.5,\ -x_{EG} - \xi_{x_{EG}} + 0.5\}$ in the last column.

[c] The probabilities in the last column are based on 10^6 samples of uncertain parameters.

According to Table 1, the joint chance constraint satisfaction probabilities of the optimal solutions (in the last column) are very close to the required probability targets (in the first column). Also, the $\hat{F}_{EG}$ computed from the first ReLU ANN are very close to the corresponding $\dot{F}_{EG}$ computed from the true model (Aspen Plus simulation). Therefore, by using the presented method, the JCCP in this case study can be reliably handled without over-conservatism.

5. Conclusions

A novel method involving the empirical quantile reformulation of JCC and the ReLU ANN approximation is proposed in this work to address JCCP problems. The presented method relies on the sampled data of objective and constraint function values, which can be obtained from explicit process model equations or black-box process simulators. Hence, the method can be used for different process optimization problems, such as problems with explicit model or black-box constraint functions. In the meantime, there is no restriction on the number of constraints in the JCC. The problem is finally converted to the solution of a deterministic MILP problem. Through a case study, it is shown that the developed method can efficiently solve a nonlinear joint chance-constrained process optimization problem without over-conservatism.

References

O. Kahrs, and W. Marquardt, 2007, The validity domain of hybrid models and its application in process optimization, Chemical Engineering and Processing: Process Intensification, 46(11), 1054-1066

B. K. Pagnoncelli, S. Ahmed, and A. Shapiro, 2009, Sample average approximation method for chance constrained programming: theory and applications, Journal of optimization theory and applications, 142(2), 399-416

W. van Ackooij, R. Henrion, A. Möller, and R. Zorgati, 2014, Joint chance constrained programming for hydro reservoir management, Optimization and Engineering, 15(2), 509-531

M. Wendt, P. Li, G. Wozny, 2002. Nonlinear chance-constrained process optimization under uncertainty. Industrial & engineering chemistry research, 41(15), 3621-3629.

B. You, E. Esche, J. Weigert, and J.-U. Repke, 2021, Joint Chance Constraint Approach based on Data-Driven Models for Optimization Under Uncertainty applied to the Williams-Otto Process, Computer Aided Chemical Engineering, Vol. 50, pp. 523-528, Elsevier.

Y. Yuan, Z. Li, and B. Huang, 2017, Robust optimization approximation for joint chance constrained optimization problem, Journal of Global Optimization, 67(4), 805-827

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50207-4

Gaussian Processes for Simulation-Based Optimization and Robust Design

Alex Durkin[a], Marcos Millan-Agorio[a], Miao Guo[a,b*]

[a]*Department of Chemical Engineering, Imperial College London, London SW7 2AZ, UK*
[b]*Department of Engineering, King's College London, London WC2R 2LS, UK*
miao.guo@imperial.ac.uk

Abstract

Gaussian Processes present a versatile surrogate modeling toolbox to address simulation-based optimization and uncertainties arising from non-converged simulations. In this work we present a black-box optimization methodology framework in which Gaussian Process Regression is used to model complex underlying process performance models and Gaussian Process Classification is used to model feasibility constraints based on converged and non-converged simulations. Additionally, we present a conservativeness parameter to enable tuning of the feasible region based on the trade-off between process performance and the risk of infeasibility due to non-converged simulations.

Keywords: Gaussian Processes, Optimization, Surrogate Modeling.

Introduction

Modeling complex process systems has long been of interest in industry and the research community (Sahinidis, 2004). As a result, commercial simulation software has been developed, enabling engineers to quickly evaluate process designs without the capital costs associated with pilot studies (Cozad et al., 2014). Such simulation-based process design can be categorised into decomposition techniques (Douglas, 1988) and optimization methods (Bhosekar and Ierapetritou, 2018). The former involves sequentially designing each stage of the process using design heuristics to obtain a quick solution. Conversely, optimization methods simultaneously design stages of the process, offering guarantees of optimality at the expense of computational cost.

Simulation-based optimization is also known as black-box optimization where the simulator is treated as a black-box, embedding complex underlying functions not accessible to the user. Black-box optimization is widely used for the optimization of complex systems and generally follows either an evolutionary approach or a surrogate-based approach. The former uses heuristics to select subsequent samples in the search space until convergence, however, this approach has problems finding globally optimal solutions. In surrogate-based optimization, sampled data is used to fit a surrogate model, representative of the underlying complex model, and then rigorous optimization methods are used on the surrogate (Jones et al., 1998). However, fitting a surrogate model introduces additional uncertainty to be considered, between surrogate model predictions and the true underlying function.

Surrogate modeling can be partitioned into 2 stages: sampling and surrogate model fitting. The former aims to sample input-output information from the black-box function whilst minimising function evaluations (Sacks et al., 1989). A trade-off between homogenous coverage of the design space and non-correlated input variables has led to quasi-random sampling techniques being developed such as Latin Hypercube Sampling (LHS) (McKay et al., 1979). For the choice of surrogate model, the trade-off between model accuracy and computational tractability within an optimization problem must be considered. Whilst linear and polynomial models offer computationally tractable surrogate models, they can fail to accurately represent non-linearities in the underlying complex model. Conversely, Gaussian Process Regression (GPR) or "kriging" has gained popularity as a more robust surrogate modeling framework, albeit with reduced computational tractability within optimization problems (Caballero and Grossmann, 2008).

An additional challenge of black-box optimization is addressing the fact that not all samples within the search space are guaranteed to return a converged result when

evaluated by the simulator. In response to this problem, feasibility constraints can be constructed to confine the optimization search space to a region where designs are more likely to have converged during sampling, and therefore are more likely to represent a feasible design. To build such feasibility constraints, classification machine learning models, such as support vector machines or Gaussian Process Classification (GPC), can be trained on the binary classification data from converged/non-converged simulations (Ibrahim et al., 2018).

In this work, we present a computational modeling framework harnessing GPR and GPC for simulation-based optimization and robust design against uncertainties in infeasibility due to non-converged simulations. Specifically, GPR was used to represent process performance in the optimization objective function, whilst GPC was used to represent process feasibility constraints. Additionally, we introduce a "conservativeness" parameter to enable tuning of the process robustness. The GPR and GPC models were embedded in an optimization problem to determine minimum cost solutions that are robust to the uncertainty from non-converged simulations.

This article proceeds with an overview of Gaussian Processes. We then present the methodology framework developed in this study followed by a demonstrative application. Finally, we discuss the results and make some concluding remarks on the implications of this work.

Gaussian Processes

Gaussian Processes (GPs) provide a framework for modeling complex black-box functions. By not assuming any model characteristics, GPs are not restricted to a fixed (and potentially incorrect) functional form, nor a finite set of parameters. Instead, GPs assume a prior probability distribution over an effectively infinite number of functions, where certain characteristics can be favored, and expert knowledge can be incorporated (for example favor smooth/periodic/noisy functions). The combination of the prior with some observed data yields a posterior distribution over functions, from which it is possible to extract a mean function and the variance in these functions. For the modeling of black-box functions, the GP posterior mean function can be used for predictive purposes whilst the GP posterior variance provides the uncertainty in these predictions.

The GP prior is defined via the specification of a covariance function of the GP which fixes some properties of the functions in the distribution and defines a set of hyperparameters that enable learning based on observed data. For a stochastic GP, $Z(\cdot)$, Eq.(1) shows the covariance function ($\mathbf{R}$ between two points $\boldsymbol{x}_i$, $\boldsymbol{x}_j$), for the squared exponential correlation function containing the tuneable parameter l_k, scaled by the tuneable process variance, σ_z^2. Specifically, l_k represents the sensitivity of the covariance function to dimension k, where a smaller value of l_k represents higher sensitivity, ensuring that even two points far away are correlated. Additionally, Eq.(1) has the property that if two points are very close, the covariance approaches the process variance, and as the distance between two points increases, the covariance approaches zero.

$$\mathbf{R}\left(Z(\boldsymbol{x}_i), Z(\boldsymbol{x}_j)\right) = \sigma_z^2 \exp\left(-\sum_{k=1}^{m} l_k \left(x_{i,k} - x_{j,k}\right)^2\right) \tag{1}$$

GPR models process input-output relationships, $y(\boldsymbol{x})$, by utilising a stochastic GP with covariance given by Eq.(1) and with expected value equal to zero, $E\left(Z(\boldsymbol{x})\right) = 0$, to model the deviations from a constant term, μ (Eq.(2)).

$$y(\boldsymbol{x}) = \mu + Z(\boldsymbol{x}) \tag{2}$$

Modeling the deviations from a constant term in this way, after optimizing the fit of hyperparameters (μ, σ_z^2, l_k) to observed data via Maximum Likelihood Estimation (MLE), it can be shown that the GPR posterior mean function (used for surrogate model predictions, $\hat{y}(\boldsymbol{x}^{\text{new}})$) and variance (used to quantify uncertainty in predictions, $s^2(\boldsymbol{x}^{\text{new}})$) are given by Eq.(3) and Eq.(4), respectively, where $\boldsymbol{r}$ is the vector of covariances between $\boldsymbol{x}^{\text{new}}$ and training data, obtained by evaluating Eq.(1) at $\mathbf{R}(\boldsymbol{x}^{\text{new}}, \boldsymbol{x})$, $\boldsymbol{y}$ is the vector of observations, and $\mathbf{1}$ is a vector of ones (6).

$$\hat{y}(\boldsymbol{x}^{\text{new}}) = \mu + \boldsymbol{r}^T \boldsymbol{R}^{-1}(\boldsymbol{y} - \mathbf{1}\mu) \tag{3}$$

$$s^2(\boldsymbol{x}^{\text{new}}) = \sigma_z^2(\mathbf{1} - \boldsymbol{r}^T \boldsymbol{R}^{-1}\boldsymbol{r}) \tag{4}$$

Eq.(3) and Eq.(4) have properties such that a new sample taken at the same point as a training observation results in a prediction equal to the value of that observation and a variance equal to zero. In other words, the GPR predictive function passes through all observations with no uncertainty. For new samples further away from training observations, the prediction approaches the constant term μ, influenced only by the nearest observations, and the prediction variance approaches the process variance.

GPC predicts the probability that a new sample $\boldsymbol{x}^{\text{new}}$ is feasible. To achieve this, the GPC model is trained on classification data such that a feasible converged simulation is labelled 1, whilst an infeasible non-converged simulation has class label 0. The prediction equation for GPC can then be generalised by Eq.(5) where the term on the left is the probability that a new sample, $\boldsymbol{x}^{\text{new}}$, has class label $t^{\text{new}} = 1$ (i.e. is feasible) given training data $\boldsymbol{x}, \boldsymbol{t}$. The term on the right of Eq.(5) is a latent function $f(\cdot)$, over which a GP prior is placed, "squashed" through the sigmoid logistic function $\sigma(\cdot)$ so that the outputs are probabilities between 0 and 1.

$$p(t^{\text{new}} = 1 \mid \boldsymbol{x}^{\text{new}}, \boldsymbol{x}, \boldsymbol{t}) = \sigma\big(f(\boldsymbol{x}^{\text{new}}, \boldsymbol{x}, \boldsymbol{t})\big) \tag{5}$$

To determine the analytical form of the GPC predictive function requires the approximation of analytically intractable integrals. Eq.(6) shows the predictive form used in this work, where the Laplace approximation and probit approximation have been used, and where v is the latent predictive variance, itself a function of $\boldsymbol{x}^{\text{new}}, \boldsymbol{x}, \boldsymbol{t}$ and the latent posterior mode $\hat{f}$ which was obtained numerically (Rasmussen and Williams, 2006).

$$p(t^{\text{new}} = 1\,) = \sigma\left(\boldsymbol{r}^T\left(\boldsymbol{t} - \sigma(\hat{f})\right)\left(1 + \frac{\pi}{8}v\right)^{-\frac{1}{2}}\right) \tag{6}$$

Method

Figure 1 shows the methodology framework developed in this study. Black-box simulation software was used to perform computer experiments. Samples were taken from the black-box simulations by varying input process design variables, $\boldsymbol{x}$, representing design decisions to optimize process performance. Output variables sampled were continuous process performance variables, $\boldsymbol{y}$, and binary classification targets representing converged/non-converged simulations, $\boldsymbol{t}$. The sampling strategy to produce these samples was "maximin" optimized LHS which ensures improved homogenous space-filling properties compared to traditional LHS (Johnson et al., 1990).

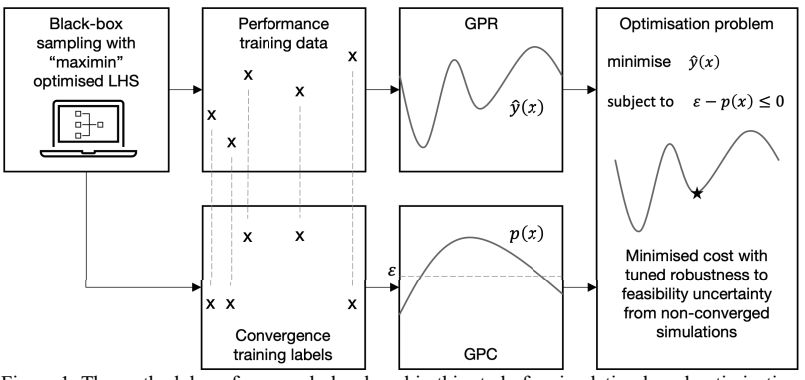


Figure 1. The methodology framework developed in this study for simulation-based optimization under uncertainty due to non-converged simulations. Computer experiments used an optimized Latin Hypercube Sampling (LHS) strategy. Gaussian Process Regression (GPR) and Gaussian Process Classification (GPC) formed process performance functions and feasibility constraints, respectively, in the resulting optimization problem.

Table 1. Optimization problem statement.

Given…	Samples	$i = 1, \ldots, n$
	Converged samples	$j = 1, \ldots, m$
	Input dimensionality	$k = 1, \ldots, q$
	Process cost training data for GPR	$x_{j,k}^{GPR}$, y_j
	Convergence target labels for GPC	$x_{i,k}^{GPC}$, t_i
	Fitted GPR parameters	μ, l_k^{GPR}, $R_{j,j'}^{-1}$
	Fitted GPC parameters	l_k^{GPC}, $\hat{f}_i$, $P_{i,i'}^{-1}$
	Conservativeness parameter	ε
Determine…	Process design variables	x_k^{new}
	Intermediate GPC variables	α_i, v
	GPC probability of feasibility	p
To optimize…	GPR predicted process cost	$\hat{y}$

Using sampled performance data and convergence target labels from computer experiments, GPR and GPC models were fitted by MLE, respectively. Specifically, only converged samples were used to train the GPR to obtain an accurate model within the feasible region without influence from non-converged samples. In this way, the complex black-box functions could be represented functionally within a mathematical optimization formulation whilst simultaneously restricting the search space to more probable feasible designs.

The functional predictive forms of the GPR and GPC models were embedded within a Non-Linear Programming (NLP) optimization problem. The matrix algebra of the GPR and GPC models, spanning vectors over training samples and input dimensions, were expanded into summation notation to ensure compatibility with NLP solver IPOPT (Wächter and Biegler, 2006), using Pyomo (Bynum et al., 2021). The optimization problem statement can be written as shown in Table 1 whilst the generalised NLP formulation is shown by Eq.(7) to Eq.(11).

Eq.(7) represents the GPR predictive function in Eq.(3) expanded into summation notation for use as the optimization objective function. Eq.(8) is the feasibility constraint where the probability of process feasibility, predicted by GPC, must be greater than the given conservativeness parameter. Eq.(9) is the implementation of Eq.(6) for the GPC predictions, expanded into summation notation for use as an optimization constraint. Eq.(10) and Eq.(11) are constraints to calculate intermediate variables within the optimization implementation of GPC.

$$\min \quad \hat{y} = \mu + \sum_{j=1}^{m} \sum_{j'=1}^{m} R_{j,j'}^{-1} \left(y_{j'} - \mu\right) \exp\left(-\sum_{k=1}^{q} l_k^{GPR} \left(x_k^{\text{new}} - x_{j,k}^{GPR}\right)^2\right) \tag{7}$$

$$\text{s.t.} \quad \varepsilon - p \leq 0 \tag{8}$$

$$p = \sigma\left(\frac{\sum_{i=1}^{n} \left(t_i - \sigma\left(\hat{f}_i\right)\right) \alpha_i}{\left(1 + \frac{\pi}{8} v\right)^{\frac{1}{2}}}\right) \tag{9}$$

$$v = 1 - \sum_{i=1}^{n} \alpha_i \sum_{i'=1}^{n} P_{i,i'}^{-1} \alpha_{i'} \tag{10}$$

$$\alpha_i = \exp\left(-\sum_{k=1}^{q} l_k^{GPC} \left(x_k^{\text{new}} - x_{i,k}^{GPC}\right)^2\right) \tag{11}$$

Results

Figure 2 shows the results of a demonstrative application of the presented methodology framework as a proof of concept. Specifically, Figure 2A shows data points sampled using "maximin" LHS from an underlying black-box function, where converged samples are shown by squares, non-converged samples are shown by circles, and the underlying function is shown (for demonstrative purposes) by the dashed line.

Figure 2B shows the trained GPR model exhibits a good fit over the feasible region, with GPR uncertainty increasing with increasing distance from converged training samples (shown by the shaded region). Figure 2B exhibits the GPR characteristic that predictions pass exactly through training samples. Additionally, as the distance from training samples increases, the GPR model predictions regress to the fitted constant term μ.

Figure 2C shows the training labels for GPC and Figure 2D shows the resulting GPC predictions used in the optimization problem feasibility constraints. A GPC prediction greater than 0.5 predicts a feasible design, whereas a GPC output less than 0.5 predicts an infeasible design. In this way, there exists a separating hyperplane where the GPC predictions are equal to 0.5 – the region separated in this way is shown by the color of the solid line in Figure 2B. Tuning the conservativeness parameter enables tightening/relaxation of the feasible region, for example, only allowing designs with a probability of feasibility greater than 0.8 or relaxing the feasible region to include all designs with a probability of feasibility greater than 0.1.

NLP solver IPOPT was used to solve the 3 optimization problems with different conservativeness parameters. For $\varepsilon = 0.5$, the GPR objective function was minimised and the solution $\hat{y}$ = -1.05 (absolute uncertainty s^2 = ±0.087, relative uncertainty ±8.3 %) found at x = 2.52 with predicted probability of feasibility at the p = 0.50 boundary (CPU time 1.45 s). For a stricter ε = 0.8, the solution was a more conservative $\hat{y}$ = -0.37 (absolute uncertainty s^2 = ±0.012, relative uncertainty ±3.2 %) found at x = 1.54 with predicted probability of feasibility at the p = 0.80 boundary (CPU time 3.91 s). Finally, by relaxing the feasible region with ε = 0.1, the solution space included the global optimum $\hat{y}$ = -1.47 (absolute uncertainty s^2 = ±0.75, relative uncertainty ±51 %) which was found at x = 3.74 with predicted probability of feasibility of p = 0.19 (CPU time 1.69 s).

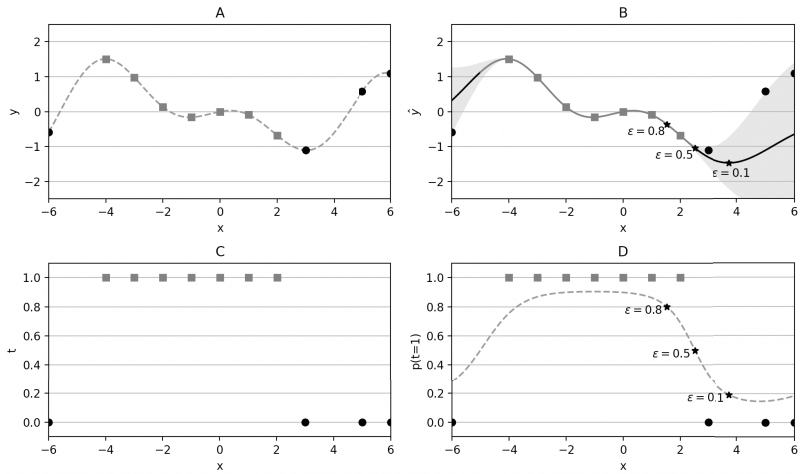


Figure 2. Results of the presented black-box optimization framework. A: samples from black-box simulations where the dashed line is the underlying function, squares show converged simulations, and circles show non-converged simulations. B: 3 optimization solutions superimposed on the GPR predictions (solid line) and uncertainty (shaded region) over the design space for 3 different conservativeness parameter values (where the solid line color shows the feasible region for conservativeness equal to 0.5). C: target values from black-box simulations. D: GPC prediction of probability of feasible designs with the 3 solutions superimposed.

Conclusions

In this work, we presented a GP-based optimization methodology framework embedding GPR for process performance and uncertainty predictions, and GPC for tuneable feasibility constraints for robust design. GPR has previously been shown to be an effective surrogate model for simulation-based optimization (Caballero and Grossmann, 2008), but we couple this, for the first time (to the best of our knowledge), with GPC feasibility constraints. We show that the probabilistic predictions from trained GPC models enable the definition and tuning of a conservativeness parameter depending on process infeasibility risk and process performance tolerances. Ongoing research includes incorporation of the GP uncertainty into a robust optimisation framework, and applications of the developed methodology to simulation-based superstructure optimization problems, formulated as mixed integer non-linear programs, for the optimization of integrative processes to recover resources from wastewater. This work contributes process systems engineering knowledge to other interesting applications of GPs to wastewater treatment research such as (Kocijan and Hvala, 2013; Oyebamiji et al., 2017).

Acknowledgements

The authors would like to acknowledge the UK Engineering and Physical Sciences Research Council (EPSRC) for providing financial support for research under the DTP "Systems modelling design for waste resource recovery" [2194316].

References

Bhosekar, A., Ierapetritou, M., 2018. Advances in surrogate based modeling, feasibility analysis, and optimization: A review. Computers & Chemical Engineering 108, 250–267.

Bynum, M.L., Hackebeil, G.A., Hart, W.E., Laird, C.D., Nicholson, B.L., Siirola, J.D., Watson, J.-P., Woodruff, D.L., 2021. Nonlinear Programming with Pyomo, in: Bynum, M.L., Hackebeil, G.A., Hart, W.E., Laird, C.D., Nicholson, B.L., Siirola, J.D., Watson, J.-P., Woodruff, D.L. (Eds.), Pyomo — Optimization Modeling in Python. Springer International Publishing, Cham, pp. 91–109.

Caballero, J.A., Grossmann, I.E., 2008. An algorithm for the use of surrogate models in modular flowsheet optimization. AIChE Journal 54, 2633–2650.

Cozad, A., Sahinidis, N.V., Miller, D.C., 2014. Learning surrogate models for simulation-based optimization. AIChE Journal 60, 2211–2227.

Douglas, J.M., 1988. Conceptual Design of Chemical Processes, Chemical engineering series. McGraw-Hill.

Ibrahim, D., Jobson, M., Li, J., Guillén-Gosálbez, G., 2018. Optimization-based design of crude oil distillation units using surrogate column models and a support vector machine. Chemical Engineering Research and Design 134, 212–225.

Johnson, M.E., Moore, L.M., Ylvisaker, D., 1990. Minimax and maximin distance designs. Journal of Statistical Planning and Inference 26, 131–148.

Jones, D.R., Schonlau, M., Welch, W.J., 1998. Efficient Global Optimization of Expensive Black-Box Functions. Journal of Global Optimization 13, 455–492.

Kocijan, J., Hvala, N., 2013. Sequencing batch-reactor control using Gaussian-process models. Bioresource Technology 137, 340–348.

McKay, M.D., Beckman, R.J., Conover, W.J., 1979. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics 21, 239–245.

Oyebamiji, O.K., Wilkinson, D.J., Jayathilake, P.G., Curtis, T.P., Rushton, S.P., Li, B., Gupta, P., 2017. Gaussian process emulation of an individual-based model simulation of microbial communities. Journal of Computational Science 22, 69–84.

Rasmussen, C.E., Williams, C.K.I., 2006. Gaussian Processes for Machine Learning. The MIT Press.

Sacks, J., Welch, W.J., Mitchell, T.J., Wynn, H.P., 1989. Design and Analysis of Computer Experiments. Statist. Sci. 4, 409–423.

Sahinidis, N.V., 2004. Optimization under uncertainty: state-of-the-art and opportunities. Computers & Chemical Engineering 28, 971–983.

Wächter, A., Biegler, L.T., 2006. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Mathematical Programming 106, 25–57.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50208-6

Machine Learning-Enabled Optimization of Force Fields for Hydrofluorocarbons

Bridgette J. Befort[a], Ryan S. DeFever[a], Edward J. Maginn[a], Alexander W. Dowling[a*]

[a]*Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, Indiana 46556, United States*
adowling@nd.edu

Abstract

In this work surrogate assisted optimization is utilized to calibrate predictive molecular models, called force fields, used in molecular simulations to reproduce the liquid density of a hydrofluorocarbon refrigerant molecule. A previous calibration workflow which relied on Gaussian process regression models and large Latin hypercube samples to screen force field parameter space is extended to include Bayesian optimization methods to efficiently guide the search for force field parameters. In comparison to the previous work, the Bayesian-based calibration workflow finds a parameter set which results in a lower objective function value than the original workflow after evaluating approximately 50% fewer parameter sets. It is envisioned that this updated workflow will facilitate rapid force field optimization enabling screening of vast molecular design space.

Keywords: Bayesian optimization, Gaussian process regression, Molecular simulation

1. Introduction

Molecular simulation is a powerful tool for studying the thermodynamic and dynamic properties of materials. For example, molecular simulation shows great promise for screening vast molecular design spaces which could be expensive or infeasible to probe experimentally. However, to utilize molecular simulation in this capacity requires accurate predictive molecular models, called force fields. Force fields use a functional form and parameters to describe the potential energy of a system and are utilized in classical molecular simulations to model intra- and intermolecular interactions. Developing generalized, or transferable, force fields to describe large swaths of chemical space has historically been a laborious endeavour, often taking months to years to complete. Though these off-the-shelf force fields offer accurate predictions for some systems, they inevitably lack accuracy across the extraordinary range of molecules found in the natural and synthetic world. Further optimization of force field parameters is often necessary to ensure the model has the required accuracy for the molecules and properties of interest (Wang and Kollman, 2001). Thus, force field optimization represents a bottleneck to applying molecular simulation to new systems.

Emerging computational frameworks promise to greatly accelerate the calibration of highly accurate, physics-based force fields from experimental data. Efforts to calibrate force fields encompass gradient-based, stochastic search, analytical, and *ad hoc* optimization approaches (Befort et al., 2021). Often, a barrier to efficiently calibrating force fields is the expense of calculating the objective function, which quantifies the difference between the simulation prediction and experimental value of a property of

interest. The time requirement of simulations, ranging from minutes-to-hours (*e.g.,* liquid density calculations) to days-to-weeks (*e.g.,* vapor-liquid equilibrium), often makes the objective function calculation cost prohibitive. This cost only increases as more objectives, state points (*e.g.*, temperatures, pressures), and parameters are incorporated in the optimization procedure. Recently, machine learning (ML) methods have been harnessed to address this challenge by mapping microscopic coordinates to a microscopic potential. This functionality enables ML force fields (Unke et al., 2021), but their black box nature often prevents physical insights that can be gained from a physics-based functional form and parameters of a traditional force field.

ML techniques also facilitate surrogate-assisted optimization. Recently, we developed Gaussian process regression (GPR) and support vector machine models to emulate molecular simulations to calibrate force fields (Befort et al., 2021), hereafter referred to as the JCIM workflow. Our surrogate-assisted optimization JCIM workflow successfully screened millions of potential parameter sets, generated through iterative batches of space filling Latin hypercube sampling (LHS). Compared to a force field calibrated via hand-tuning, our workflow enabled the screening of $\mathcal{O}(10^3)$ times more parameter sets while requiring five times fewer simulations. However, this semi-automated workflow required generating large parameter set samples each iteration as well as user input to select which parameter sets to consider in the next iteration (*i.e.*, batch of molecular simulations).

In this work, we explore automating our ML-enabled force field calibration framework by leveraging Bayesian optimization (BO) (Wang and Dowling, 2022) to intelligently propose new parameter sets. Instead of relying on large $\mathcal{O}(10^5 - 10^6)$ LHS batches of parameter sets and user-driven decisions to screen parameter space, BO automatically balances the search for optimal parameter sets between regions which improve the molecular simulation agreement with experimental data (exploitation) and regions which, if sampled, will reduce the uncertainty of the ML model (exploration). As a demonstration case, we optimize force field parameters for a hydrofluorocarbon (HFC) refrigerant molecule, difluoromethane (HFC-32). HFCs are a motivating application because they are subject to recent mandates which require the phaseout of high global warming potential (GWP) materials. Molecular simulation can aid in the sustainable implementation of this phaseout, but this is contingent upon highly accurate force field models for HFCs. Here, we consider optimization of a force field for HFC-32 which is widely used in many refrigerants and will play a significant role in the transition to next-generation refrigerants due to its low GWP relative to other commonly used HFCs.

2. Methods

2.1. Force Field Model

This work calibrates a classical molecular mechanics force field for HFC-32 with the functional form:

$$U(\boldsymbol{r}) = \sum_{\text{bonds}} k_r(r - r_0)^2 + \sum_{\text{angles}} k_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedrals}} \upsilon_n[1 + \cos(\eta\phi - \gamma)] + \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \sum_i \sum_{j>i} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$$

Here, U is the potential energy and $\boldsymbol{r}$ is the vector of position coordinates within the configuration space. The first three terms in this equation represent intramolecular

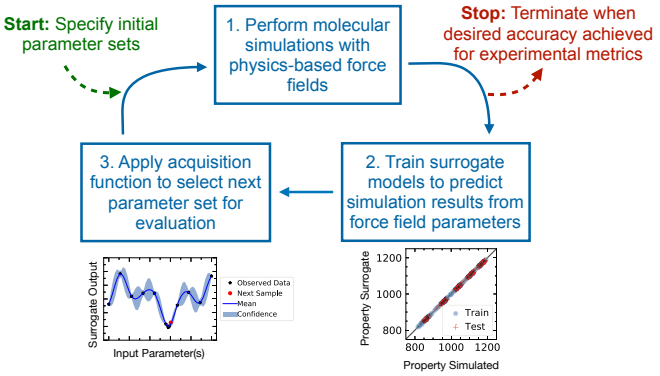


Figure 1: Overview of the proposed BO-enabled force field calibration workflow.

interactions and the fourth term represents Coulombic intermolecular interactions. The parameters in these terms are not calibrated and are reported in Befort et al. (2021). The final term of this force field functional form contains σ_{ij} and ε_{ij} which parameterize the Lennard-Jones potential describing the van der Waals repulsion-dispersion intermolecular interactions between atoms i and j. Here, we focus on rapidly generating an accurate force field, *i.e.*, improving $U(\boldsymbol{r})$, for HFC-32, by calibrating the like-interaction Lennard-Jones parameters, σ_{ii} and ε_{ii}, to reproduce experimental HFC-32 liquid density. These parameters are calibrated for the three atom types (C, F, and H) of HFC-32, resulting in six total fitting parameters. A description of the model, parameters, and general system setup can be found in Befort et al. (2021). System changes in this work include: first, performing simulations using the LAMMPS molecular dynamics package and, second, expanding the parameter bounds as follows (σ in Å, ε in kcal/mol): $3.0 \leq \sigma_C \leq 4.0$, $2.5 \leq \sigma_F \leq 3.5$, $1.7 \leq \sigma_H \leq 2.7$, $0.06 \leq \varepsilon_C \leq 0.16$, $0.04 \leq \varepsilon_F \leq 0.14$, $0.0 \leq \varepsilon_H \leq 0.05$. While this paper only focuses on optimizing the intermolecular Lennard-Jones parameters, which are the least accurate when force fields are traditionally parameterized using quantum calculations, we emphasize the proposed BO calibration workflow is applicable to any parameters in the force field.

2.2. Bayesian Optimization Workflow

The goal of this work is to refine $U(\boldsymbol{r})$ by optimizing force field parameters, ζ, such that the objective function, $f(\zeta)$, is minimized. For this case study, $\zeta = (\sigma_C, \sigma_F, \sigma_H, \varepsilon_C, \varepsilon_F, \varepsilon_H)$. Figure 1 shows the BO-enabled force field calibration workflow. First, ten initial parameter sets are generated via LHS. In step one, molecular simulations compute the liquid density, $\mathbf{y}^{sim}$, of HFC-32 from $U(\boldsymbol{r})$ at multiple state points of interest. Depending upon the quality of $U(\boldsymbol{r})$, $\mathbf{y}^{\mathrm{sim}}$ may or may not be close to the experimental values, $\mathbf{y}^{exp}$, and this discrepancy is quantified as the mean squared error objective function, $f(\zeta) = \sum_{i=1}^{n}\left(y_i^{sim}(\zeta) - y_i^{exp}\right)^2$, where n is the number of state points considered. After the initial molecular simulations, for each parameter set, $\mathbf{y}^{sim}$ and the subsequent objective function $f(\zeta)$ are computed from the simulation output. At some state points the parameters used in the simulation are so poor that the vapor, not liquid, density of HFC-32 is the simulation result, leading to a discontinuity in objective function value that may be difficult for a surrogate model to capture. To prevent this discontinuity, if a simulation outputs a density lower than the critical density of HFC-32, the simulation density is reported as the critical density. This formulation results in a sufficiently poor objective function value, indicating

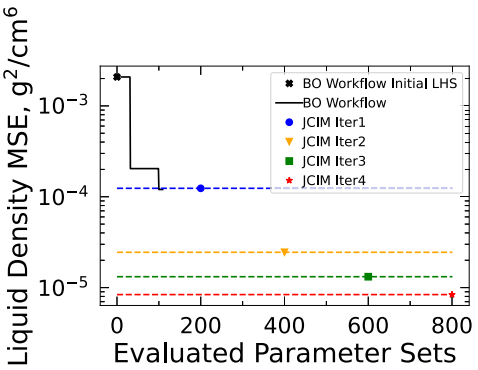


Figure 2: Comparison of the best objective function value after evaluating a certain number of parameter sets using the proposed BO workflow (black **x**, black line) versus our prior JCIM workflow (colored markers, dashed lines).

a poor parameter set, while preventing a discontinuity. In step two, a GPR model is trained to predict $f(\zeta)$ as a function of calibrated parameters ζ (Befort et al., 2021). In step three a BO acquisition function is optimized to determine the next parameter set to evaluate. Here, the expected improvement (EI) acquisition function is used to select the next optimal parameter set for simulation (Wang and Dowling, 2022). Upon applying the EI acquisition function, a new parameter set is generated which is used in new simulations, and the workflow continues iteratively until the desired simulation accuracy is reached or no improvement can be achieved in the objective function.

3. Results

We begin by comparing our automated BO workflow to our prior semi-automated JCIM workflow. Figure 2 plots the best (lowest) objective function value found after simulating 110 trial parameter sets (*i.e.,* initial ten LHS parameter sets plus one hundred EI-generated samples) in the BO workflow compared to the best objective function values found after evaluating 200, 400, 600, and 800 total parameter sets in the JCIM workflow. Figure 2 shows the improvement in objective function value for both workflows as more parameter sets are evaluated and the surrogate models are trained on more data. After 101 parameter sets are evaluated in the BO-based workflow, the objective function is 1.20 x 10^{-4} g^2/cm^6 while the lowest objective function for the JCIM workflow's initial 200 parameter sets, which were generated via LHS, was 1.24 x 10^{-4} g^2/cm^6. This indicates that the BO workflow can achieve a lower objective function value after evaluating approximately 50% fewer parameter sets, and therefore performing less simulations, than a space filling sample of parameter space. We hypothesize this is the result of the adaptive nature of BO acquisition function, which can effectively explore and exploit parameter space to more efficiently find optimal parameter sets. We expect that as more simulations are performed, the GPR models in the BO workflow will improve such that this workflow will additionally require fewer simulations to surpass the objective function values found after evaluating 400, 600, and 800 samples in the JCIM workflow.

GPR model improvement for the BO workflow is seen in Figure 3, which shows the absolute error between the GPR model prediction of the objective and the actual simulation result for each trial parameter set. The standard deviation in the GPR model prediction is plotted and shows a decreasing trend. The discrepancy between the GPR

model prediction and actual simulation result also shows a decreasing, although less obvious, trend. This indicates GPR model improvement as more training data and regions of parameter space are sampled via the guidance of BO. Figure 4 compares the GPR model predictions and the actual simulation results of the objective for each evaluated parameter set, with the GPR model uncertainty plotted as error bars. This figure provides an example of how the BO-based workflow balances exploration and exploitation, showing how GPR model predictions for certain trial parameter sets result in objectives very close to simulation results (exploitation) while other predictions are significantly different than simulations and have high uncertainty, indicating exploration. Thus, instead of relying on GPR models to generate batches of new trial parameter sets in the original workflow, the BO-based workflow systematically samples parameter sets to gain information about the parameter regions which result in the lowest objective function values. We believe this approach both reduces the burden of the user by automatically selecting new parameter sets to sample and improves the efficiency of the workflow by decreasing the number of parameter set samples and subsequent simulations required to calibrate force fields.

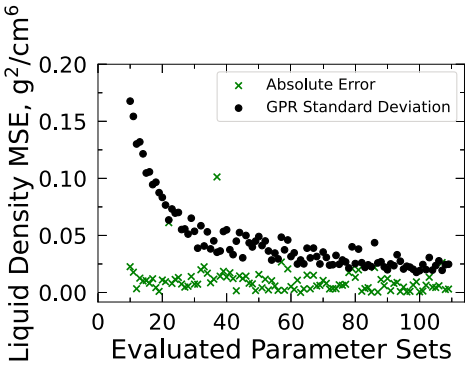


Figure 3: Absolute error between GPR model and simulation (green **x**) and GPR prediction uncertainty (black •) change as more parameter sets are evaluated.

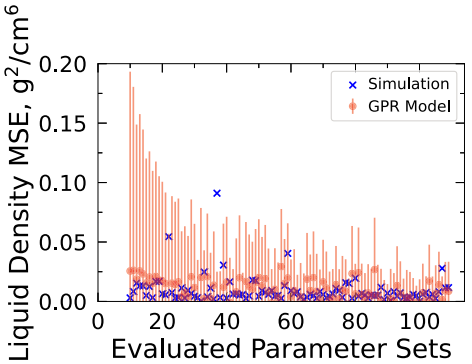


Figure 4: Comparison of GPR model prediction mean (red •) and standard deviation (red | , error bar) versus molecular simulation results (blue **x**) as more parameter sets are evaluated.

Further analysis is required to benchmark the BO-based calibration workflow. The BO results shown in Figure 2 were generated using the gradient based L-BFGS (scipy) optimizer to calibrate the GPR model hyperparameters for the first 80 iterations. Then, the hyperparameter optimization failed due to a Cholesky factorization error (the GP kernel became negative semi-definite) and was switched to an ADAM (BOTorch) optimizer. Reproducing this result using only the BOTorch optimizer is a work in progress; with only the BOTorch optimizer, the GPR model error remains $\mathcal{O}(10^{-2})$ whereas the JCIM workflow was $\mathcal{O}(10^{-4})$ after 200 samples. Ongoing work is investigating the differences in trained GP hyperparameters and overall BO performance using these two optimizers. Additional opportunities for further improving this framework include determining the minimum amount of data and initial parameter set samples necessary for efficient GPR model improvement, exploring various formulations for the objective function, kernel function, and the vapor-liquid density discontinuity, and evaluating the capabilities of various GPR model optimizers. We expect these analyses to improve the overall efficiency of the BO calibration workflow.

4. Conclusions

In this paper, we prototype a fully automated BO framework for force field calibration. Results show that after evaluating 101 parameter set samples with the BO workflow, the lowest mean squared error between simulation and experimental values for the liquid density of HFC-32 is 1.20 x 10^{-4} g^2/cm^6. This objective was 0.04 x 10^{-4} g^2/cm^6 smaller than the best objective found in the initial 200 parameter set LHS used in our prior JCIM workflow. This result suggests the BO techniques enhance the efficiency of force field calibration. Additionally, BO has enabled automated sampling of parameter space removing the need for user decisions for generating trial parameter sets. We expect that as more parameter sets are sampled, improvement within the GPR models will continue to show that fewer simulations will yield equally accurate force fields as the original workflow. Ultimately, this framework can be used to develop accurate force fields for multiple HFCs and other classes of molecules for which accurate molecular models are lacking.

5. Acknowledgements

We acknowledge support from the National Science Foundation grant CBET-1917474 and the University of Notre Dame. BB acknowledges support from the Richard and Peggy Notebaert Premier Fellowship.

6. References

Befort, B. J., DeFever, R. S., Tow, G. M., Dowling, A. W., & Maginn, E. J. (2021). Machine Learning Directed Optimization of Classical Molecular Modeling Force Fields, *J. Chem. Inf. Model*, 61(9), 440-4414.

Unke, O. T., Chmiela, S., Sauceda, H. E., Gastegger, M., Poltavsky, I., Schütt, K. T., Tkatchenko, A., & Müller, K. R. (2021). Machine learning force fields. *Chemical Reviews*.

Wang, J., & Kollman, P. A. (2001). Automatic parameterization of force field by systematic search and genetic algorithms. *J. Comp. Chem.*, *22*(12), 1219-1228.

Wang, K., & Dowling, A. W. (2022). Bayesian optimization for chemical products and functional materials. *Current Opinion in Chemical Engineering*, *36*, 100728.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50209-8

Design of an Event-Driven Rescheduling Algorithm via Surrogate-based Optimization

Teemu J. Ikonen[a,*], Keijo Heljanko[b,c], Iiro Harjunkoski[a,d]

[a]*Aalto University, Department of Chemical and Metallurgical Engineering, PO Box 16100, 00076 Aalto, Finland*
[b]*University of Helsinki, Department of Computer Science, PO Box 68, 00014 University of Helsinki, Finland*
[c]*Helsinki Institute for Information Technology (HIIT), Helsinki, Finland*
[d]*Hitachi Energy Research, Kallstadter Strasse 1, 68309 Mannheim, Germany*
teemu.ikonen@aalto.fi

Abstract

In event-driven rescheduling, new re-optimization procedures are triggered when obtaining new information that indicates the current schedule to be outdated. Critical design aspects of such an algorithm are the definition of the trigger event and the allocated computing time for a new rescheduling procedure. We treat both of these design aspects as continuous control parameters. Nevertheless, finding the best-suited control parameter combination for a given operating environment may be computationally expensive, as it requires simulating the process with many candidate combinations. We use surrogate-based optimization to reduce the computing cost of optimizing the control parameters. We demonstrate the method on real-time rebalancing of a bike sharing system and investigate the sensitivity of the optimized parameters to changes in the operating environment.

Keywords: optimization, event-driven rescheduling, surrogate modelling, logistics, bike sharing rebalancing.

1. Introduction

Online process scheduling is real-time decision-making of operational decision, such as the set of tasks to be executed, their timing and resource allocation. As processes are inherently stochastic, reaction or anticipation is required to deviations (e.g., in batch durations and material yields), disturbances and new orders. Gupta et al. (2016) review the literature of online scheduling and propose a framework for the classification of the online scheduling design aspects. One of the design aspects is the *re-computation trigger*, which is also referred to as the *when-to-schedule decision* (Sabuncuoglu and Kizilisik, 2003). Commonly used methods to trigger a re-computation procedure are *periodic* and *event-driven* (also *event-triggered*) *rescheduling*, and their hybrids.

In this work, our focus is on event-driven rescheduling. In the literature, Touretzky et al. (2017) propose a framework where process level deviations and disturbances trigger re-computing procedures at the scheduling level. Katragjini et al. (2013) investigate heuristic schedule repair algorithms on flow shop scheduling problems. New rescheduling procedures are triggered if new events (e.g., a machine disruption, a new order, or a deviation in a task processing time) affect the current schedule. Gupta et al. (2016) demonstrate a case where a rescheduling procedure is needed even if a trigger

event is not present. Pattison et al. (2017) propose a closed-loop production rescheduling framework combining the two categories of re-optimization triggers. Periodic rescheduling is conducted when receiving updates of price and demand forecasts, whereas event-driven rescheduling is conducted in the case of a market or process disturbance.

The design of an event-driven rescheduling algorithm requires a definition of *an event* that triggers a re-optimization procedure. One approach is to define a quantitative metric for the importance of the event, such that when this metric exceeds a given threshold a re-optimization is triggered. Typically, event-driven rescheduling involves also other continuous control parameters, such as the allocated computing time and the used prediction horizon length. The evaluation of the closed-loop performance of just one control parameter combination requires simulation of the process and repeatedly performed optimization procedures. Thus, optimization of these control parameters may require a long computing time, as many candidate combinations need to be evaluated.

In our earlier work, we propose surrogate-based optimization to be used to reduce the computational cost of optimizing (continuous) control parameters of a periodic rescheduling algorithm (Ikonen et al., 2021). In this paper, we extend the work to the design of an event-driven rescheduling algorithm. We investigate three continuous control parameters, including the threshold of event importance and the allocated computing time. Further, we investigate how sensitive the optimized parameters are to small changes in the operating environment.

2. Bike sharing rebalancing

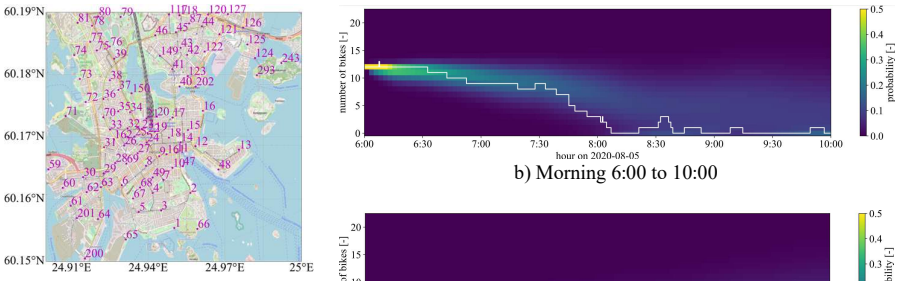
Urban bike sharing systems are expanding in cities around the world. The users of the system can pickup a bike for a short-term use from (typically fixed location) stations and return them to a station in the same system after cycling. As the pickup and return rates at different stations are asymmetric, the operators of the system need to rebalance the distribution of bikes by, e.g., trucks.

We investigate event-driven rescheduling on the dynamic bike sharing rebalancing problem (BRP), in which the rebalancing actions are performed during the day when the system is in use. We have chosen the application because in it new information of the state of the system is obtained frequently and it (like typical online production scheduling) involves solving mixed-integer programming (MIP) problems within a limited time. Last but not least, real process data of the bike sharing systems is publicly available.

We use the framework by Schuijbroek et al. (2017) to optimize the rebalancing actions. The framework consists of 1) a prediction model for the bike levels at different stations, 2) a MIP model to decompose the routing problem into smaller problems, each of which includes only one vehicle, and 3) an MIP model for optimization of the rebalancing actions of a vehicle. In this work, we formulate the dynamic BRP for only a single vehicle. Thus, we use the probabilistic prediction model and the MIP model for rebalancing operations from the framework by Schuijbroek et al. (2017).

The prediction model for the bike inventory at station $i \in S$ is based on the $M_t/M_t/1/K$ queuing system, where the arrivals of new 'customers' (here: bikes) and their service are both non-stationary Markovian processes. Both events are assumed to be exponentially distributed with time-dependent rates, which are determined based on

historical data. One machine is serving the 'customers' and there is a total of K waiting spaces (here: the capacity C_i of station i). The transient probabilities for the bike levels can be solved by the Kolmogorov forward equations and the fourth order Runge-Kutta method. Figure 1 shows the predictions of the bike levels at a station, located in a residential area, in the Helsinki bike sharing system in the morning and afternoon of August 5, 2020. The predictions are based on the data recorded in June 2020. The white path shows the actual realization of the bike level on August 5, 2020.



© OpenStreetMap contributors

a) 92 stations lying within coordinate ranges [60.15…60.19] °N and [24.90…25.00] °E.

b) Morning 6:00 to 10:00

c) Afternoon 14:00 to 18:00

Figure 1: The selected stations from the Helsinki bike sharing system, Finland, (Subfigure **a**) and four-hour probabilistic predictions (**b** and **c**) made at 6:00 and 14:00 on August 5, 2020 at the Toinen Linja bike sharing station (number 149 in Subfigure **a**) (Helsinki Region Transport, 2016) (Kainu, 2017). The white paths show the actual bike levels.

The MIP model for the rebalancing actions minimizes the rebalancing time to bring the system to a state where all stations fulfil a given *service level requirement* (SLR) β for both pickup and return demand during a predefined prediction horizon T. The service level is the fraction of satisfied pickup/return demand. It can be determined based on the probabilistic prediction for the station levels. For detailed descriptions of the prediction and MIP model, the reader may consult the papers by Schuijbroek et al. (2017) and Ikonen et al. (2021).

We trigger a new rebalancing procedure when the service level of a station, having no scheduled visit, drops below a certain threshold β_{trig} (abbreviated as SL trigger). This value, as well as the SLR β and the allocated computing time t_c for solving the MIP model, are the continuous control parameters of our event-driven rescheduling algorithm. The objective is to minimize the total time the stations in the system, S, are either empty of full, $\Sigma_{i \in S} t_i^{\text{ef}}$, where t_i^{ef} is the time station i is empty/full. The next section introduces surrogate-based optimization, which we use to optimize the control parameters.

3. Surrogate-based optimization

Surrogate-based optimization is a sample efficient search method, suitable for black box type objective functions that are expensive to evaluate. Figure 2 shows a generic illustration of surrogate-based optimization. First, after identifying important decision variables, the initial sampling plan is generated using a design of experiment method. Second, each candidate solution is evaluated and used to construct a surrogate of the

objective function in the search space. Third, an iterative process is started where new candidate solutions (also referred to as *infill points*) are identified based on an infill criterion and evaluated. The obtained data is used to update the surrogate. The iterative process is continued until a predefined termination criterion. We refer the reader to the paper by Forrester and Keane (2009) for a review of surrogate-based optimization and to that by McBride and Sundmacher (2019) for review of surrogate modelling in chemical process engineering.

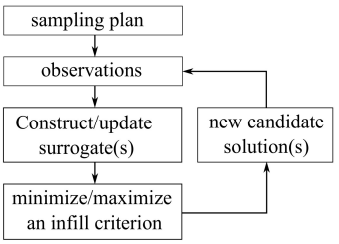


Figure 2: An illustration of surrogate-based optimization.

In this work, we use the efficient global optimization (EGO) algorithm by Jones et al. (1998). We generate the initial sampling plan with n_0 points using Latin hypercube sampling. We then identify the first n_{ei} infill points by maximizing the expected improvement and the following $n_{\hat{\mathrm{y}}}$ points by seeking the point where the predicted value from the surrogate is the smallest. We use the implementation of the EGO method by Bouhlel et al. (2019).

4. Results

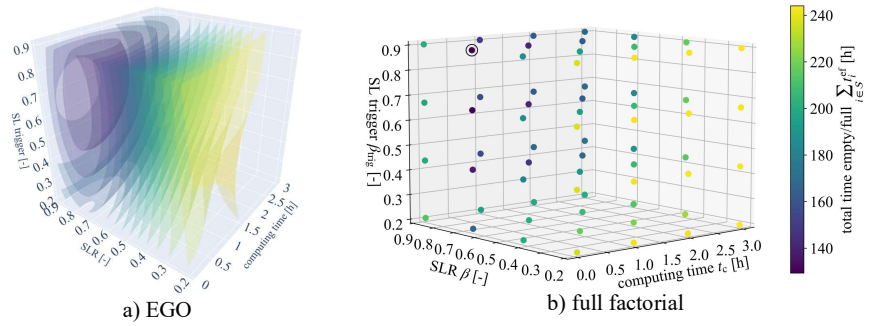
We evaluate the proposed method on optimization problems formulated based on real process data recorded in Helsinki bike sharing system (Helsinki Regional Transport, 2016, Kainu, 2017). We consider a subset of the stations (92), the coordinates of which lay within [60.15…60.19] °N and [24.90…25.00] °E (see Figure 1a), the rebalancing of which is to be conducted by one vehicle. The predictions of station levels are based on data recorded in June 2020. When optimizing the control parameters, we evaluate each control parameter combination on 9 to 10 weekdays in August-September 2020. The objective function value is the average of the obtained results. Each day, the rebalancing is conducted from 6:00 to 15:00. We evaluate the objective $\Sigma_{i \in S} t_i^{\mathrm{ef}}$ from 6:00 to 22:00.

In order to investigate also the sensitivity of the results, we consider three different time windows (TW1: August 3 - 14 (data from August 11 is missing), TW2: August 17 - 28, TW3: August 31 - September 11). Weekends are excluded. TW2 is comparable to the case studied by periodic rescheduling in Section 4.1 of Ikonen et al. (2021). We use the prediction horizon of $T = 1.738$ h, which was the optimized value with the periodic rescheduling algorithm. Further details of the test case are given in Ikonen et al. (2021).

The search space is defined by the following bounds: SL trigger $\beta_{\mathrm{trig}} \in [0.2 \ldots 0.9]$, SLR $\beta \in [0.2 \ldots 0.9]$, computing time $t_{\mathrm{c}} \in [0.05 \ldots 3.0]$ h. We allocate a total of 32 evaluations for each optimization procedure by the EGO method ($n_0 = 15$, $n_{\hat{\mathrm{y}}} = 15$, $n_{\hat{\mathrm{y}}} = 2$). We use full factorial sampling as a reference method, for which we allocate 64 evaluations (a $4 \times 4 \times 4$ grid). The evaluations were performed on a high-performance computing facility, such that each run was allocated one thread of an Intel Xeon E5 2680/2690 v3 node with 4 GB of memory. We solve the MIP models using Gurobi 9.1.2.

Figure 3 visualizes the obtained results by the EGO method and full factorial sampling for TW2. Table 1 shows the numerical results for all three time windows. On the studied time windows, EGO method yields 1.2 to 5% smaller objective function values than full factorial sampling. In the optimized parameter combinations, obtained by the

EGO method, the SLR β has the smallest variation (all values are within 1.3% from each other). The optimized SL trigger β_{trig} and allocated computing time t_c have more variation depending on the time window. Figure 4 shows the realized and planned route of the vehicle at the first two time points it receives operations updates on August 17, 2020 (TW2). The used parameters are those obtained by the EGO method.

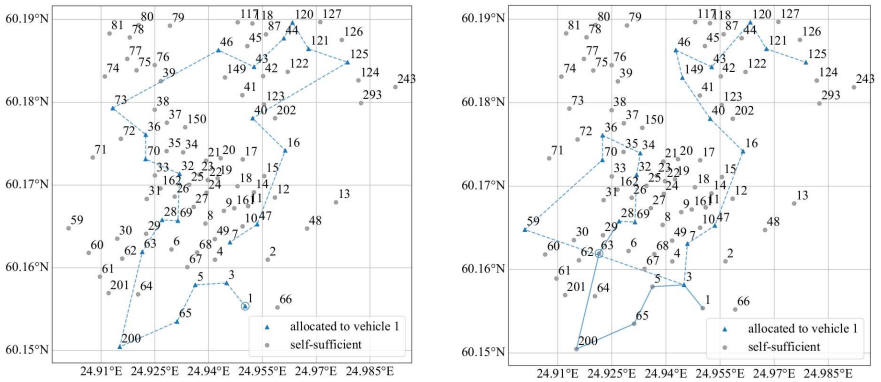


a) EGO b) full factorial

Figure 3: Optimization results by the EGO method and the full factorial sampling for TW2. Subfigure (**a**) shows isosurfaces obtained from a surrogate that is trained with all observed data. In Subfigure (**b**), the best parameter combination is circled.

Table 1: The optimized control parameters (i.e., service level trigger β_{trig}, service level requirement β, computing time t_c) for the three time windows by EGO the method and full factorial sampling.

time window	days	method	evals.	objective $\Sigma_{i \in S} t_i^{\text{ef}}$ [h]	parameters β_{trig} [-]	β [-]	t_c [h]
		EGO	32	120.62	0.701	0.730	0.050
TW1	9	full factorial	64	125.54	0.667	0.667	0.050
		EGO	32	127.75	0.722	0.732	0.420
TW2	10	full factorial	64	129.33	0.900	0.667	0.050
		EGO	32	95.53	0.810	0.739	0.370
TW3	10	full factorial	64	100.63	0.667	0.667	0.050



a) 6:00:00 b) 6:25:12

Figure 4: Realized (solid line) and planned route (dashed line) of the vehicle when obtaining the first two operations updates on August 17, 2020.

5. Conclusions

In this work, we investigated surrogate-based optimization of three parameters of an event-driven rescheduling algorithm, using the EGO method. On the three tested time windows, the EGO method yielded 1.2 to 5% smaller objective function values than full factorial sampling, when allocated 50% fewer function evaluations. The optimized service level requirement β has only small variation between the tested time windows, whereas the other two optimized parameters have larger variation.

Acknowledgement: Financial support from the Academy of Finland is gratefully acknowledged (project RELOOP, decision number 330388). The calculations presented in this work were performed using computer resources within the Aalto University, School of Science, "Science-IT" project.

References

M. A. Bouhlel, J. T. Hwang, N. Bartoli, R. Lafage, J. Morlier, & J. R. Martins, 2019, A Python surrogate modeling framework with derivatives, *Advances in Engineering Software*, 135, 102662.

A. I. Forrester, & A. J. Keane, 2009, Recent advances in surrogate-based optimization. *Progress in aerospace sciences*, 45(1-3), 50-79.

D. Gupta, C. T. Maravelias, and J. M. Wassick, 2016, From rescheduling to online scheduling, *Chemical Engineering Research and Design*, 116:83–97.

Helsinki Region Transport, 2016, City bike stations' Origin-Destination (OD) and location data [accessed on November 3, 2020], https://www.avoindata.fi/data/en_GB/organization/helsingin-seudun-liikenne.

T. J. Ikonen, K. Heljanko, I. Harjunkoski, 2021, Surrogate-based optimization of a periodic rescheduling algorithm, *AIChE Journal* (submitted).

D. R. Jones, M. Schonlau, and W. J. Welch, 1998, Efficient global optimization of expensive black-box functions, *Journal of Global Optimization*, 13(4):455–492.

M. Kainu, 2017, City bikes of Helsinki: Open data of bike levels at stations during seasons 2017, 2018, 2019, 2020 ja 2021 (translated from Finnish) [accessed on February 22, 2021], https://data.markuskainu.fi/opendata/kaupunkipyorat/.

K. Katragjini, E. Vallada, and R. Ruiz, 2013, Flow shop rescheduling under different types of disruption, *International Journal of Production Research*, 51(3):780–797.

K. McBride and K. Sundmacher, 2019, Overview of surrogate modeling in chemical process engineering, *Chemie Ingenieur Technik*, 91(3):228–239.

R. C. Pattison, C. R. Touretzky, I. Harjunkoski, and M. Baldea, 2017, Moving horizon closed-loop production scheduling using dynamic process models, *AIChE Journal*, 63(2):639–651.

I. Sabuncuoglu and O. B. Kizilisik, 2003, Reactive scheduling in a dynamic and stochastic FMS environment, *International Journal of Production Research*, 41(17):4211–4231.

J. Schuijbroek, R. C. Hampshire, and W.-J. Van Hoeve, 2017, Inventory rebalancing and vehicle routing inbike sharing systems, *European Journal of Operational Research*, 257(3):992–1004.

C. R. Touretzky, I. Harjunkoski, and M. Baldea, 2017, Dynamic models and fault diagnosis-based triggers for closed-loop scheduling, *AIChE Journal*, 63(6):1959–1973.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50210-4

A two-stage network optimization for sustainable treated wastewater planning

Fatima-zahra Lahlou[a], Sarah Namany[a], Hamish Mackey[a], Tareq Al-Ansari[a,b]

[a]*Divison of Sustainable Development, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar*
[b]*Divison of Engineering Management and Decisison Sciences, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar*
**talansari@hbku.edu.qa*

Abstract

The current global availability of water resources is able to sustain human activities, however, its uneven distribution causes scarcities in some regions of the world, leading to the hindrance of industrial and agricultural operations. In addition to that, water undergoes multiple volatilities and dynamics owing to its dependence on precipitations and weather conditions which are now accentuated due to climate change. To overcome water shortage, water scarce countries have deployed tremendous efforts in enhancing their water sectors by developing novel technologies substituting the renewable water base. Wastewater reuse represents a sustainable opportunity to alleviate water scarcity in arid regions by providing additional reserves. Nevertheless, public perception is the major obstacle to implement such practices due to social constraints. This study presents a two-stage methodology to evaluate the socio-economic and environmental aspects of reusing treated wastewater (TW) generated from different municipal wastewater treatment plants in an arid climate country. The first step consists of assessing the social acceptance of TW use in different applications to preselect the highly socially welcomed ones. Selected uses are then assessed economically and environmentally using a multi-objective network optimization model that aims to reduce the cost associated with the treatment process and the transportation of TW, and to minimize the engendered environmental burden. The TW allocation is constrained by the social acceptance, the capacity of treatment and production along with the quality of the TW. Firefighting, athletic fields irrigation and street cleaning were the uses that obtained the highest social acceptance amongst the studied sample. As for the environmental and economic aspects, the network involving a high contribution of treated municipal wastewater represents the optimum plan.

Keywords: Treated wastewater, social acceptance, multi-objective network optimization, sustainability.

1. Introduction

1.1 Background

Water security is the ability to maintain a continuous access to sufficient and safe water resources to preserve ecosystems and humans' wellbeing for a guaranteed social development and economic prosperity. While water is the backbone of all the vital industrial and economic activities and processes such as agriculture and power generation, it represents a threatened commodity due to several factors. In fact, the ever-increasing population, along with its associated growing demand for water resources, induce tremendous pressure on the water base leading to scarcities in many regions, particularly the arid ones. Climate change is also another issue that further intensifies

water scarcity leading to water insecurity. It is, indeed, the direct cause of volatile rainfall patterns, droughts, floods, and many sudden environmental events that reduce the replenishment of renewable water sources. As a response to the multiple risks governing water security, alternative water sources have been introduced to ensure the satisfaction of the current and future water demands. Desalination is one of the exhaustively deployed technologies to provide water in water-scarce regions. For instance, the Golf Cooperation Council Countries (GCC) which suffers from very limited water reserves, are principally adopting desalination to cover most of their water needs. Although this technique is efficient in alleviating water insecurity, it represents an energy intensive option that engenders significant environmental impact. As part of the efforts to shift to more sustainable water provision technologies, wastewater reuse represents a promising substitute to reduce the environmental and economic costs of desalination methods. In fact, producing water from a treatment facility can cost nine times less that providing it from a thermal desalination plant. However, social perception is still a hinder hampering the deployment of this technology in all domains, counting agriculture and drinking, and the usage is only restricted to activities not including direct human consumption or involvement such as landscaping and firefighting.

1.2 Literature review

The investigation of the potential of utilising TW as an alternate water source was significantly investigated in literature. Studies have addressed the topic from different perspectives representing the three sustainability pillars, counting economic, social and environmental aspects (Lahlou et al., 2021). Considering the economic viewpoint, works have mainly shed the light on the feasibility analysis and the economic benefits of wastewater use in different domains. In this regard, Lahlou et al. (2020a; 2020b), with a focus on agricultural activities, have examined the potential of utilising TW for irrigation purposes such that the nutritional intake of the crops and their water requirement are both satisfied simultaneously. Minhas et al. (2015) have also proven the efficiency of using TW in minimising fertilisers in agriculture through supplying sufficient amounts of Nitrogen and Phosphorus(Minhas et al., 2015). Wang et al. (2020) suggested two-stage model based on a genetic algorithm that optimises the efficiency of a system integrating wastewater reuse and power generation. Adopting an environmental perspective, Canaj et al. (2021) conducted a Life Cycle Analysis (LCA) to assess the environmental viability of the adoption of TW for irrigation purposes. Similarly, Almanaseer et al. (2020) investigated the impact of the wastewater treatment and reuse on water basins by quantifying the microbial biological of the surrounding environment of the plant. As for the social aspect, the focus of TW studies was mainly to assess the social perception and acceptance of adopting TW in certain applications. For instance, Lahlou et al. (2021) examined through a survey analysis, the social perception on TW usage generated from oil and gas industries, in addition to the impact of newly-identified factors on the reuse acceptance. All these studies have looked at the TW utilisation from a single perspective or considering the combination of two sustainability pillars, however, in order to holistically assess the sustainability performance of wastewater reuse, there is a need to aggregate the three aspects in one model. In this paper, a two-stage approach is adopted to determine the optimal allocation of wastewater from different sources into diverse sectors considering economic and environmental constraints and relying on the social perception.

2. Methodology

2.1 Problem Formulation

The two-stage methodology consists first of conducting a survey to assess the acceptance level of reusing treated sewage effluent generated from municipal WWTPs. The results from the social study can then feed into the second stage of the methodology which is based on a multi-objective optimization model. The first objective of the model is represented by equation 1. It aims at reducing the cost associated with the transportation of the TW from the different sources to the different sinks which were identified during the initial stage. The second objective is the minimization of the global warming potential which is engendered by moving the water. If one of the applications requires N fertilization, the amount of carbon footprint offset using the TW for fertigation is taken into consideration in the carbon footprint calculation. The second objective is formulated in equation 2. Finally, the last objective, equation 3, is to maximize the total nitrogen uptake by the sinks that require this resource in order to reduce the material and environmental costs associated with the energy intensive commercial fertilizers.

$$Cost = \sum_{i=1}^{I}\sum_{j=1}^{J} x_{ij}\, d_{ij}\partial \quad (1)$$

$$CF = \sum_{i=1}^{I}\sum_{j=1}^{J} x_{ij} d_{ij}\zeta E - \sum_{i=1}^{I}\sum_{z=1}^{Z} x_{ij} n_i N_\zeta \quad (2)$$

$$N_T = \sum_{i=1}^{I}\sum_{z=1}^{Z} x_{ij} n_{ij} \quad (3)$$

Where:

- x_{ij} is the decision variable such that i and j are the indices for the source and the sink, respectively;
- R_j is the total water requirement for sink j [m^3];
- I and J represent the total number of sources and sinks, respectively;
- d_{ij} represents the distance from course i to sink j [km];
- ∂ is the average price to transport 1 m^3 of water across 1 km using a pipeline [$USD.m^{-3}km^{-1}$];
- E is the energy required to transport 1 m^3 of water across 1 km using a pipeline [$kWh.m^{-3}km^{-1}$];
- ζ represents the carbon footprint associated with producing 1kWh of energy [$CO_{2\text{-}eq}.kWh^{-1}$];
- n_i represents the N concentration in source N [kg/m3];
- N_ζ represents the carbon footprint associated with the production packaging and transportation of 1kg of N fertilizer [$CO_{2\text{-}eq}.kg_N$];
- z represents the indices of the sinks which require N fertilization.

Subject to the following set of constraints:

$$\sum_j^J x_{ij} R_j \leq \xi_i\,,\ \forall i \in (1, I)$$

$$\sum_j^J x_{ij} = 1\,,\ \forall i \in (1, I)$$

Where ξ_i is the total capacity of the source i.

2.2 Case study

The State of Qatar is one of the most water scarce countries in the world, thereafter, effort is required to develop water resources management while respecting the triple bottom line of sustainable development. In the context of Qatar, developing such water planning

framework involving the reuse of alternative water resources as a function of social acceptance represents an opportunity for water stress alleviation at reduced environmental and economic costs.

The public acceptance for TW reuse was assessed using a survey which initial purpose was to investigate newly identified factors, mainly maternal altruism, and their association with the public acceptance of reusing treated industrial wastewater and bio-solids generated from oil and gas industries (F.Z. Lahlou et al., 2021). The findings of part of the survey which focused on the willingness to use TW from municipal WWTPs will feed into this paper. In this study, Doha and Um-Salal municipalities are chosen which are two of the most populated ones in the State of Qatar. There are three different WWTP in the chosen municipalities. While two of them perform tertiary treatment with N and P removal, the third WWTP performs tertiary treatment without N and P removal. Hence, the N concentration is higher in the latter.

Table 1: Water requirements of different wastewater applications (sinks) and their distances from WWTPs (sources). [1]Price, 2019; [2]Bauer et al., 2020; [3]Clark Tanks, n.d.; [4]Shublaq and Sleiti, 2020

Sinks	**Water Requirements**	**Distances from**		
		Doha North WWTP	Lusail WWTP	Doha south WWTP
	($m^3.year^{-1}$)	(km)		
Athletic fields irrigation	1,664[1]			
Al Tamama Stadium		43.2	27.9	4.5
Ras Abu Abud Stadium		45.3	28.8	19.3
Street Cleaning	927,465[2]			
Ministry of Municipality and Environment		37.9	16.9	22.5
Firefighting	290,400[3]			
Fire Station		36.3	16.9	23.2
District Cooling	8,486,200[4]			
Plant P1		36.3	18.7	25.9
Plant P2		40.2	18.4	21.4
Plant P3		36.5	20.4	19.7
Plant Pearl		37.4	15.6	33.3

3 Results

The survey was answered by 1040 respondents which socio-demographic characteristics are significantly representative of the population of the state of Qatar in terms of gender and age group. The applications which scored the highest public acceptance for TW are athletic fields irrigation, street cleaning, firefighting, and district cooling as at least 70% declared to be willing to use TMW for these applications.

Water requirements of the different sinks considered and their distances from the three various sources of WWTP are listed in Table 1. The sinks that require the highest amount of water are the district cooling plants of which each uses over 8 million m^3 of water per year.

The results of the developed multi-objective optimization model are illustrated in Figure 1. The optimum solution comes at low economic and environmental cost as opposed to the current scenario which involves the use of energy intensive desalinated water in some of the applications. The optimum solution allocates TMW to the eight different sinks for a yearly cost of 71 thousand USD. The associated carbon footprint emissions is equivalent to less than 3700 t-$CO_{2\text{-eq}}$.$year^{-1}$. This optimum solution also allocates almost 10% of the nitrogen fertilization requirements of the stadium (Taylor et al., n.d.). In order to achieve these results, only two WWTP's water resources are required which are Doha South and

Doha North. As a matter of fact, the produced water of Lusail WWTP is not required as the two other plants provide sufficient water resources at lower economic and environmental costs. Figure 2 illustrates the distribution of water from the different sources to the different sinks. Most of the water comes from Doha South due to its location which happens to be closer to all the sinks considered in this study. For the case of Doha North, it is the major contributor for District cooling plant P2 and Al-Tammama Stadium.

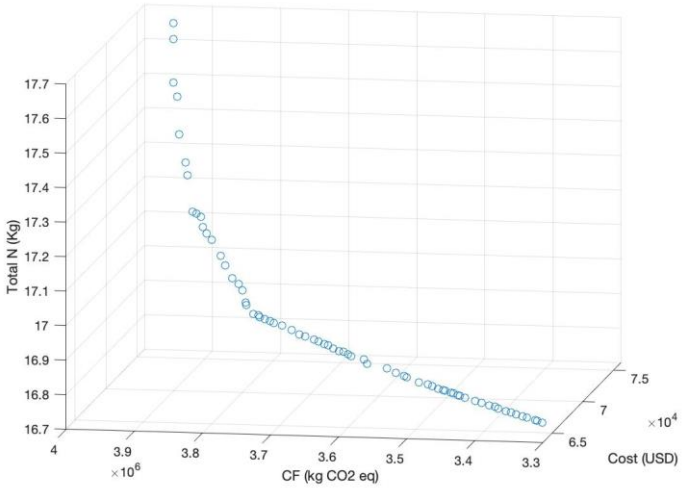


Figure 1: Three-dimensional pareto front water allocation model.

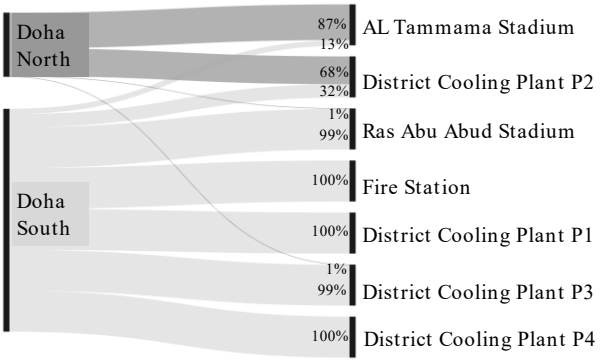


Figure 2: Summary of resource allocation to the different sinks.

4 Conclusions

In the light of the ever-increasing socio-economic and environmental stressors that threaten the availability of the renewable water base and considering the environmental burden associated with some alternative energy-intensive sources such as desalination, there is an imminent need to consider other sustainable options that can meet the demand while reducing the emissions. In this paper, the usage of TW as an alternate water supplier is investigated using a two-stage method. First, the acceptance of utilising TW in different

domains was assessed by means of a survey analysis. Afterwards, a multi-objective optimisation model was conducted to optimally allocate water between different accepted water sources and multiple application sinks counting firefighting, district cooling, street cleaning and stadiums' irrigation. The optimal solution generated exhibits relatively reduced economic and environmental costs which do not exceed 71 thousand USD and 3700 t-$CO_{2\text{-eq}}$.year^{-1} while supplying the stadiums with over 10% of their nitrogen requirements.

5 References

N. Almanaseer, M. Hindiyeh, R. Al-assaf, 2020, Hydrological and environmental impact of wastewater treatment and reuse on Zarqa river basin in Jordan, Environ., MDPI, 7.

S. Bauer, H.J. Linke, M. Wagner, 2020, Optimizing water-reuse and increasing water-saving potentials by linking treated industrial and municipal wastewater for a sustainable urban development, Water Sci. Technol., 81, 1927–1940,

K. Canaj, A. Mehmeti, D. Morrone, P. Toma, M. Todorović, 2021, Life cycle-based evaluation of environmental impacts and external costs of treated wastewater reuse for irrigation: A case study in southern Italy, J. Clean. Prod., 293, 126142,

C Tanks, n.d. Water Storage Requirements for Fire-Fighting Services [WWW Document].

F. Lahlou, H. R. Mackey, T. Al-Ansari, 2021, Wastewater reuse for livestock feed irrigation as a sustainable practice: A socio-environmental-economic review, J. Clean. Prod., 294, 126331,

F. Lahlou, H. R. Mackey, G. McKay, T. Al-Ansari, 2021, Reuse of treated industrial wastewater and bio-solids from oil and gas industries: Exploring new factors of public acceptance, Water Resour. Ind., 26, 100159,

F. Lahlou, H.R. Mackey, G. McKay, U. Onwusogh, T. Al-Ansari, 2020a, Water planning framework for alfalfa fields using treated wastewater fertigation in Qatar: An energy-water-food nexus approach, Comput. Chem. Eng. 141, 106999

F. Lahlou, S. Namany, H.R. Mackey, T. Al-Ansari, 2020b, Treated Industrial Wastewater as a Water and Nutrients Source for Tomatoes Cultivation: an Optimisation Approach, Computer Aided Chemical Engineering, 48, 1819–1824

P .S. Minhas, Khajanchi-Lal, R.K. Yadav, S.K. Dubey, R.K. Chaturvedi, 2015, Long term impact of waste water irrigation and nutrient rates: I, Performance, sustainability and produce quality of peri urban cropping systems. Agric. Water Manag,

M. Price, 2019, Irrigation Fundamentals, Water Well J

M. Shublaq, A.K. Sleiti, 2020, Experimental analysis of water evaporation losses in cooling towers using filters, Appl. Therm. Eng. 175, 115418,

G.R. Taylor, R.H. White, S. Abernathy, D. Smith, n.d., Athletic fields and water conservation.

Y. Wang, J. Li, S. Wang, J. Yang, C. Qi, H. Guo, X. Liu, H. Zhang, 2020. Operational optimization of wastewater reuse integrated energy system. Energy 200.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50211-6

Surrogate Modeling for Superstructure Optimization with Generalized Disjunctive Programming

H. A. Pedrozo[a,b]; S. B. Rodriguez Reartes[a,b]; A. R. Vecchietti[c],

M. S. Diaz[a,b,*], I. E. Grossmann[d]

[a]*Planta Piloto de Ingeniería Química (PLAPIQUI CONICET-UNS), Camino La Carrindanga km. 7, Bahía Blanca, Argentina*
[b]*Departamento de Ingeniería Química, Universidad Nacional del Sur (UNS), Bahía Blanca, Argentina*
[c]*Institute of Design and Development (INGAR CONICET-UTN), Avellaneda 3657, Santa Fe, Argentina*
[d]*Department of Chemical Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA*
sdiaz@plapiqui.edu.ar

Abstract

In this work, we propose an iterative framework to solve superstructure design problems, which includes surrogate models, with a custom implementation of the Logic-based Outer- Approximation algorithm (L-bOA). We build surrogate models (SM) using the machine learning software ALAMO exploiting its capability for selecting low-complexity basis functions to accurately fit sample data. To improve and validate the SM, we apply the Error Maximization Sampling (EMS) strategy in the exploration step. In this step, we formulate mathematical problems that are solved through Derivative Free Optimization (DFO) techniques. The following step applies the L-bOA algorithm to solve the GDP synthesis problem. As several NLP subproblems are solved to determine the optimal solution in L-bOA in the exploitation step, the corresponding optimal points are added to the SM training set. In case that an NLP subproblem turns out to be infeasible, we solve the Euclidean Distance Minimization (EDM) problem to find the closest feasible point to the former infeasible point. In this way, the entire information from NLP subproblems is exploited. As original model output variables are required, we solve EDM problems using DFO strategies. The proposed methodology is applied to a methanol synthesis problem, which shows robustness and efficiency to determine the correct optimal scheme and errors less than 0.2% in operating variables.

Keywords: superstructure optimization; surrogate models; disjunctive programming; derivative free optimization

1. Introduction

Advances in computers and mathematical modeling have enabled the detailed representation of process systems, and thus, the development of fundamental tools for decision making in process design. This scenario also presents new challenges. In mathematical programing, the standard method to formulate a problem is to declare all process unit equations to perform the optimization. However, when formulating highly

accurate models, some constraints or even the objective analytic function may not be available if they are evaluated through simulators or special programs. These functions that are not analytically available, are referred to as black-box models. When a mathematical problem includes both, explicit and black-box equations, it is referred to as hybrid or grey box model. A common approach to address this kind of problems includes building surrogate models (SMs) to replace the black-box models. SMs are simplified functions that can estimate output data from a set of input variables, requiring small CPU times.

When working with surrogate models, there is a trade-off between exploration and exploitation steps. Exploration strategies improve the global performance of the SM in the entire feasible region to reduce the probability of excluding the global optimum. On the other hand, exploitation-based methods refine the SM in regions where optima could be potentially found.

The interest of the Process Systems Engineering (PSE) community in developing efficient methods to address the formulation and solution of black/grey box problems has increased significantly in recent years (Bhosekar and Ierapetritou, 2018). Kim and Boukouvala (2020) developed a surrogate-based optimization procedure to solve mixed-integer nonlinear problems focused on avoiding the binary variable relaxation. Pedrozo et al. (2021a) proposed an iterative framework to address hybrid problems, replacing highly nonlinear equations for SM in order to reduce problem complexity. Thus, it was assumed that the analytic function was available for the exploration and exploitation steps.

In this work, we include Derivative Free Optimization (DFO) techniques (Zhao et al., 2021) in the exploration and exploitation steps to avoid using the analytic functions. Numerical results show that the strategy is efficient and accurate to address the synthesis problems and the generation and refinement of SMs.

2. Methodology

The proposed optimization framework is outlined in Fig. 1. Initially, lower and upper bounds are set for the input variables of each SM. The Latin Hypercube Sampling (LHS) technique is employed in MATLAB to generate sampling data. Output variables corresponding to each sampling point are obtained by performing simulations of the true or original model. When working with hybrid problems, a filtering step is required to discard infeasible sample points. Then, an initial SM is built in the machine learning software ALAMO (Wilson and Sahinidis, 2017) considering simple algebraic regression functions (SARFs).

Since the accuracy of this initial SM may not be good enough in all sampling points, we evaluate the corresponding relative errors, and we add Gaussians Radial Basis Functions (GRBFs) to represent those points whose errors are greater than a tolerance. In this way, we build the first SM based on both, SARFs and GRBFs, and then, we carry out the first exploration step. The Error Maximization Sampling (EMS) (Wilson and Sahinidis, 2017) strategy is applied in the exploration step. This method consists of maximizing the relative error of the SM in the feasible region. In this work, this optimization is performed through the DFO solver (Powell, 2009), which makes use of black-box simulation models. As a result, low-accuracy points of the domain are identified, and then interpolated by means of GRBF to improve the SM performance in that region, until the relative error is less than a tolerance or a maximum number of EMS problems is solved.

In the following step, we solve the hybrid model-based Generalized Disjunctive Programming (GDP) problem in GAMS. A custom implementation of the Logic-based Outer-Approximation (L-bOA) algorithm is employed (Pedrozo et al., 2021b, Pedrozo et

al., 2020). We exploit the information of the L-bOA subproblems to refine the SM in the exploitation step. The feasible NLP subproblem solutions are compared to the rigorous black-box simulations to assess the SMs accuracy in that region. As some NLP subproblems or black-box simulations might be infeasible due to the performance of the SMs, we formulate an optimization problem to determine the feasible sampling point that minimizes the Euclidean distance to the NLP subproblem solution, and this point is then added to the training set. This optimization problem is also solved using DFO solvers (Powell, 2009).

The iterative algorithm, which is shown in Fig. 1, stops when the specified convergence criterion is met. Otherwise, the exploration step is carried out again (the number of major iterations of the algorithm is equal to the times the GDP problem is solved).

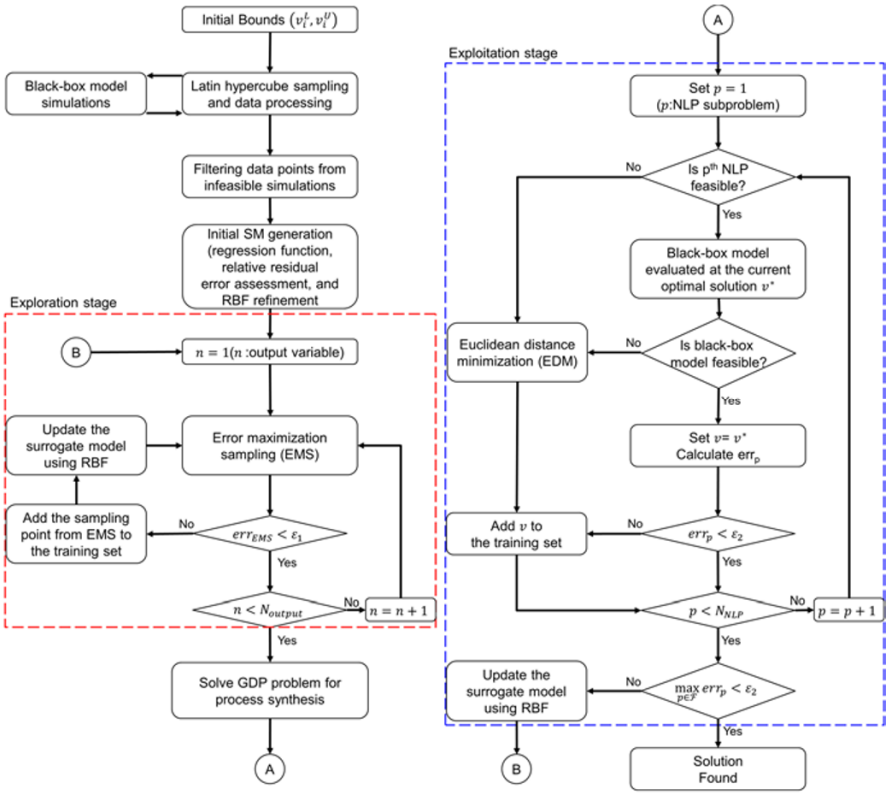


Figure 1: Iterative optimization framework

2.1. Software resources

The solution procedure is automated using MATLAB as a core for data transferring (see Fig. 2). In this way, ALAMO is run from MATLAB to generate the corresponding initial SMs. These functions and their derivatives are transferred to GAMS to formulate the hybrid GDP problem for process synthesis, and to solve it with the custom implementation of the L-bOA algorithm. To improve the SMs, we solve DFO problems for black-box models in the exploration and exploitation steps. In these cases, we employ the algorithms developed by Powell (2009), through the package provided by Ragonneau and Zhang (2021). Since these algorithms do not explicitly handle constraints, the Bound Optimization BY Quadratic Approximation (BOBYQA) algorithm is used.

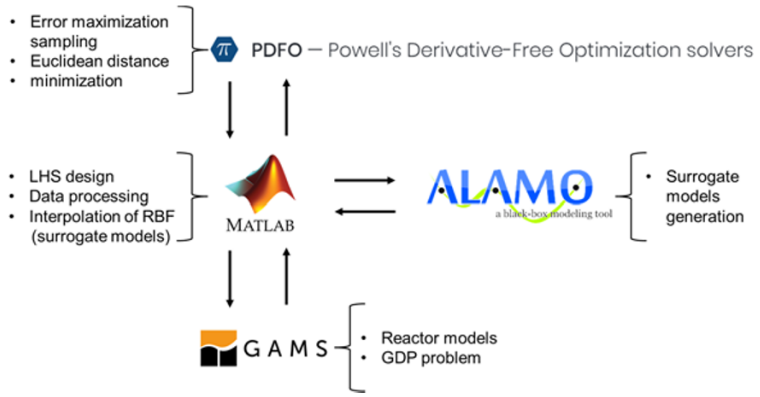


Figure 2: Software integration

3. Case Study

The methanol synthesis problem (Chen and Grossmann, 2019) is used as case study to test the proposed iterative algorithm. Figure 3 shows the process superstructure, where discrete decisions are represented using dashed lines for both, equipment and streams. The objective function is profit maximization.

In order to illustrate the algorithm, reactor models (units 9 and 10) are replaced by surrogate models to calculate conversion in each reactor. In this way, a hybrid formulation, which includes first principles and two SMs, is obtained.

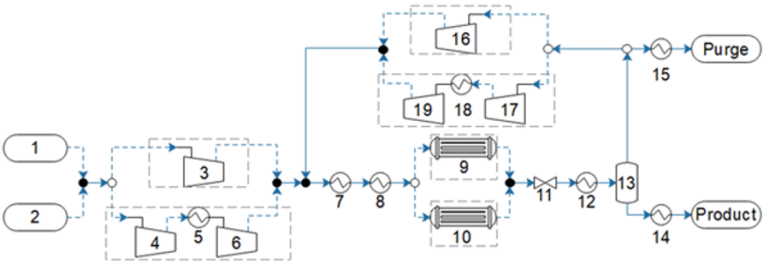


Figure 3: Superstructure for methanol synthesis (Chen and Grossmann, 2019)

4. Results

In order to show the robustness of the method and to consider the random component of the sampling technique, the problem is solved using ten different initial sampling data sets. In addition, we test 100 and 1,000 initial sampling points to assess the impact of the initial SM in the algorithm performance.

We observe that the iterative algorithm of Fig. 1 determines the optimal solution of the problem in each run (1,840 M$/y), and the error in the objective value is less than 0.2 % even in the worst case. Moreover, we observe that for the runs with 100 initial sampling points, the algorithm generally requires two major iterations (in 7 runs of 10, Fig.4) to satisfy the convergence criterion. Thus, these SMs are refined only one time in the neighbourhood of the optimal solution during the exploitation step to make them accurate enough. However, in the worst case, four major iterations are required to meet the convergence criterion. On the other hand, considering large initial sampling data (1,000 points), the proposed method generally converges in one iteration (in 5 runs of 10, Fig.4).

Accordingly, the initial SM after exploration step has enough accuracy, so no data points are included in the exploitation step. In the worst cases, four iterations are also required. This analysis indicates that we cannot guarantee the quality of the initial SM. Even working with a large initial sampling data set, SM refinement during the exploration and exploitation steps can be required to achieve the desired accuracy of the generated SMs.
Regarding the algorithm performance, Fig. 5 shows the corresponding CPU time distributions. On average, 11.9 and 2.5 minutes are the total CPU time for 100 and 1,000 initial sampling points, respectively. The exploration step is the most time consuming, followed by the exploitation step, while CPU times associated with the initial fit and the GDP problem solution are negligible. These results are related to the use of DFO strategies in the refinement steps. Solving either an optimization problem for the exploration or exploitation step with DFO methods, requires 40 s approximately. Thus, the quality of initial SMs strongly influences the method's performance. When the algorithm is run with a large initial sampling data set (1000 points), the SMs require less refinement, and consequently, fewer problems must be solved using DFO strategies, as compared to the case of using 100 initial sampling points. These results are in agreement with those from Wilson and Sahinidis (2017).
When comparing this strategy with the case of using NLP solvers (CONOPT) for the exploration and exploitation steps (Pedrozo et al. 2021a), there is a significant increase in CPU time, i.e., 15 s vs. 2.5 min for 1,000 initial sampling points on average.

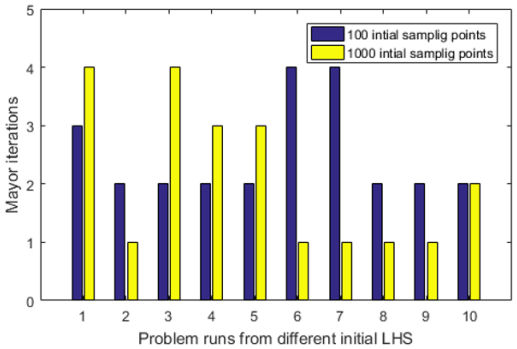


Figure 4: Major iterations of the iterative framework from different initial LHS sets

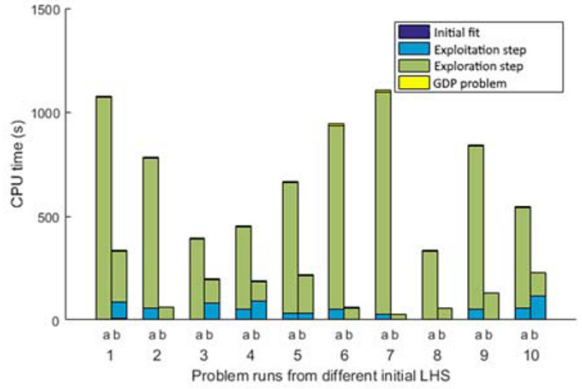


Figure 5: CPU time distribution. a) 100 initial sampling points. b) 1000 initial data sampling points. CPU times corresponding to initial fit and GDP problem solution are less than six seconds, so they are not easily distinguishable in the figure

5. Conclusions

In this work, we propose an algorithm for SMs generation and refinement using DFO strategies in the exploration and exploitation steps for the synthesis of process flowsheets using Generalized Disjunctive Programming with surrogate models. The algorithm has been tested with a methanol synthesis case study. The optimization tool has been proven to be robust and effective in generating solutions with relative errors lower than 0.2 % for the objective function in the worst cases, and obtaining the same optimal flowsheet as the rigorous model. The CPU time can be reduced by using a larger initial sampling point set. This strategy paves the way to efficiently refine SMs by the use of black-box models and DFO solvers.

References

Bhosekar, A., & Ierapetritou, M. (2018). Advances in surrogate based modeling, feasibility analysis, and optimization: A review. Comp. & Chem. Eng., 108, 250-267.

Chen, Q., & Grossmann, I. (2019). Modern modeling paradigms using generalized disjunctive programming. Processes, 7(11), 839.

Kim, S. H., & Boukouvala, F. (2020). Surrogate-based optimization for mixed-integer nonlinear problems. Comp. & Chem. Eng., 140, 106847.

M. J. D. Powell, The BOBYQA algorithm for bound constrained optimization without derivatives, Technical Report DAMTP 2009/NA06, Department of Applied Mathematics and Theoretical Physics, Cambridge University, Cambridge, UK, 2009

Pedrozo, H. A., Reartes, S. R., Bernal, D. E., Vecchietti, A. R., Díaz, M. S., & Grossmann, I. E. (2021a). Hybrid model generation for superstructure optimization with Generalized Disjunctive Programming. Comp. & Chem. Eng., 154, 107473.

Pedrozo, H. A., Reartes, S. R., Vecchietti, A. R., Díaz, M. S., & Grossmann, I. E. (2021b). Optimal design of ethylene and propylene coproduction plants with generalized disjunctive programming and state equipment network models. Comp. & Chem. Eng., 149, 107295.

Pedrozo, H. A., Rodriguez Reartes, S., Diaz, M. S., Vecchietti, A. R., Grossmann, I. E. (2020), Coproduction of ethylene and propylene based on ethane and propane feedstocks, Computer Aided Chemical Engineering, 48, 907-912. https://doi.org/10.1016/B978-0-12-823377-1.50152-X

T. M. Ragonneau and Z. Zhang, PDFO: Cross-Platform Interfaces for Powell's Derivative-Free Optimization Solvers (Version 1.1), available at https://www.pdfo.net

Wilson, Zachary T., and Nikolaos V. Sahinidis. "The ALAMO approach to machine learning." Comp. & Chem. Eng., 106 (2017): 785-795.

Zhao, F., Grossmann, I. E., García-Muñoz, S., & Stamatis, S. D. (2021). Flexibility index of black-box models with parameter uncertainty through derivative-free optimization. AIChE Journal, 67(5), e17189.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50212-8

Educational computer-aided tools towards Industry 4.0: recommendations and BioVL

Carina L. Gargalo[a], Simoneta Caño[a], Fiammetta Caccavale[a], Krist V. Gernaey[a], Ulrich Krühne[a*]

[a]*Process and Systems Engineering Centre (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søltofts Plads, Building 228 A, 2800 Kgs. Lyngby, Denmark*
ulkr@kt.dtu.dk

Abstract

Industry 4.0 is no longer a distant concept but an ongoing paradigm shift. Thus, it seems essential to prepare future engineers by fine-tuning our current educational approach. For example, advanced modelling, programming, and data analysis are now becoming fundamental skills. Therefore, it is essential to integrate these elements/subjects in the graduate and undergraduate curriculums. For a dynamic and active experience, educational computer-aided tools could be an excellent platform to teach fundamental engineering concepts and other essential tools for digitalization.

In this work, we propose the customization of the current curriculum and introduce those prerequisites for Industry 4.0's realization through an on-line and open-source educational computer-aided platform/simulator (BioVL). Hence, BioVL is built upon: (i) the identification of learning requirements; (ii) a learning design; and, (iii) a motivation strategy that includes gaming elements and an agile microlearning approach. BioVL (available at www.biovl.com) is in its prototype stage and under continuous development and refinement. Some of the ongoing and future steps include implementing AI-powered adaptive learning so that the students can receive immediate feedback and prevent error propagation.

To conclude, we encourage building on the pillars of a successful transition towards a digitalized industry by upgrading the curriculum and integrating on-line educational computer-aided tools to prepare future engineers to be fluent in data analysis and process modelling.

Keywords: biomanufacturing, digital models, digital platform, BioVL, education

1. Introduction

Industry 4.0, also frequently referred to as the Industrial Internet of Things or smart manufacturing, is no longer a distant concept but an ongoing paradigm shift. Among other processing industries, the biomanufacturing/biochemical industry now has the opportunity to adopt the technologies supporting smart manufacturing and intelligent automation. This leads increasingly to the introduction of digital solutions that focus on data communication/transfer, process modelling, and AI to predict and optimize the behavior of the process at all life cycle stages in real-time (Gargalo et al., 2021).
Henceforth, for this to be a reality, there is a high demand for qualified engineers. Today's engineers need to be ready for such developments, which poses a challenge since the typical engineering undergraduate and graduate curriculum is not yet ready. There is a

need to fine-tune the current educational approach in order to prepare engineers for a digitalized industry (Cameron et al., 2019; Narayanan et al., 2020). However, existing studies in the PSE field do not present neither propose a comprehensive perspective on how and what needs to change regarding education. Thus, we must answer a critical question: what do (bio)chemical engineers need to learn to contribute positively to making Industry 4.0 a reality?

Industry 4.0 entails implementing the concept of digital twins, which is, in a perfect world, a flawless digital replica of the physical process (Figure 1). It is a cyber-physical loop system where the different units in a manufacturing process interact with one another, share information, simulate outcomes that strengthen decision-making, and thus make adaptive decisions without, or with minimal, human intervention Presently in its prototype stage, BioVL (www.biovl.com, Figure 2) has been built to support blended learning. The main intention of this software platform is to provide more modelling and programming exercises to graduate and undergraduate students in the chemical and biochemical engineering study line. It focuses on explaining, solving, and enabling the modification of conditions and parameters in mechanistic models for bioprocesses-focused education (Deloitte, 2021; Gargalo et al., 2021; Narayanan et al., 2020; Udugama et al., 2021). Some of its enabling tools and methods are big data, data sharing, the Internet of Things, artificial intelligence (AI), and the development of process models, among many others. All the tools mentioned above share a common element: the demand for a computer platform/approach. Hence, we are convinced that educational computer-aided tools have the potential to become the preferred platform to teach, through an active experience, not only engineering fundamental concepts but also the essential tools for digitalization such as data analysis and basic and advanced modelling. For example, in a smart factory, engineers must now know how to combine process knowledge with the benefits brought by advanced data analysis (i.e., big data analytics). Big data analytics allows examining large data sets to uncover hidden patterns and trends and quickly identify a faulty batch or any other process faults in order to make informed decisions.

Although the available methods and tools for data analysis and other digitalization enablers are numerous and highly complex, they are not different among fields (e.g., mechanical engineering, business). However, process models are system-specific and highly depend on conditions and process boundaries. Therefore, the teaching of basic and advanced modelling should be rigorous and requires the selection and proper implementation of process models. This entails that, firstly, the students must understand the system and its requirements, as well as the adequate modelling strategy (e.g., first principles, hybrid model, surrogate model, etc.). Thus, when the appropriate model is chosen (depending on the complexity level required), it can be derived and described through a set of mathematical expressions to be further applied for process control and optimization, among other tasks. Therefore, for future engineers to be an active part of the Industry 4.0 movement and implementation, it is essential that there is a detailed understanding of process modelling and more complex computer-aided methods. In this study, we recommend and propose that the students acquire practical experience throughout their curriculum in order to be able to: (i) observe and recognize trends and patterns in the data; (ii) propose a model; (iii) implement it; (iv) test it; and finally, (v) use the model (e.g., for optimization purposes). Consequently, an educational computer-aided platform can be used which embeds these steps and provides a structural and theoretical frame for learning the needed methods and tools (e.g., big data, AI, etc.). Thus, the authors have developed *BioVL* (**Bio**process **V**irtual **L**aboratory), an educational platform whose goal is to tackle these challenges (Figure 1) and train the users in applied

modelling and programming (Caño De Las Heras et al., 2019). *BioVL*, presently in its prototype stage, is an open-source online simulator that aims to explain, solve, and modify bioprocess models and support the users in creating their own models. Features, as well as software design and targets, are discussed at length in the following sections. We believe this is a valuable tool to face the knowledge needs and challenges for the transition and continuous operation of smart biomanufacturing facilities.

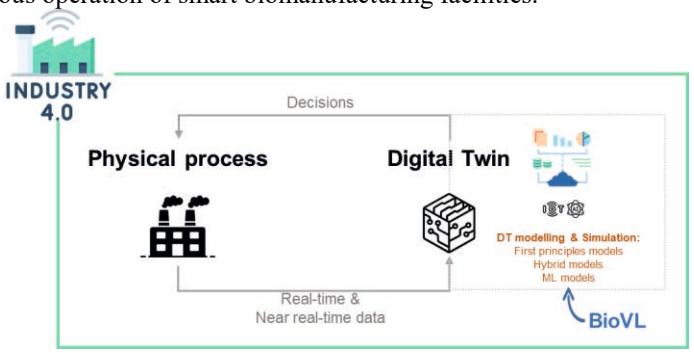


Figure 1: Industry 4.0 and the simplified relationship between the physical process and its Digital Twin. Orange: Knowledge gap in the (bio)chemical engineering curriculum (modelling and programming). Blue: BioVL contribution. Graphic illustration used elements from www.flaticon.com.

2. Methodology

As previously stated, educational platforms must become active enablers of the ongoing digital transformation (Caño De Las Heras et al., 2021b; Gargalo et al., 2021; Narayanan et al., 2020). To develop such a platform, as detailed in (Caño de las Heras et al., 2021), there are three main steps: identification of learning requirements, formulation and implementation of the learning design, and validation. The educational platform developed in this work is based on the mentioned steps, as described in the following sub-sections.

2.1. Identification of Learning requirements

Previous studies by the authors (Caño De Las Heras et al., 2021b, 2021c), as well as others (Balamuralithara & Woods, 2009; Dyrberg et al., 2017; Feisel & Rosa, 2005), have collected and quantified the students' perception on building on essential skills (e.g., programming) and on the use of simulators in their education. The study by (Caño De Las Heras et al., 2021b, 2021c) has revealed that, according to the students' opinion, the programming content in the curriculum is not sufficient to cover the industry's future needs. Furthermore, (Caño De Las Heras et al., 2021b) also reported that Python is the preferred language for modelling and optimization subjects. Further, based on the students' and teachers' perceptions, other topics are lacking, such as the use of collaborative learning; the students miss the social elements of traditional education (Caño de Las Heras et al., 2020) and/or wish for a more dynamic and iterative simulator for open-ended exploration/investigation.

The mentioned learning needs are used as the foundation for the learning design definition.

2.2. Learning Design formulation

A learning design is developed to overcome the mentioned pedagogical needs. It is built upon the (i) learning goals, (ii) background describing how the learning will occur, (iii)

learning content, theory and methods used, (iv) evaluation method, and (v) learning prerequisites (Caño de las Heras et al., 2021; Caño De Las Heras et al., 2021b; Hiim, H and Hippe, 1997). This didactic framework was first proposed (Hiim, H and Hippe, 1997) and later expanded (Weitze, 2016). Besides, the learning goals include the higher levels of the Bloom taxonomy associated with meta-cognitive knowledge (Caño De Las Heras et al., 2021b; Krathwohl, 2002).

The learning design is embedded in our educational computer-aided platform (BioVL). A central objective of BioVL is to teach applied modelling and programming in Python to graduate and undergraduate (bio)chemical students/engineers and potentially technicians. Hence, BioVL has several features such as multimedia resources, questionnaires, and activities described in detail in the Results and Discussion section.

2.3. Learning Design verification

Learning design is mostly based on "if…, then…" rules (Berggren et al., 2005) and commonly requires verification that it is valuable for the students. Therefore, many validation strategies have been previously proposed. However, in this work, we decided to intrinsically involve the students during different steps of the process. This strategy, known as co-participatory design, integrates the students as co-designers for content and platform development. It has also been proven that students can develop a sense of ownership and empowerment, while misunderstandings, confusions, and false impressions are minimized (Yamauchi, 2012).

3. Results & discussion: BioVL

Presently in its prototype stage, BioVL (www.biovl.com, Figure 2) has been built to support blended learning. The main intention of this software platform is to provide more modelling and programming exercises to graduate and undergraduate students in the chemical and biochemical engineering study line. It focuses on explaining, solving, and enabling the modification of conditions and parameters in mechanistic models for bioprocesses-focused education (Caño De Las Heras et al., 2021b). BioVL includes the following features: (a) collaborative learning through a chatbot; (b) information about bioprocess model formulation; (c) a simulator where realistic operational problems might occur requiring the student to propose a solution (Caño De Las Heras et al., 2019, 2021b); (d) a library of mechanistic models in the form of a functional database (Caño De Las Heras et al., 2021a); and, (e) a set of coding tasks in Python specifically related to bioprocess problems. Furthermore, to make the software engaging and easy to use, it employs gamification elements and agile microlearning (learning content is broken down into smaller segments that can be achived in minutes). In addition, students are active agents in the design of the content and layout of the software through co-participatory design. Due to the highlighted collaboration with students and teachers to make the best possible software, BioVL is far from finished. Future steps for BioVL's continuous improvement include introducing and teaching the students (i) more advanced modelling strategies by including optimization exercises; (ii) to develop a suitable control strategy for the system; and (iii) data analysis and its practical aspects by, for instance, introducing small experiments for the development of hybrid models. Furthermore, we are in the first stages of implementing AI-powered adaptive learning so that the students can receive immediate feedback and prevent error propagation.

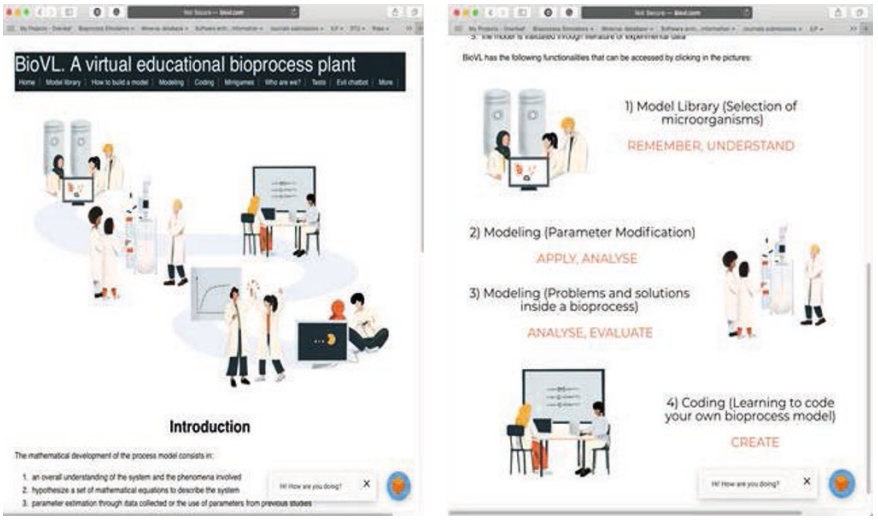


Figure 2: BioVL welcome screen. More at www.biovl.com.

4. Conclusions & Future perspectives

Educational computer-aided tools should be the preferred platform to teach, through an active experience, not only engineering fundamental concepts but also the essential tools for digitalization such as data analysis and advanced modeling. Engineers need to combine their process knowledge with the benefits of advanced data analysis to actively and dynamically contribute to the ongoing digitalization efforts and effective operation. Although AI methods do not vary among fields, process models are system-specific. Thus, background knowledge is essential for the successful selection, implementation/formulation, and use of process models (e.g., for advanced control and optimization purposes). Therefore, it is fundamental to integrate the teaching of basic and advanced modeling, as well as programming, in the graduate and undergraduate curriculums. To tackle the mentioned prerequisites of Industry 4.0-related activities, in this work, we propose to expand/customize the current curriculum (learning requirements), customize it to the current needs (learning design), and dynamically include the students in the process (learning verification). To achieve this, the authors have developed an online and open-source educational platform (BioVL) that, among other strategies, integrates gamification elements and an agile microlearning as a motivation approach. BioVL is in its prototype stage and under continuous development and refinement. Some of the ongoing and future steps include implementing AI-powered adaptive learning so that the students can receive immediate feedback and prevent error propagation. To conclude, we encourage building on the pillars of a successful transition towards a digitalized industry by upgrading the curriculum by integrating online educational computer-aided tools to empower future engineers to be fluent in data analysis and process modelling.

References

Balamuralithara, B., & Woods, P. C. (2009). Virtual laboratories in engineering education: the simulation lab and remote lab. *Computer Applications in Engineering Education*, *17*(1), 108–118.

Berggren, A., Burgos, D., Fontana, J. M., Hinkelman, D., Hung, V., Hursh, A., & Tielemans, G.

(2005). Practical and Pedagogical Issues for Teacher Adoption of IMS Learning Design Standards in Moodle LMS. *Journal of Interactive Media in Education*, *2005*(1), 3.

Cameron, I. T., Engell, S., Georgakis, C., Asprion, N., Bonvin, D., Gao, F., Gerogiorgis, D. I., Grossmann, I. E., Macchietto, S., Preisig, H. A., & Young, B. R. (2019). Education in Process Systems Engineering: Why it matters more than ever and how it can be structured. *Computers and Chemical Engineering*, *126*, 102–112.

Caño de las Heras, S., Gargalo, C. L., Weitze, C. L., Mansouri, S. S., Gernaey, K. V., & Krühne, U. (2021). A framework for the development of Pedagogical Process Simulators (P2Si) using explanatory models and gamification. *Computers and Chemical Engineering*, *151*.

Caño de Las Heras, S., Jones, Mark Nicholas Gernaey, K. V., Krühne, U., & Mansouri, S. S. (2020). An E-learning Bot for Bioprocess Systems Engineering. *Computer Aided Chemical Engineering*, *48*, 2023–2028.

Caño De Las Heras, S., Kensington-Miller, B., Young, B., Gonzalez, V., Krühne, U., Mansouri, S. S., & Baroutian, S. (2019). Facilitating learning by failure through a pedagogical model-based tool for bioprocesses. In *29th European Symposium on Computer Aided Process Engineering* (Vol. 46). Elsevier Masson SAS.

Caño De Las Heras, S., Kensington-Miller, B., Young, B., Gonzalez, V., Krühne, U., Mansouri, S. S., & Baroutian, S. (2021a). Minerva: a non-relational bioprocesses modeling database supportedby an ontology. *Submitted to Chemical Engineering Journal*.

Caño De Las Heras, S., Kensington-Miller, B., Young, B., Gonzalez, V., Krühne, U., Mansouri, S. S., & Baroutian, S. (2021b). Programming skills across the (bio)engineering curriculum – a students' perspective. *Computer Aided Chemical Engineering*, *50*, 2039–2044.

Caño De Las Heras, S., Kensington-Miller, B., Young, B., Gonzalez, V., Krühne, U., Mansouri, S. S., & Baroutian, S. (2021c). Benefits and Challenges of a Virtual Laboratory in Chemical and Biochemical Engineering: Students' Experiences in Fermentation. *Journal of Chemical Education*, *98*(3), 866–875.

Deloitte. (2021). *Tech Trends 2021 | Deloitte Insights*. https://www2.deloitte.com/us/en/insights/focus/tech-trends.html

Dyrberg, N. R., Treusch, A. H., & Wiegand, C. (2017). Virtual laboratories in science education: students' motivation and experiences in two tertiary biology courses. *Journal of Biological Education*, *51*(4), 358–374.

Feisel, L. D., & Rosa, A. J. (2005). The role of the laboratory in undergraduate engineering education. *Journal of Engineering Education*, *94*(1), 121–130.

Gargalo, C. L., de Las Heras, S. C., Jones, M. N., Udugama, I., Mansouri, S. S., Krühne, U., & Gernaey, K. V. (2021). Towards the Development of Digital Twins for the Bio-manufacturing Industry. *Advances in Biochemical Engineering/Biotechnology*, *176*, 1–34.

Hiim, H and Hippe, E. (1997). Learning through experience, understanding and action. *Gyldendal Uddannelse*.

Krathwohl, D. R. (2002). A revision of bloom's taxonomy: An overview. *Theory into Practice*, *41*(4), 212–218.

Narayanan, H., Luna, M. F., von Stosch, M., Cruz Bournazou, M. N., Polotti, G., Morbidelli, M., Butté, A., & Sokolov, M. (2020). Bioprocessing in the Digital Age: The Role of Process Models. *Biotechnology Journal*, *15*(1), 1–10.

Udugama, I. A., Lopez, P. C., Gargalo, C. L., Li, X., Bayer, C., & Gernaey, K. V. (2021). Digital Twin in biomanufacturing: challenges and opportunities towards its implementation. *Systems Microbiology and Biomanufacturing*, *1*(3), 257–274.

Weitze, C. L. (2016). Designing for learning and play: The smiley model as a framework. *Interaction Design and Architecture(S)*, *29*(1), 52–75.

Yamauchi, Y. (2012). Participatory Design. In *Field Informatics* (pp. 123–138). Springer Berlin Heidelberg.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50213-X

Process Superstructure Optimization through Discrete Steepest Descent Optimization: a GDP Analysis and Applications in Process Intensification

David E. Bernal[a*], Daniel Ovalle[a,c*], David A. Liñán[b], Luis A. Ricardez-Sandoval[b], Jorge M. Gómez[c], and Ignacio E. Grossmann[a]
[a] *Department of Chemical Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh PA 15213, United States of America*
[b] *Department of Chemical Engineering, University of Waterloo, Ontario N2L 3G1, Canada*
[c] *Departamento de Ingeniería Química y de Alimentos, Universidad de Los Andes, Bogotá 111711, Colombia*
[*] *These authors contributed equally to this work.*
grossmann@cmu.edu

Abstract

This manuscript introduces a Logic-based Discrete-Steepest Descent Algorithm (LD-SDA) to tackle problems arising from process superstructure optimization. These problems often appear in Process Systems Engineering and become challenging when addressing Process Intensification applications. The current algorithm considers a disjunctive interpretation of these optimization problems through Generalized Disjunctive Programming (GDP). This formulation allows further analysis of the solution method as a tailored approach for GDP and results in a general open-source implementation of the method relying on the modeling paradigm Pyomo.GDP. Complementing our previous studies in the subject, we compare the LD-SDA against other well-known GDP solution methods and a D-SDA that does not consider the disjunctive nature of these problems. The results showcase the advantages of LD-SDA when dealing with superstructure problems arising from process intensification.

Keywords: superstructure optimization; process intensification; convex discrete analysis.

1. Introduction

The optimal design of processes is a challenge the Process Systems Engineering (PSE) community faces. Chemical processes require a systematic procedure to find optimal designs to remain competitive. Recent developments from Process Intensification (PI) have shown to be promising alternatives to traditional processes by integrating and interconnecting units and achieving superior processes in terms of economic, environmental, and efficiency objectives (Sitter et al., 2019). Different process flowsheets can be integrated into a single process superstructure, where potential units and interconnections are considered. Superstructure models allow the units and interconnections' equations to be constraints in optimization problems.

Since these equations can involve nonlinear functions and depend on both continuous (e.g., flowrates or temperatures) and discrete variables (e.g., equipment choice, interconnection location), the mathematical models become Mixed-Integer Nonlinear Programs (MINLP). The solution to these optimization problems is challenging given their combinatorial and nonconvex nature. Generalized Disjunctive Programming (GDP)

has been proposed to tackle specific modeling and solution challenges of MINLP. In GDP, the modeling capabilities of traditional mathematical programming are extended by introducing Boolean variables involved in propositions and disjunctions.

The novelty of this work is to frame a Discrete-Steepest Descent Algorithm (D-SDA) for the solution of discrete nonlinear problems within the scope of GDP and use it to address process superstructure problems with ordered interconnections. Such problems arise in PI applications, such as studying a series of units with interunit refluxes, e.g., tray-by-tray distillation column models.

2. Generalized Disjunctive Programming

In general, a GDP problem can be written as

$$\begin{aligned} &\min_{\mathbf{x},\mathbf{Y},\mathbf{z}} f(\mathbf{x},\mathbf{z}) + \textstyle\sum_{k\in K} c_k \\ &\text{s.t. } \mathbf{g}(\mathbf{x},\mathbf{z}) \le 0;\ \boldsymbol{\Omega}(\mathbf{Y}) = \text{True} \\ &\veebar_{i\in D_k} \begin{bmatrix} Y_{ik} \\ \mathbf{r}_{ik}(\mathbf{x},\mathbf{z}) \le 0;\ c_k = \gamma_{ik} \end{bmatrix} \forall\, k \in K \\ &\mathbf{x} \in X \subseteq \mathbb{R}^{n_x};\ \mathbf{Y} \in \{\text{True}, \text{False}\}^{n_y};\ \mathbf{z} \in Z \subseteq \mathbb{Z}^{n_z};\ \mathbf{c} \subseteq \mathbb{R}^{|K|} \end{aligned} \tag{1}$$

Where the continuous variables are denoted by the n_x-dimensional vector $\mathbf{x}$ bounded by the finite set X, and the discrete variables are denoted by the n_z-dimensional vector $\mathbf{z}$, bounded by the finite set Z. The function $f: \mathbb{R}^{n_x} \times \mathbb{R}^{n_z} \to \mathbb{R}$ is the objective function, and the vector function $\mathbf{g}: \mathbb{R}^{n_x} \times \mathbb{R}^{n_z} \to \mathbb{R}^{n_l}$ denotes the global inequality constraints. $\mathbf{Y}$ is a n_y-dimensional vector of logic variables, where for each disjunct $i \in D_k$ of each disjunction ($\veebar$) $k \in K$ the individual logic variable Y_{ik} enforce the set of inequalities $\mathbf{r}_{ik}: \mathbb{R}^{n_x} \times \mathbb{R}^{n_z} \to \mathbb{R}^{n_{ik}}$ and the fixed cost γ_{ik}. Logical constraints $\boldsymbol{\Omega}: \{\text{True}, \text{False}\}^{n_y} \to \{\text{True}, \text{False}\}$ encode logical relationships among the logical variables.

Besides offering a more intuitive modeling paradigm of discrete problems through disjunctions, a GDP model can inform computational solution tools of the original problem's underlying structure of the original problem, thus leading to improved solving performance. The GDP framework has successfully addressed problems derived from process superstructure optimization (Chen et al., 2021).

GDP problems are often solved by reformulating them as MINLP problems, by adding a binary variable y_{ik} for each Boolean variable Y_{ik}, and reformulating the constraints $\mathbf{r}_{ik}$ within the disjunctions to be enforced when the corresponding variable $y_{ik} = 1$ or trivially satisfied otherwise. The two best-known cases are the Big-M and the Hull reformulation, for which the Big-M case requires fewer continuous variables while the Hull reformulation is always at least as tight as the Big-M reformulation.

The tailored solution methods for GDP are usually based on generalizing algorithms for MINLP. The optimization problems are decomposed in a way where the discrete variables are fixed into what we call a discrete combination and allow to solve the problem only in terms of the continuous variables. Different methods are used to select the combination of these discrete variables, including branching across the different values the discrete variables can take (i.e., Branch-and-Bound) or solving a linear approximation of the original problem (Kronqvist et al., 2019). For GDP algorithms, contrary to the case in MINLP, these (possibly Mixed-Integer) Nonlinear Programming (NLP) subproblems that arise when fixing a particular discrete combination, now including the logical variables,

only include the constraints that concern the logical variables within each combination. Namely, for a given logical combination $\widehat{\mathbf{Y}}$ the subproblem becomes

$$\begin{aligned} \min_{\mathbf{x},\mathbf{z}} \; & f(\mathbf{x},\mathbf{z}) + \sum_{ik \text{ if } \widehat{Y}_{ik}=\text{True}} \gamma_{ik} \\ \text{s.t. } & \mathbf{g}(\mathbf{x},\mathbf{z}) \leq 0 \\ & \mathbf{r}_{ik}(\mathbf{x},\mathbf{z}) \leq 0 \text{ if } \widehat{Y}_{ik} = \text{True } \forall i \in D_k, k \in K \\ & \mathbf{x} \in X \subseteq \mathbb{R}^{n_x}, \mathbf{z} \in Z \subseteq \mathbb{Z}^{n_z} \end{aligned} \tag{2}$$

Notice that in the most general case, the problem in Eq.(2) is an MINLP, although in most applications, $n_z = 0$, leading to the problem being an NLP. This problem avoids evaluating numerically challenging nonlinear equations whenever its corresponding logical variables are irrelevant (i.e., "zero-flow" issues).

The different tailored algorithms for GDP are defined in the strategy to find the logical combination $\widehat{\mathbf{Y}}$ such that subproblems as in Eq.(2) solve to the optimal solution Eq.(1). One alternative is using gradient-based linearizations of the nonlinear constraints at the optimal solution of Eq.(2) to approximate the original problem feasible region. This defines a Mixed-Integer Linear Program (MILP) whose optimal solution returns values for the integer combinations. This method is known as the Logic-based Outer-Approximation (LOA) method. One can also systematically explore the values of the Boolean variables in a search tree where the nodes correspond to partial fixations of these variables, whose solutions provide bounds to the optimal solution, in a method called the Logic-based Branch-and-Bound (LBB) method (Chen et al., 2021). Both methods seek to find potentially optimal combinations of logical variables efficiently.

3. Discrete Steepest Descent Optimization

In a previous study, we presented the D-SDA (Liñán et al., 2020a) based on the theory of discrete convex analysis (Murota, 1998). The algorithm aims to solve Mixed-Binary Nonlinear Programs (MBNLP) and relies on reformulating the original discrete problem, in terms of binary variables, into a problem of integer choices, referred to as external variables. This reformulation was designed for binary variables defined in an ordered set constrained to an assignment constraint, meaning that only one of these ordered binary variables can be 1. These external variables, which are no longer representable in the original problem constraints, provide a concise representation of the discrete feasible region. This structure often appears in process superstructure optimization problems, e.g., when a set of binary variables define the location of a reflux stream within a stages sequence, implicitly defining the existence of left-over stages after them.

Exploring discrete neighborhoods of the external variables provides the D-SDA with an efficient approach to choose which combination of the discrete variables should be considered to solve the subproblems that appear by fixing such values, NLPs in this case, thus efficiently searching the combinatorial space of the discrete variables. The D-SDA uses the integrally local optimality as a termination criterion (Murota, 1998), enabling the efficient solution of process superstructure optimization problems.

When considering a series of continuously stirred tank reactors (CSTR), the D-SDA outperforms MINLP solvers in solution time and quality (Liñán et al., 2020a). Furthermore, we applied the algorithm to the optimal design of a PI application involving reactive distillation, where we tackled the production of Ethyl tert-butyl ether (ETBE) from iso-butene and ethanol through the optimal design of a catalytic distillation column.

The D-SDA revealed a better performance against MINLP solvers when optimizing an economic objective in this problem (Liñán et al., 2020b). This allowed us to consider more complex models for this system, i.e., modeling multi-scale phenomena through a rate-based model for mass and energy transfer (Liñán et al., 2021).

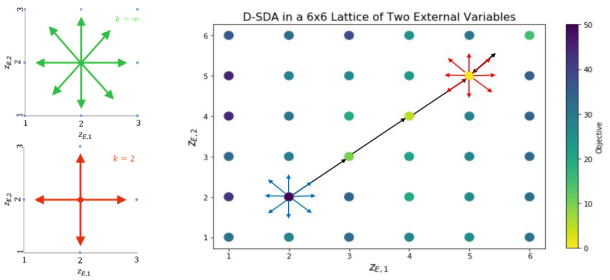


Figure 1. Pictorial representation of (a) different search neighborhoods in external variables lattice and (b) D-SDA with the neighbor and line search using $k = \infty$

4. Discrete-steepest descent optimization as a disjunctive algorithm LD-SDA

The problem in Eq. (1) suggests that the structure fitting for the D-SDA algorithm appears naturally in GDP, namely the disjunctions ($\vee_{i \in D_k} Y_{ik}$) enforce the assignment constraint, $\text{ExactlyOne}(Y_{1k}, \dots, Y_{|D_k|k})$, which would enable the reformulation with external variables if $\{Y_{ik} | i \in D_k\}$ represents an ordered decision. The reformulation is also possible across other ordered sets of Boolean variables, i.e., $\vee_{S(a) \in S} Y_{S(a)}$ defined over the ordered set $S = \{s_1, \dots, s_{|S|}\}$ with the element of S in its a-th position denoted $S(a)$. These sets are common within process superstructure problems and the reformulation into external variables $\mathbf{z}_E \in \{1, \dots, |S|\} \subseteq \mathbb{Z}$ becomes $Y_{S(a)} \Leftrightarrow z_E = a$. Notice that this reformulation allows the external variables to be interpreted as indicators of position within the set.

Such a reformulation allows us to map the Boolean variables into a lattice of integer variables, on which we can perform exploration based on ideas from discrete convex analysis (Murota, 1998). This leads to the Logic-based D-SDA (LD-SDA), which compares the objective function of each problem solved at a lattice site with its neighbors, defined by either a $k = 2$ or $k = \infty$ norm as seen in Fig.(1a), together with a line search along the direction provided by the best objective improvement after a complete neighbor search, as seen in Fig.(1b). The local optimality of the solution determines the stopping criterion compared to its neighbors (Murota, 1998).

From a GDP perspective, the external variables delineate a branching rule in the disjunctions, informing the problem structure. Notice that the more $\text{ExactlyOne}(Y_S)$ constraints the problem have, the more effective this reformulation is, with the limiting case of not having any other apart from the disjunctions, making it equivalent to LBB.

5. Numerical Experiments

We implement this method in open-source code using Python, available in https://github.com/bernalde/dsda-gdp. This code automatically transforms Pyomo.GDP (Chen et al., 2021) models, reformulates the disjunctions and the logical constraints $\text{ExactlyOne}(Y_{S(1)}, \dots, Y_{S(|S|)})$ automatically and solves the models using LD-SDA. We present the following two case studies after solving these problems with the solvers in GAMS 34.2 and using an Intel Core i7-7700 @ 3.6GHz PC with 16 Gb of RAM Memory.

5.1. Continuously Stirred Tank Reactors in series superstructure

We consider a superstructure of N_T CSTR in series where its total volume is minimized given an autocatalytic reaction $A + B \rightarrow 2B$ with 1st order reaction rate. This example is generalized from the one presented in (Liñán et al., 2020a). This example is illustrative given that we have an analytical solution at the limit of $N_T \rightarrow \infty$ equivalent to the Plug Flow reactor, and that we can explore the behavior in instances varying the value of N_T. Fig.(2) presents a scheme of the problem and its GDP formulation, together with its external variables reformulation. For this problem, there is a locally optimal solution with five reactors and reflux before the first reactor, $\mathbf{z}_E = (5,1)$.

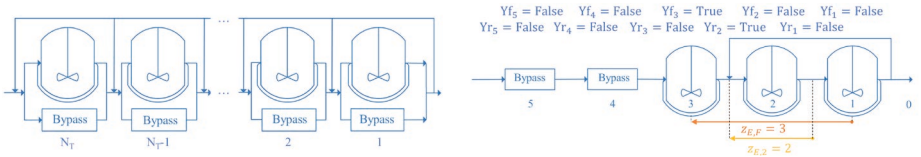


Figure 2. Scheme of (a) CSTR reactor superstructure and (b) Case with $N_T = 5$ and reformulation using external variables $z_{E,1}$ the number of reactors, $z_{E,2}$ the relative position of the reflux

We considered a set of different solver approaches to this problem with $N_T = [5, \dots, 25]$, including reformulating it into MINLP via Big-M and Hull reformulations, using LBB, LOA, and GLOA, and LD-SDA with two different norms, as seen in Fig.(3). We also include the total enumerations through the external variable reformulation.

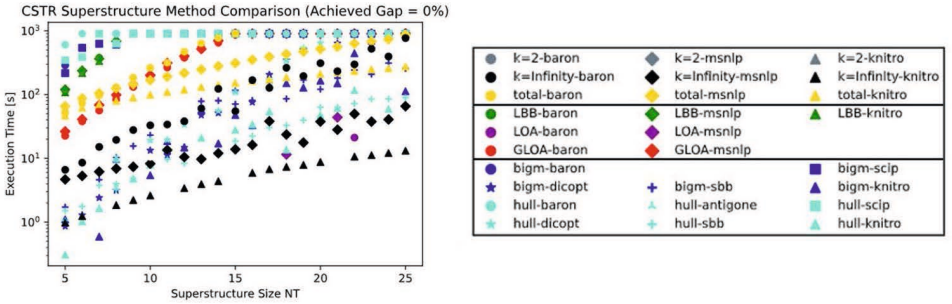


Figure 3. Execution time to achieve global minimum vs N_T for optimization of CSTR superstructure problem using different combinations of NLP solvers and reformulation methods.

From Fig.(3), one can see that LD-SDA provides the most efficient methods to solve this problem to global optimality. For this problem, the $k = 2$ norm neighborhood does not obtain the optimal solution. Notice that the external variable reformulation leads to a better search procedure, as seen when a total enumeration in the external variable space can be more efficient than other GDP solution alternatives.

5.2. Rate-based catalytic distillation column

The economic objective maximization of a catalytic distillation column to produce ETBE from butenes and methanol was solved using a D-SDA (Liñán et al., 2021). This test case is relevant since it deals with a PI design problem, where several traditional optimization methods fail even to compute feasible solutions (Liñán et al., 2021). The derivation of the D-SDA method was initially motivated to address this PI superstructure optimization problems, leading to a series of papers as seen in (Liñán et al., 2020b, 2020a, 2021).

The previous D-SDA would tackle the problem as an MBNLP, fixing and unfixing binary variables and including constraints of the form $y g(x) \leq 0$ to enforce the logic constraints.

Here we show that considering the problem from a disjunctive point of view as in LD-SDA leads to the solution of subproblems as in Eq.(2) instead of including irrelevant and numerically challenging nonlinear constraints.

Table 1. Execution time of Catalytic distillation optimal design problem from (Liñán et al., 2021)

	D-SDA: (Liñán et al., 2021)				**LD-SDA: This work**			
NLP Solver	CONOPT		MSNLP		CONOPT		MSNLP	
Neighborhood	k=2	k=inf	k=2	k=inf	k=2	k=inf	k=2	k=inf
Time [s]	367.1	16880.0	3626.0	102030.7	118.7	6751.1	2000.0	38532.5

As seen in Table (1), the proposed LD-SDA method leads to speedups up to 3x in this challenging PI problem. We could obtain the same solution to all subproblems more efficiently, given that only the relevant constraints were included for each problem. Adding to the fact that the previous results using the D-SDA were already beating state-of-the-art MINLP solution methods shows the advantages of the LD-SDA.

6. Conclusions

The current manuscript presents the usage of a disjunctive discrete steepest descent optimization algorithm LD-SDA to tackle process superstructure problems. This algorithm is presented from the perspective of Generalized Disjunctive Programming solution methods, showing its relationship with existing algorithms for GDP. Moreover, this allowed for the algorithm to be implemented in Python and through the modeling paradigm of Pyomo.GDP. With this implementation, we solved problems of superstructure optimization, a series of CSTR volume minimization, and a rate-based catalytic distillation column economical design more efficiently than other proposed solution methods, including MINLP reformulations, GDP-tailored algorithms, and a D-SDA aimed at MBNLP problems. The results in this manuscript show how LD-SDA becomes a valuable tool to address process superstructure problems, of which many challenges instances arise from PI applications.

References

Chen, Q., Johnson, E. S., Bernal, D. E., Valentin, R., Kale, S., Bates, J., Siirola, J. D., & Grossmann, I. E. (2021). Pyomo.GDP: an ecosystem for logic based modeling and optimization development. *Optimization and Engineering*.

Kronqvist, J., Bernal, D. E., Lundell, A., & Grossmann, I. E. (2019). A review and comparison of solvers for convex MINLP. *Optimization and Engineering*, 1–59.

Liñán, D. A., Bernal, D. E., Gómez, J. M., & Ricardez-Sandoval, L. A. (2021). Optimal synthesis and design of catalytic distillation columns: A rate-based modeling approach. *Chemical Engineering Science*, *231*, 116294.

Liñán, D. A., Bernal, D. E., Ricardez-Sandoval, L. A., & Gómez, J. M. (2020a). Optimal design of superstructures for placing units and streams with multiple and ordered available locations. Part I: A new mathematical framework. *Computers & Chemical Engineering*, *137*, 106794.

Liñán, D. A., Bernal, D. E., Ricardez-Sandoval, L. A., & Gómez, J. M. (2020b). Optimal design of superstructures for placing units and streams with multiple and ordered available locations. Part II: Rigorous design of catalytic distillation columns. *Computers & Chemical Engineering*, *139*, 106845.

Murota, K. (1998). Discrete convex analysis. *Mathematical Programming, Series B*.

Sitter, S., Chen, Q., & Grossmann, I. E. (2019). An overview of process intensification methods. In *Current Opinion in Chemical Engineering*.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50214-1

Advances in Generalized Disjunctive and Mixed-Integer Nonlinear Programming Algorithms and Software for Superstructure Optimization

David E. Bernal[a], Yunshan Liu[a], Michael L. Bynum[b], Carl D. Laird[a], John D. Siirola[b], and Ignacio E. Grossmann[a]

[a]*Department of Chemical Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh PA 15213, United States of America*
[b]*Discrete Mathematics and Optimization, Sandia National Laboratories, 1515 Eubank SE, Albuquerque NM, 87185, United States of America*
grossmann@cmu.edu

Abstract

This manuscript presents the recent advances in Mixed-Integer Nonlinear Programming (MINLP) and Generalized Disjunctive Programming (GDP) with a particular scope for superstructure optimization within Process Systems Engineering (PSE). We present an environment of open-source software packages written in Python and based on the algebraic modeling language Pyomo. These packages include MindtPy, a solver for MINLP that implements decomposition algorithms for such problems, CORAMIN, a toolset for MINLP algorithms providing relaxation generators for nonlinear constraints, Pyomo.GDP, a modeling extension for Generalized Disjunctive Programming that allows users to represent their problem as a GDP natively, and GDPOpt, a collection of algorithms explicitly tailored for GDP problems. Combining these tools has allowed us to solve several problems relevant to PSE, which we have gathered in an easily installable and accessible library, GDPLib. We show two examples of these models and how the flexibility of modeling given by Pyomo.GDP allows for efficient solutions to these complex optimization problems. Finally, we show an example of integrating these tools with the framework IDAES PSE, leading to optimal process synthesis and conceptual design with advanced multi-scale PSE modeling systems.

Keywords: superstructure optimization; generalized disjunctive programming; MINLP.

1. Introduction

Process superstructure optimization is a challenging problem within Process Systems Engineering (PSE). Using a mathematical programming formulation, the problem of superstructure optimization can be written as a set of constraints to be satisfied by selecting the values of variables while optimizing an objective function. These variables can be continuous, and represent properties of a process (e.g., temperature, pressure), or discrete, representing discrete choices (e.g., selecting a piece of given equipment). The constraints include the superstructure model equations (e.g., mass and energy balances, thermodynamic equations), and the objective function is a goal to reach by selecting the decision variables (e.g., maximize profit, minimize environmental impact). It is particularly challenging to obtain the globally optimal solutions to problems whose nonlinear constraints describe a non-convex region (Kronqvist et al., 2019).

This paper covers two different modeling paradigms for such superstructure optimization problems, Mixed-Integer Nonlinear Programming (MINLP), which relies only on algebraic functions of discrete and continuous variables, and Generalized Disjunctive Programming (GDP) that considers disjunctions, logical variables, and constraints for these problems. After mentioning the different solution methods for each paradigm, we present a set of examples of process superstructure optimization problems that have been modeled and solved using each of these approaches. The modeling has been done through the open-source software Pyomo.GDP (Chen et al., 2021), and made available as part of the problem library GDPLib. Finally, we demonstrate how these modeling tools and solution methods can be applied to a more intricate model created through the advanced modeling framework IDAES PSE (Miller et al., 2018).

2. Mixed-Integer Nonlinear Programming

The optimization models obtained from process superstructures have traditionally been written in algebraic equations and variables with both continuous and discrete domains. Mathematical optimization models with these characteristics are known as Mixed-Integer Nonlinear Programs (MINLP). The solution methods for this challenging type of optimization problem usually rely on the separate treatment of the two sources of complexity of the problem, the nonlinearity of the constraints and the integer variables' discreteness. Among the best-known deterministic solution strategies for these problems, we can count the Branch-and-Bound (BB) method and decomposition methods.

Both methods rely on finding bounds to the optimal objective function value through relaxations and restrictions of the original problem. A relaxation accounts for a different optimization problem whose feasible region is larger than the one of the original problem, whose solution is an optimistic bound of the optimal solution. Among these relaxations, the usual ones are continuous relaxations, where the discreteness of the integer variables is ignored, yielding a continuous problem, and linear relaxations, where the nonlinearities of the problems are replaced by linear feasible region that encompasses the feasible region of the original problem. These relaxations are not unique, and the successful solution of these problems can strongly depend on how close the relaxation approximates the original feasible region, known as its tightness, and other factors such as its size. The restriction of the original problem usually appears when fixing the value of some of the discrete variables, leading to a continuous problem in a lower-dimensional space. In the case that the original problem's objective is minimized, the optimal solution of a relaxation yields a lower bound of the optimal objective function, while any feasible solution to the problem, usually found through a partial or total fixing of the discrete variables and an optimization on the remaining variables, leads to an upper bound of the optimal solution.

In the BB method, starting from the solution of the continuous relaxation of the original MINLP, one systematically enforces values on the discrete variables to explore increasingly smaller and restricted subproblems. The solution of specific subproblems allows the derivation of extra inequalities that can help better approximate the original problem's feasible region, improving the quality of the lower bound obtained by solving it. Although effort has been made to derive strong inequalities, or cuts, for the nonlinear case and this can be generalized for branching on continuous variables, the BB method is better known for its highly successful implementations in modern solvers when addressing Mixed-Integer Linear Programming (MILP).

The decomposition methods for MINLP usually rely on MILP relaxations, which can be efficiently solved through BB, and continuous subproblems. As with BB, the MILP relaxations can be improved iteratively using cuts derived from the solution of the continuous restrictions. Eventually, if the relaxations are rigorous and each subproblem is solved optimally, either the decomposition methods or the BB methods will find the optimal solution in a finite number of steps/iterations. These two algorithms are the main ingredients of most known MINLP solvers (Kronqvist et al., 2019).

Considering this, a solver for MINLP is usually a meta-solver, where the solution of the original problem relies on other solvers to tackle subproblems. This observation has led us to develop the open-source Mixed-Integer Nonlinear Decomposition Toolbox in Pyomo - MindtPy (Bernal et al., 2018). This solver uses the interface that the Python-based algebraic modeling language Pyomo (Hart et al., 2017) has to solvers of continuous problems and MILP solvers and provides a flexible implementation of several of the decomposition methods known in the literature, such as the Outer-Approximation method. Furthermore, it includes implementations of heuristic techniques and enhancements such as single-tree solution methods and regularization-based algorithms. Furthermore, the derivation of strong relaxations to nonlinear terms is vital to solving these problems efficiently. We have also developed the open-source software CORAMIN (Bynum et al., 2019), which generates easily refinable relaxations of a Pyomo model's nonlinear constraints. These relaxations can be integrated within MINLP algorithms.

The convergence to the optimal solution of the MINLP is guaranteed when the relaxations are valid and can be further refined after each subsequent iteration and when the continuous subproblems are guaranteed to be solved to global optimality. This is easier to achieve when assuming well-behaved nonlinearities, e.g., convexity in the nonlinear functions. In this case, the linear relaxation can be found through the 1st-order Taylor expansions of the nonlinear functions in a method known as the Outer-Approximation (OA). When the convexity assumption is not satisfied, disciplined relaxations can still be derived as implemented in CORAMIN or through generalized McCormick relaxations (Scott et al., 2011) available in the software MC++. These allow our methods to solve even non-convex MINLP problems using a Global Outer-Approximation (GOA).

3. Generalized Disjunctive Programming

A more natural framework to represent superstructure optimization problems is Generalized Disjunctive Programming (GDP), which extends the modeling capabilities of traditional mathematical programming by incorporating logic variables involved in propositions and disjunctions. In general, a GDP problem can be written as

$$\begin{aligned} &\min_{\mathbf{x},\mathbf{Y},\mathbf{z}} \mathrm{f}(\mathbf{x},\mathbf{z}) \\ &\mathrm{s.t.}\ \mathbf{g}(\mathbf{x},\mathbf{z}) \le 0;\ \Omega(\mathbf{Y}) = \mathrm{True} \\ &\vee_{\mathrm{i}\in D_k} \begin{bmatrix} \mathrm{Y_{ik}} \\ \mathbf{r_{ik}}(\mathbf{x},\mathbf{z}) \le 0 \end{bmatrix} \forall\, \mathrm{k} \in \mathrm{K} \\ &\mathbf{x} \in \mathrm{X} \subseteq \mathbb{R}^{n_x};\ \mathbf{Y} \in \{\mathrm{True},\mathrm{False}\}^{n_y};\ \mathbf{z} \in \mathrm{Z} \subseteq \mathbb{Z}^{n_z} \end{aligned} \tag{1}$$

where the continuous variables are denoted by the n_x-dimensional vector $\mathbf{x}$ within a bounded set X, the discrete variables are denoted by the n_z-dimensional vector $\mathbf{z}$ within a bounded set Z, the function $f: \mathbb{R}^{n_x} \times \mathbb{R}^{n_z} \to \mathbb{R}$ is the objective function, and the vector function $\mathbf{g}: \mathbb{R}^{n_x} \times \mathbb{R}^{n_z} \to \mathbb{R}^{n_g}$ denotes the global constraints. Besides, the logical

structure of the problem includes $\mathbf{Y}$ as a n_y-dimensional vector of logic variables, where for each disjunct $i \in D_k$ of each disjunction $k \in K$ the set of inequalities $\mathbf{r}_{ik}: \mathbb{R}^{n_x} \times \mathbb{R}^{n_z} \rightarrow \mathbb{R}^{n_{ik}}$ are enforced by the individual logic variables Y_{ik}, and $\Omega: \{\text{True}, \text{False}\}^{n_y} \rightarrow \{\text{True}, \text{False}\}$ that encodes logical relationships among the logical variables. Notice that if the disjunctions set K is empty, Eq.(1) represents an MINLP problem.

The modeling framework of GDP allows for a more intuitive representation of the problems arising from superstructure optimization. The structure of such a representation can be exploited by a tailored solution algorithm to solve these problems more efficiently. First, reformulations can convert a GDP into MINLP, such as the Big-M and the Hull reformulations. These reformulations include all the constraints in the GDP and enforce or make trivially satisfiable constraints depending on the values of newly introduced binary variables $y_{ik} \in \{0,1\} \leftrightarrow Y_{ik} \in \{\text{True}, \text{False}\}$. The capability of writing a model directly as a GDP is included in the open-source code Pyomo.GDP (Chen et al., 2021).

We present two algorithms that generalize ideas from MINLP for GDP: the Logic-based Branch-and-Bound (LBB) and the Logic-based Outer-Approximation (LOA) algorithms. Like their MINLP counterpart, these algorithms have a search strategy for the values of the discrete variables, including the logic variables $\mathbf{Y}$, but consider the logical constraint Ω, pruning it. On the other hand, by leveraging the existing structure provided by the disjunctive formulation of the GDP, some algorithms selectively remove constraints that are not involved in each combination of the logical variables while exploring that combination. This approach is beneficial given the numerical issues that can appear from evaluating nonlinear constraints on vanishing variables, i.e., "zero-flow".

Similar to the MINLP case, if the subproblems are solved to optimality, for example, through the global solvers mentioned earlier, the relaxations of the nonlinear constraints are built rigorously, using MC++ or tailored relaxations as those in CORAMIN, these algorithms can solve non-convex problems to global optimality. We call this method Global Logic-based Outer Approximation (GLOA). These algorithms are implemented in the open-source GDP solver GDPOpt (Chen et al., 2021).

4. GDPLib, the library for GDP models

Finally, we highlight the usability of our framework and solution methods by solving different process superstructure problems. These problems have been previously presented in the literature, mainly as MINLPs. We included several instances as Pyomo.GDP problems in the repository github.com/grossmann-group/gdplib.

The library currently contains nine different examples of process or unit superstructure optimization, including a Methanol production process (Türkay & Grossmann, 1996), a Hydrodealkylation (HDA) process to produce Toluene (Kocis & Grossmann, 1989), a biofuel processing network, a heat exchanger network evaluating modular process design, a plant capacity expansion model, a synthesis gas production plant from methane, a Kaibel distillation column, and a tray distillation column design. Several examples include a few test cases leading to 25 GDP problems related to PSE. These instances range from $n_x \in [6,31968]$, $n_y \in [2,516]$, $n_z \in [0,5040]$, $n_g \in [30,14927]$.

We show two cases with more detail related to process superstructures. The first detailed case for the process superstructure optimization is the profit maximization of a methanol production process (Türkay & Grossmann, 1996). Mass balances define the global constraints of the problem, and there are 4 disjunctions in this problem as seen in Fig.(1a),

one associated with the feed choice, another one choosing between a high-cost and high-conversion reactor or a low-cost and low-conversion reactor, and having a single or two-stage compression for the feed and the recycle. This problem involves 285 variables, of which 8 are Boolean and the remaining continuous. The total number of configurations is 2^4 but GDPOpt using LOA requires only two iterations to find the optimal solution.

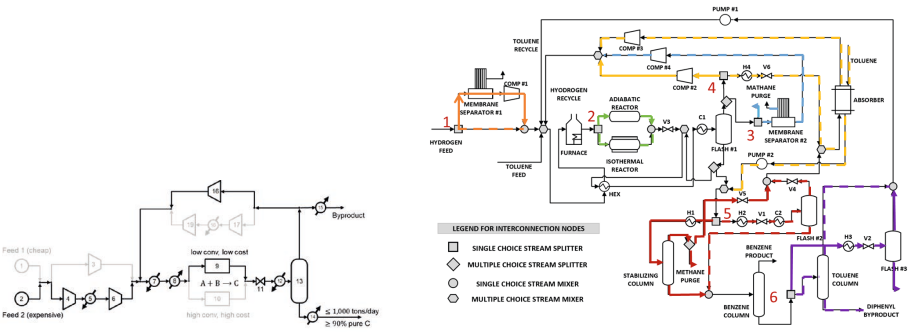


Figure 1. Flowsheet superstructure for (a) Methanol production and (b) Hydrodealkylation of Toluene. Alternatives highlighted correspond to optimal solutions.

The second example is the profit maximization of the Toluene production through the HDA process (Kocis & Grossmann, 1989). This superstructure, shown in Fig.(1b), considers 6 disjunctions: the choice to pretreat the hydrogen feed, whether to use an adiabatic or isothermal reactor, whether to purge or treat a methane stream for it to be recycled, considering the installation of an absorber or recycle a vapor stream, using a stabilizing column or a flash to remove extra methane from the process, and whether to use a distillation column to separate Toluene or a flash to obtain Diphenyl as a byproduct. There are 733 variables, with 12 being Boolean and the rest continuous, and 728 constraints, of which 12 are nonlinear. The MINLP transformation of this model is quite challenging for solvers, and both ANTIGONE and BARON fail to solve this problem to global optimality. Through LOA, GDPOpt was able to find the globally optimal solution, verified via enumeration. When using IPOPT and Gurobi as subsolvers, LOA converged to that solution in 17 iterations and 1 minute of computation in a standard desktop.

We highlight that the modeling paradigms and algorithms presented here, given their roots in Pyomo, can be used within more complicated process modeling alternatives. That is the case of the next-generation multi-scale PSE framework IDAES. By being based on Python and Pyomo, this framework leads to supporting our implementations natively. This results in the potential use of detailed process models, including property and thermodynamic packages and a disjunctive framework. We do this through the Methanol production example, which has been reimplemented as an IDAES PSE model and is available in GDPLib. The integration with IDAES PSE allows considering rigorous thermodynamic properties, resulting in more challenging optimization problems.

5. Conclusions

This paper presents two modeling paradigms for process superstructure optimization problems: Mixed-Integer Nonlinear Programming (MINLP) and Generalized Disjunctive Programming (GDP). MINLP is traditionally used and for which powerful solvers have been developed. On the other hand, GDP can be not only transformed into MINLP through different reformulations, leading to a difference in solution performance but can

also be solved directly by algorithms that take advantage of the disjunctive and logical structure encoded in it. We have developed open-source software tools for such models to be implemented, as is the case for Pyomo.GDP, and solved, which is the case for MindtPy and GDPOpt, within the algebraic modeling language in Python, Pyomo.

Moreover, this paper presents process superstructure optimization problems that have been implemented using GDP. These Python implementations are available in the repository GDPLib and are freely available and installable through the package manager pip (pip install gdplib). We show two examples of flowsheet superstructure optimization, namely Methanol and Toluene production processes. Finally, we include the Methanol production process case implemented using the IDAES PSE framework. This example highlights the applicability of this paper's modeling and algorithmic ideas to an advanced process modeling framework, enabling conceptual process design through superstructure optimization integrated with rigorous property models and unit operations blocks.

We hope this library leads process designers to adopt these modeling paradigms and algorithm developers to use these examples as a testbed to improve solution methods for these optimization problems. We envision more models becoming part of GDPLib, leading to a richer resource for the process design and optimization communities.

Disclaimer: Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

References

Bernal, D. E., Chen, Q., Gong, F., & Grossmann, I. E. (2018). Mixed-Integer Nonlinear Decomposition Toolbox for Pyomo (MindtPy). In *Computer Aided Chemical Engineering*.

Bynum, M., Castillo, A., Laird, C., Watson, J.-P., & USDOE. (2019). *Coramin v. 0.1 Beta, Version v. 0.1*.

Chen, Q., Johnson, E. S., Bernal, D. E., Valentin, R., Kale, S., Bates, J., Siirola, J. D., & Grossmann, I. E. (2021). Pyomo. GDP: an ecosystem for logic based modeling and optimization development. *Optimization and Engineering*, 1–36.

Hart, W. E., Laird, C. D., Watson, J.-P., Woodruff, D. L., Hackebeil, G. A., Nicholson, B. L., Siirola, J. D., & others. (2017). *Pyomo-optimization modeling in Python* (Vol. 67). Springer.

Kocis, G. R., & Grossmann, I. E. (1989). Computational experience with dicopt solving MINLP problems in process systems engineering. *Computers and Chemical Engineering*, *13*(3), 307–315.

Kronqvist, J., Bernal, D. E., Lundell, A., & Grossmann, I. E. (2019). A review and comparison of solvers for convex MINLP. *Optimization and Engineering*, 1–59.

Miller, D. C., Siirola, J. D., Agarwal, D., Burgard, A. P., Lee, A., Eslick, J. C., Nicholson, B., Laird, C., Biegler, L. T., Bhattacharyya, D., Sahinidis, N. V., Grossmann, I. E., Gounaris, C. E., & Gunter, D. (2018). Next Generation Multi-Scale Process Systems Engineering Framework. In *Computer Aided Chemical Engineering*.

Scott, J. K., Stuber, M. D., & Barton, P. I. (2011). Generalized McCormick relaxations. *Journal of Global Optimization*.

Türkay, M., & Grossmann, I. E. (1996). Logic-based MINLP algorithms for the optimal synthesis of process networks. *Computers & Chemical Engineering*, *20*(8), 959–978.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50215-3

Designing Novel Structured Packings by Topology Optimization and Additive Manufacturing

Andreas Lange[a*], Georg Fieg[a]

[a]*Hamburg University of Technology, Am Schwarzenberg-Campus 4, 21073 Hamburg, Germany*
a-lange@tuhh.de

Abstract

Additive manufacturing (AM) reveals a completely new freedom in design and development of structured packings for thermal separation columns. This potential might lead to the next generation of high-performance packings, but it can only be fully used if novel design methods are developed. One of these innovative design methods is presented in this contribution. A topology optimization approach based on the coupling of a stochastic optimization algorithm and computational fluid dynamics (CFD) simulations is applied to generatively design structured packings. By its application, novel structured packing shapes may be found. Binary elements, which are either defined as packing material or as empty elements, are considered as design variables in a defined design space. A multi-objective genetic algorithm with tailored process- and manufacturing-related constraints is used to identify the best packing material distribution within the column shell, revealing minimized pressure drop and maximized surface area. In this paper, the optimization tool and CFD model are presented before selected results of an exemplary topology optimization study are given. The objective of this study is the development of a packing element for a lab scale distillation column. The promising results prove the viability of the design method, showing that it is possible to generatively design structured packings algorithm-based and without any well-defined initial packing geometries as starting point.

Keywords: Generative Design; Topology Optimization; Structured Packings; Additive Manufacturing

1. Introduction

Structured packings are column internals, which are widely applied in thermal separation apparatuses like distillation columns to intensify the mass transfer between the present fluid phases. Besides maximized separation efficiency, minimal pressure drop and maximal capacity are essential parameters in leading to a successful separation process. Improvements in these key parameters directly affect the performance of the entire column, resulting in a more efficient and sustainable separation process.

The three highlighted performance parameters are directly correlated to the flow behaviour of the fluid phases, which is in turn strongly affected by the specific shape of the structured packing. For this reason, the shape of structured packings has been of research interest for decades (Spiegel and Duss, 2014), but previous developments were restricted by the limitations of conventional manufacturing methods. Additive manufacturing as a class of innovative manufacturing processes makes it now possible to realize novel packing shapes, which are not feasible to manufacture by conventional

methods due to the high complexity of the shapes. Accordingly, a new freedom in design is available, which can result in the next generation of high-performance structured packings using AM as the key technology. However, the main challenge is the identification of novel packing shapes in a systematic and targeted way. Therefore, advanced design methods based on mathematical optimization techniques and tailored computational tools for the targeted identification and systematic development of structured packings are essential. These design methods and tools need to use the potential of the high freedom in design, while simultaneously considering the specific constraints of both the additive manufacturing process and thermal separation applications.

In general, three different structural optimization approaches, classified as sizing (Neukäufer et al., 2019), shape optimization (Lange and Fieg, 2020; A. Courtais et al., 2020), or topology optimization (Rosinha et al., 2015) approaches can be applied to structurally optimize or identify novel packings. In the following, a method for the generative and hence algorithm-based design of structured packings is presented. Using a topology optimization approach, a multi-objective genetic algorithm is applied to develop novel structured packings with minimized pressure drop and maximized surface area. Rigorous CFD simulations are conducted to evaluate the key performance parameters of the packings and determine the values of the objectives. To present the design method, the optimization problem is defined and the topology optimization approach is presented in the next chapter. Then, first results of an optimization study are shown in chapter 3.

2. Generative Packing Design

For the design of structured packings several objectives and constraints need to be considered to meet the specific demands of the separation task as well as the additive manufacturing process. The minimization of the pressure drop across the packing as well as the maximization of the separation efficiency are essential objectives in packing design, especially in the case of vacuum distillation applications. This results in a complex constrained multi-objective optimization problem. To solve this task and to targeted identify novel structured packing shapes, a topology optimization algorithm and a rigorous analysis model for the evaluation of the values of the objective functions are needed. In the following the optimization problem is formulated before focussing on the optimization tool and the parametrisation of an exemplary topology optimization study.

2.1. Optimization Problem Formulation

For the formulation of the optimization problem the design variables and the design space need to be defined first, before focussing on the objective functions. The design space is defined by the given separation column diameter and the defined height of a single packing element. The resulting volume is discretized by a finite number of volume elements, where each element can either act as a packing material element or an empty element. The entirety of the material elements builds the structured packing, whereas fluid can flow through the empty elements. This behaviour is described by a binary code, defining a material element as a one and a fluid element as a zero. Then, each packing can be described by a matrix E of binary elements, where n describes the number of binary elements e_{ijk} in the direction of the cartesian coordinates of x, y and z. In turn, the matrix E can be reduced to a vector x_E.

$$E: (1, \dots, n_x) \times (1, \dots, n_y) \times (1, \dots, n_z) \rightarrow S \quad (i, j, k) \mapsto e_{ijk} \quad e_{ijk} \in [0,1] \tag{1}$$

$$x_{E_{i+(j-1)\cdot n_y+(k-1)\cdot n_x \cdot n_y}} = e_{ijk} \tag{2}$$

The resulting matrix E would describe a packing element with a squared cross-sectional area. Due to the fact that separation columns have a circular cross-sectional area, the coordinates of the packing are projected on the unit circle.

$$x_{Circ} = x \cdot \sqrt{1 - \frac{y^2}{2}}, \qquad y_{Circ} = y \cdot \sqrt{1 - \frac{x^2}{2}} \tag{3}$$

Based on the vector x_E of binary elements as design variables, the objective function is defined. Besides a minimum pressure drop across the packing, the separation efficiency should be maximized. Assuming perfect wetting of the packing surface, perfect mixing of the present phases, an infinitesimal small liquid film thickness on the packing and no dead volumes, the geometric surface area is equal to the interfacial mass transfer area. In this case, the separation efficiency of the packing is a function of the geometric surface area. Also, in the assumed case the multiphase system can be simplified as a first approximation to a single-phase vapour system. Accordingly, the objective is the minimization of the dry pressure drop $\Delta p_{P,d}$ and the reciprocal of the geometric surface area A_P of the packing.

$$f(x_{E,Opt}) = \min_{x_E} \left(\Delta p_{P,d}, \frac{1}{A_P} \right) \tag{4}$$

The geometric surface area A_P of the packing is defined as the sum of all faces of the material elements. These faces are either aligned parallelly to the main flow direction of the vapour phase or orthogonally directed to the main flow. It can be assumed, that the impact of the latter on the pressure drop is higher than the former. To consider the effect of the different face orientation of every single material element on the dry pressure drop, the objective function is complemented by a weighting factor β. This constant can take values between 0 and 1. The relevant surface area is then defined as the sum of the parallel $A_{P,par}$ and orthogonal $A_{P,orth}$ surface area weighted by β and one minus β. Accordingly, the objective function reads as follows.

$$f(x_{E,Opt}) = \min_{x_E} \left(\Delta p_{P,d}, \frac{1}{\beta \cdot A_{P,par} + (1-\beta) \cdot A_{P,orth}} \right); \qquad 0 \leq \beta \leq 1 \tag{5}$$

After defining the design space and the objective function of the optimization problem, the basic details of the optimization tool will be presented in the following.

2.2. Optimization Tool

The optimization tool is a combination of a developed modular control program implemented in MATLAB, using the PlatEMO platform (Tian et al., 2017) and OpenFOAM. The coupling between the optimization platform and the analysis model implemented in OpenFOAM is done by the tessellated packing data in STL-data format via the control program. The stochastic multi-objective genetic Strength Pareto

Evolutionary Algorithm 2 (SPEA2) is used as optimization algorithm, which was developed by Zitzler et al. (2001). As a genetic algorithm it mimics the evolutional theory by evaluating a fitness function and using the operators mutation, crossover and selection to generate a new generation of individuals. Also, process and manufacturing related constraints are considered. Every material element needs to be connected to another material element or to the boundaries of the design space, because unconnected material cannot be manufactured. Also, a minimum number of free flow channels needs to be available, because otherwise the mass transfer would be near zero. If these constraints are violated, the individual of the generation is punished.

2.3. Analysis Model

To predict the stationary flow behaviour of air flowing through the packing and to evaluate the resulting pressure drop over the created packings, rigorous CFD simulations are necessary. A laminar single-phase air flow at ambient conditions passes the packing from the lower to the upper end. Simplified Navier-Stokes equations in a finite-volume method implementation are solved using the SIMPLE-algorithm. The fluid is considered as incompressible and isothermal. No-slip condition at the surface of the packing and the column wall are assumed. A constant velocity at the inlet of the packing which correlates to a certain gas load or F-factor is defined and at the outlet a constant pressure is assumed.

The pressure drop is evaluated over the inlet and outlet of the packing as an area-averaged mean value. The total surface area of the packing as well as the surface area parallel and orthogonal to the main flow direction are also determined. OpenFOAM is used to conduct pre- and post-processing and for solving the model equations. Due to the high number of necessary CFD simulations for each evaluated generation, the simulations are performed at the North-German Supercomputing Alliance (HLRN).

2.4. Topology Optimization Study

In an exemplary optimization study, a packing for a lab scale distillation column should be generatively designed by the presented topology optimization approach. The column diameter is 54 mm and the height of a single packing element is 100 mm. The edge length of the quadratic material elements is chosen to 2.3 mm, assuming 23 elements in x- and y-direction and 43 material elements in z-direction. A preferred porosity of 0.8 for the randomly created initial generation is defined. The population size is 300 individuals per generation and the weighting factor β is 1/3. All CFD simulations are conducted at an F-factor at the inlet of the packing of 0.5 $\sqrt{\text{Pa}}$. In the following chapter, the results of the topology optimization study are presented.

3. Results

Selected results of the topology optimization study are shown in Figure 1 by plotting the specific pressure drop against the specific surface area of the evaluated packings for the three selected generations 0, 9, and 19. Having a look at the initial generation, the diversity of the individuals is quite low, due to the defined preferred packing porosity. Up to generation 20, the range of specific surface areas of the individuals rises up to a range of 370 m^2/m^3 to almost 420 m^2/m^3. Hence, the diversity also rises significantly. From generation to generation, the front is moving constantly into the lower right direction

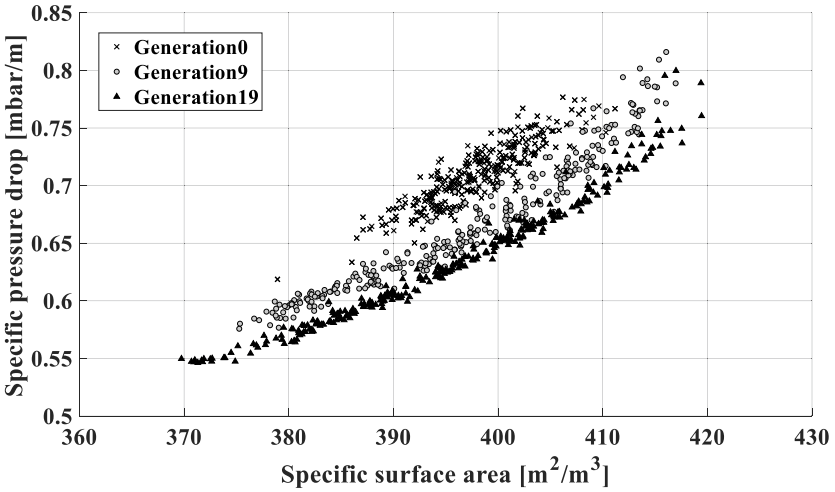


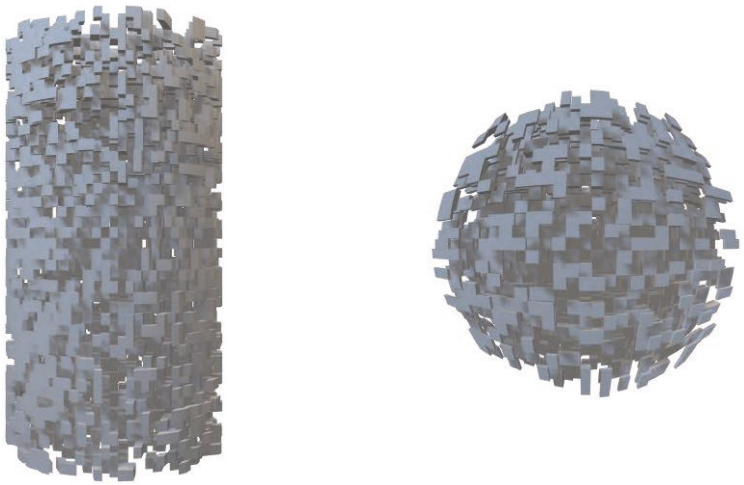
Figure 1: Specific dry pressure drop plotted against the specific surface area of the evaluated packings of the selected generations 0, 9 and 19

showing a good convergence behaviour. The best compromise individual regarding dry pressure drop and specific surface area reveals a specific dry pressure drop of 0.546 mbar/m and a specific surface area of 371.3 m^2/m^3. The appearance of this packing is shown in Figure 2 as STL-data in the side and top view. It becomes clear, that as expected, the packing consists of quite large material elements. Repetitive or regular parts are not visible yet. Hence, the optimization run needs to be continued. However, it can be stated that a packing element for the lab scale distillation column is successfully generatively designed, algorithm-based without any starting geometry.

In further studies, the edge length of the material elements will be lowered. Also, optimization studies for the development of tailored packings for specific separation tasks depending on the assumed boundary conditions and the selection procedure of the packings will be performed.

4. Conclusions

In this paper, an innovative method for the generative design of structured packings as a tool for the systematic identification of novel structured packing shapes was presented. The method is based on a topology optimization approach using the multi-objective genetic algorithm SPEA2 in combination with CFD simulations for the evaluation of the packings. A first optimization study for the development of a packing element for a lab scale distillation column was presented. First results show that it is possible to successfully design structured packings generatively without a well-performing packing as a starting point.



(a) (b)

Figure 2: STL-data of the packing element which shows the best compromise between pressure drop and surface area of generation 19 shown in (a) side view and (b) top view

Acknowledgement

The authors gratefully acknowledge the support from the North-German Supercomputing Alliance (HLRN).

References

A. Lange, G. Fieg, 2020, Systematische Entwicklung von formoptimierten additiv gefertigten strukturierten Packungen, Chemie Ingenieur Technik 92, No. 9, p. 1299

A. Lange, G. Fieg, 2021, Systematische Entwicklung neuartiger strukturierter Packungen – Generatives Design neuartiger additiv fertigbarer strukturierter Packungen für thermische Trennapparate, North-German Supercomputing Alliance (HLRN), Project ID hhi00035

A. Courtais, F. Lesage, Y. Privat, C. Pelaingre, A. Latifi, 2020, Shape Optimization of a Fixed-bed Reactor using Additive Manufacturing, Computer Aided Chemical Engineering 48, p. 1939-1944

J. Neukäufer, F. Hanusch, M. Kutscherauer, S. Rehfeldt, H. Klein, T. Grützner, 2019, Methodology for the Development of Additively Manufactured Packings in Thermal Separation Technology, Chemical Engineering & Technology 42, No. 9, p. 1970-1977

I. P. Rosinha, K. V. Gernaey, J. M. Woodley, U. Krühne, 2015, Topology optimization for biocatalytic microreactor configurations, Computer Aided Chemical Engineering 37, p. 1463-1468

L. Spiegel, M. Duss, 2014, Structured Packings, p. 146-151 in Distillation: Equipment and Processes (edited by A. Gorak, Z. Olujic), Academic Press

Y. Tian, R. Cheng, X. Zhang, Y. Jin, 2017, PlatEMO: A MATLAB Platform for Evolutionary Multi-Objective Optimization, IEEE Computational Intelligence Magazine 12, No. 4, p. 73-87

E. Zitzler, M. Laumanns, L. Thiele, 2001, SPEA2: Improving the strength pareto evolutionary algorithm, TIK-Report 103, Zurich, Switzerland

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50216-5

Multi-Objective Bayesian Optimization for Design and Operating of Fluidized Bed Reactor

Minsu Kim[a], Sunghyun Cho[b], Areum Han[c], Yoojin Han[d], Joseph Sang-Il Kwon[e], Jonggeol Na[c,*], Il Moon[a,*]

[a]*Department of Chemical and Biomolecular Engineering, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, 03722, Republic of Korea*

[b]*Harvard John A. Paulson School of Engineering & Applied Sciences, Science and Engineering Complex 3.401, 150 Western Avenue, United States*
[c]*Department of Chemical Engineering and Materials Science, Ewha Womans University, Seoul 03760, Republic of Korea*

[d]*Graduate Institute of Ferrous & Energy Materials Technology, Postech, 77 Cheongam-Ro, Nam-Gu, Pohang, 37673, Republic of Korea*

[e]*Texas A&M Energy Institute, Texas A&M University, College Station, TX 77845, USA*

Abstract

Treating the explosive waste using fluidized bed reactor, the design and operating conditions have highly effect to emission of pollutants (e.g. nitrogen oxides). Although it is possible to reduce the amount of pollutants through additional unit processes and extreme design and operating conditions, there are many practical difficulties because it causes an increase in cost. In addition, because of the explosive properties of waste, designing the process through actual experiments has many risks. Therefore, computational fluid dynamics (CFD) is used to simulate the reactor with high accuracy and to observe the internal temperature characteristics. While CFD shows high accuracy, it is difficult to obtain sufficient data for optimization because it requires a long computation time. Bayesian optimization repeats surrogate model optimization and infill criteria optimization and adaptively constructs the surrogate model. It shows good performance for time-consuming or expensive experiments. This study is to identify the design and operating conditions that minimize nitrogen oxides and cost through multi-objective Bayesian optimization. Multi-objective optimization problems generally do not have a single global optimization solution, but multiple solutions. This set of solutions forms a pareto front, which derives various solutions and gives decision-makers many options.

Keywords: explosive waste, nitrogen oxides, Bayesian optimization, multi-objective

1. Introduction

Explosive waste is treated via fluidized bed reactors. In this process, major air pollutant(e.g. nitrogen oxides, NOx) such as nitrogen oxides are generated. Taking environmental benefits through extremely high costs may present impractical solution. The fluidized bed reactor has the advantage of being easy to control the temperature of the reactor, so it was judged to be suitable for the incineration process of explosive waste. In addition, a hot spot with a rapid increase in temperature causes a significant

amount of nitrogen oxide to occur instantaneously, and computational fluid dynamics was used to simulate this phenomenon with high accuracy. It aims to minimize the emission of nitrogen oxides and the cost of the process by optimizing the design of the reactor as well as the operating conditions of the incineration process. In the case of changing the design of the reactor, it takes a lot of time to set up the mesh of the system a new, so there are many practical difficulties in optimization. The surrogate model is widely used in the engineering field as a method of simulating the underlying function inherent in an actual system with high accuracy through data. However, the surrogate model, whose accuracy is determined by the quality and quantity of data, has many difficulties in the data collection stage when sufficient data are not secured. Therefore, the importance of the design of experiment is further emphasized in this case. Adaptive sampling is a suitable method when it is difficult to obtain high-quality data due to cost and time issues. Bayesian optimization is a method that simultaneously performs adaptive design of experiment and optimization. In the case of multi-objective optimization that minimizes NOx emission and cost at the same time, the next experiment point is allocated through optimization of filling criteria, and a pareto front composed of a set of points that maximizes (or minimizes) the value of the Gaussian process-based surrogate model is drawn. Through this, several optimal solutions can be obtained, and it can help a lot in decision-making in real-world problem.

2. Mathematical model : Fluidized Bed Reactor

2.1. Governing equation

The mass and momentum transport phenomenon is based on Eq. (1) (mass transport) and Eq. (2) (momentum transport).

$$\frac{\partial \theta_f \rho_f}{\partial t} + \nabla \cdot \left(\theta_f \rho_f \boldsymbol{u}_f\right) = \delta \dot{m}_p \tag{1}$$

$$\frac{\partial \left(\theta_f \rho_f \boldsymbol{u}_f\right)}{\partial t} + \nabla \cdot \left(\theta_f \rho_f \boldsymbol{u}_f \boldsymbol{u}_f\right) = -\nabla p + \boldsymbol{F} + \theta_f \rho_f \boldsymbol{g} + \nabla \cdot \left(\theta_f \boldsymbol{\tau}_f\right) \tag{2}$$

In case of Newtonian fluid, the stress is calculated by Eq. (3) and the acceleration of the particle is calculated by Eq. (4).

$$\tau_{f,ij} = 2\mu S_{ij} - \frac{2}{3}\mu \delta_{ij} \frac{\partial u_k}{\partial x_k} \tag{3}$$

$$\frac{d\boldsymbol{u}_p}{dt} = D_p\left(\boldsymbol{u}_f - \boldsymbol{u}_p\right) - \frac{1}{\rho_p}\nabla p + \boldsymbol{g} - \frac{1}{\theta_p \rho_p}\nabla \tau_p + \frac{\bar{\boldsymbol{u}}_p - \boldsymbol{u}_p}{\tau_D} \tag{4}$$

The drag function D_p is expressed by Eq. (6) and it is calculated using drag coefficient C_d(Eq. (7))and the Reynold number Re.

$$D_p = \frac{3}{8} C_d \frac{\rho_f \left|\boldsymbol{u}_f - \boldsymbol{u}_p\right|}{\rho_p r_p}, \qquad Re = \frac{2\rho_f r_p \left|\boldsymbol{u}_f - \boldsymbol{u}_p\right|}{\mu_f} \tag{5}$$

To calculate the drag coefficient, the Wen-Yu/Ergun blend model that combines the Wen-Yu model and the Ergun model is used and the specific information about these models is explained in (*C. Wen et al., S. Ergun et al.*). Several gas phase pollutants such

as NOx and CO occur through a complex mechanism consisting of 131 reactions and a total of 25 components(*N. E. Ermolin et al, S. Cho et al(2020).*).

2.2. Cost estimation

In order to adjust the temperature of the air flowing into the fluidized bed reactor, a compressor and a heater are additionally needed in front of the fluidized bed reactor. The cost according to the operating conditions of the heater and the compressor and the gas mixture (oxygen/nitrogen) was considered. Each compressor is an isentropic type, with isentropic efficiency of 0.72 and mechanical efficiency of 1.0. It operates on electricity and is 0.0775 $/kWh. The heater is a fired heater operated at a cost of 0.0153$/kWh. The cost is calculated via Aspen Plus and the specific information about cost estimation is described in (*S. Cho et al(2021).*)

3. Methodology : Multi-objective Bayesian optimization

3.1. Gaussian process

The Gaussian Process(GP) is a Bayesian regression approach that predicts a system value by quantifying uncertainty of data. Let $x_n = \{x_i | i = 1, 2, \cdots, n\}$ be the input of the training set, and $y_n = \{y_i | i = 1, 2, \cdots, n\}$ the output of the training set. If we do not know the Bayesian statistics of a certain system, it is assumed that x_n, and y_n are randomly selected from a specific prior probability distribution. GP regression is a non-parametric approach that assumes the prior probability distribution as a multivariate normal distribution (*N*), and the mean (μ) and variance (=*Kernel, K*) are expressed as functions. Therefore, it is possible to calculate the distribution $p(\mathcal{F})$ of the function $\mathcal{F}$ through GP. A detailed description of GP is described in (*Rasmussen and Williams*).

$$p(\mathcal{F}) = N(\mu(x), K(x, x')) \tag{8}$$

3.2 *Multi-objective Bayesian optimization : q-Expected Hypervolume Improvement*

Multi-objective optimization does not have only one global optimum but forms a set of several global optimums, which is called a pareto front. Bayesian optimization consists of two stages, and it proceeds while optimizing a certain function in each stage. First, a surrogate model is generated by quantifying uncertainty through the GP model. The maximum value of the objective function is implemented through GP optimization.

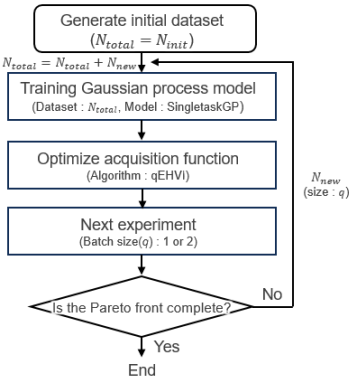


Figure 1. Multi-objective Bayesian optimization loop

The key to Bayesian optimization is to know the next design of experiment. The Nadir point (=Reference point) gives information about the range of objectives in the Pareto front stage in multi-objective optimization (Fig.2.(b)). These are the worst objective values determined by the decision maker, and since it is difficult to determine compared to the ideal point (=Utopia), it must be decided by domain knowledge (*Kalyanmoy Deb et al*).

Optimize the infill criteria that determine the next design of experiment and receive feedback on the next experiment with the most information. *q*-Expected Hypervolume Improvement (*q*-EHVI) (*Samuel Daulton et al*) is shown in Fig. 2. Find the experiments that maximizes the hypervolume in the initial objective function scatter (Fig.2.(a)) constructed as shown in Fig.2.(c) and repeat the optimization loop as shown in Fig.1. to maximize the hypervolume. As a result, when a part of the pareto front appears as shown in Fig.2.(e), the loop is terminated.

The number of next experiments has a trade-off between the quality of the data and the overall research time and cost. As the number increases, the research process time decreases, but the quality of the data decreases, so more experiments may be required than the optimal design of experiment. On the other hand, if the number of experiments is reduced, the sampling time increases, but the cost can ultimately be reduced because the data quality is good. Therefore, one or two experiment(s) is preferred.

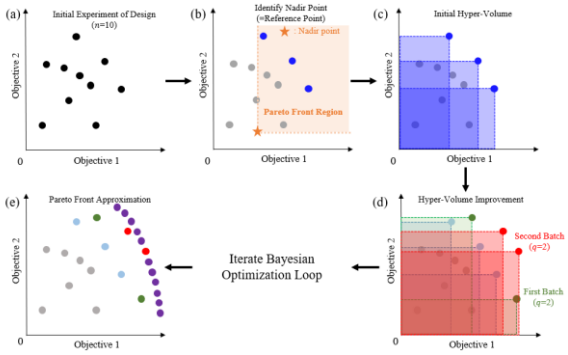


Figure 2. Sequence of *q*-EHVI algorithm and pareto front

4. Results

The reference point was selected based on domain knowledge among the factors that have a huge influence on the quality of the solution set identified through multi-objective Bayesian optimization. Based on Republic of Korea's air pollutant emission standard of 90 ppm, 100 ppm was selected with a margin of 10 ppm, and the cost standard was assumed to be 8,000,000 dollars. 100 initial data sets for GP model generation were selected, and the distribution of two objective values in the initial dataset is shown in Fig. 3. The reactor was a cylinder-type fluidized bed reactor, and 3 variables related to reactor design were selected and 7 variables related to operating were selected (Table 1.). In addition, although the objective is to minimize both NOx emission and cost, it has been transformed into a maximization problem by applying a negative sign for the efficient application of the optimization algorithm.

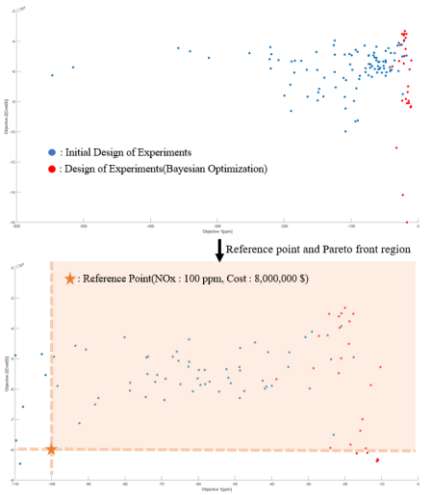


Figure 3. Reference point and Pareto front region

Through iteration of the Bayesian optimization loop, the maximum value of each objective function value farther from the reference point can be obtained compared to the initial dataset, and it shows a convex shape. Also, except for some points, a solution set is obtained within the preset reference point range.

Category	Parameter (Unit)	Lower bound	Upper bound
Operating parameter	Inlet gas velocity(m/s)	1	3.5
	Temperature(K)	400	800
	Pressure(bar)	2	4
	Particle radius(mm)	2	4
	Particle mass ratio	0.25	0.75
	Filled sand ratio	0.1	0.7
	O2 composition ratio in inlet gas	0.1	0.8
Design parameter	Fluidized bed diameter(m)	0.2	1
	Fluidized bed height(m)	1	5
	Sparger area/Fluidized bed cross sectional area(m^2)	0.1	4

Table 1. Operating and Design parameters

The objective function values for 40 dataset extracted through Bayesian optimization excluding the initial dataset are shown in Fig. 4. In order to make the range of the two objective functions the same, it was converted through a standard score (Z score). Most of the two objective function values show a trade-off relationship. However, in the 31st experiment (Red mark), it can be seen that both values are high. This can be seen as a stage of exploration to reduce the uncertainty of the surrogate model rather than performing exploitation that maximizes the values of the two objective functions. Also, for various solution sets, the particle radius is mostly calculated as the lower bound of 2 mm. That is, when minimizing both cost and NOx emission, the smaller the particles of explosive waste, the better.

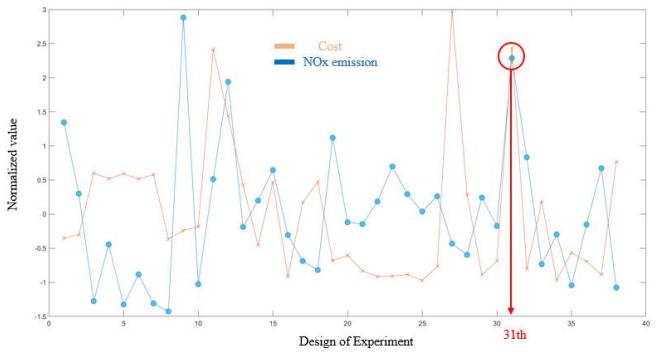


Figure 4. Trade-off between cost and NOx emission

5. Conclusions

Here, we optimize the amount of NOx emission and the cost of operation and design simultaneously to identify the practical reactor design and operating conditions. When the design of the reactor is changed, it is necessary to reconfigure and execute the mesh of the CFD structure, which is a highly time-consuming task. To overcome this problem, the GP model was adaptively improved through Bayesian optimization. This can be used very efficiently not only in CFD but also in actual experimental situations.

6. Acknowledgement

This research was supported by the MOTIE (Ministry of Trade, Industry, and Energy) in Korea, under the Fostering Global Talents for Innovative Growth Program (P0008747) supervised by the Korea Institute for Advancement of Technology (KIAT)

References

Robert Durichen, 2019, Multi-task Gaussian process models for biomedical applications, IEEE-EMBS International Conference on Biomedical and Health Informatics

Samuel Daulton, 2020, Differentiable Expected Hypervolume Improvement for Parallel Multi-Objective Bayesian Optimization, *NeurIPS 2020*

Kaifeng Yang, 2019, Multi-Objective Bayesian Global Optimization using expected hypervolume improvement gradient, *Swarm and Evolutionary Computation*, 44, 945-956

Eric Brochu, 2010, A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning, *arXiv*

C. Wen et al., 1966, Mechanics of fluidization, *Chem. Eng. Prog. Symp.*, 62, 100-111

S.Ergun et al., 1949, Fluid Flow through Randomly Packed Columns and Fluidized Beds, *Ind. Eng. Chem.*, 41, 1179-1184

S. Cho et al., 2020, Finding the best operating condition in a novel process for explosive waste incineration using fluidized bed reactors, *Comput. Chem. Eng.*, 142, 107054

Carl E. Rasmussen and Christopher Williams, Gaussian Processes for Machine Learning. *MIT press*, 2006.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50217-7

Analysis of Optimization Algorithms for Real-Time Optimization Applied on the Model of a Fluid Catalytic Cracking Unit

Adriana L. Rodríguez[a], Carlos A. M. Riascos[a]*

[a]*Univesidad Nacional de Colombia, Chemical and Environmental Department, Group of Process System Engineerig, Avenida Carrera 30, 45-03, Bogotá 111321, Colombia. camartinezri@unal.edu.co*

Abstract

On-line optimization of Fluid Catalytic Cracking Units (FCCU) is essential in modern refineries, because of high commercial value of cracking products. Within Real-Time Optimization (RTO) the on-line optimization module plays a fundamental role since operational conditions must be frequently adjusted. The performance of this module must guarantee good solutions in a sufficiently short time to allow on-line implementation.

This work presents an assessment of some deterministic and stochastic optimization algorithms into a three-layer RTO strategy for a virtual FCCU. The implementation begins with definition of an economic objective function, a sensitivity analysis to evaluate how profit is affected by variations on operational variables and concludes with a performance comparison of some algorithms.

Results from deterministic algorithms confirm that there are multiple local solutions in FCCU optimal operation problem, and deterministic search methods could be trapped in these solutions. On the other hand, GA allowed to obtain satisfactory solution, but with the default values for stopping parameters, it doesn´t generate the solution in the available time for an RTO implementation. As alternative to obtain good solution in a reasonable time, some GA parameters were adjusted and a hybrid strategy was assessed. With these approaches, execution time was satisfactorily reduced, and the hybrid strategy was confirmed as an interesting option for optimization modules in on-line applications.

Keywords: On-line Optimization; Optimization Methods Assessing; GA Parameter Adjusting; Hybrid Algorithms.

1. Introduction

Optimization has received important attention as a tool for increasing competition, reducing production costs, improving quality, and satisfying environmental requirements and regulations. The use of optimization has been extended in the engineering of process systems since it is not easy to find an optimal solution that leads to more profitable and competitive operating conditions (Biegler and Grossmann, 2004).

Optimization in process design and equipment specification is usually performed prior to the implementation of the process (off-line). On the other hand, optimization of operating conditions is carried out monthly, weekly, daily, hourly, or even, in the extreme, every minute. In that way, on-line optimization makes decisions in different levels of the organization for improving the performance of industrial processes during

their execution. RTO is an on-line optimization executed at an intermediate level, between the production scheduling and the control system (Rodríguez, 2019). While scheduling defines detailed programming to fulfill the strategic planning, the RTO defines the operational conditions that provide the highest performance of the process, and the control system must adjust the operation of the process to these optimal conditions.

Since disturbances affect the process performance, RTO develops a continuous reevaluation of any alteration on operating conditions, so that the economic productivity of the process is maximized. RTO cycle starts as soon as a steady state is detected after an external disturbance or a control action and, because of new disturbances could affect the process at any time, this cycle needs to be concluded as fast as possible.

In oil and gas industry, Fluid Catalytic Cracking (FCC) optimization has led to a significant increase in the performance. In FCCU applications, the on-line optimization problem usually comprises the maximization of the operational profit subject to a set of constraints and a non-linear programming problem (NLP), with additional complexity generated by a strong interaction between riser and regenerator (Fernández et al., 2007). Therefore, FCCU is a good candidate for RTO applications. This work presents a brief analysis of some optimization algorithms as candidates for implementing the optimization module into an RTO.

2. Methodology

This study was developed on the MATLAB platform, taking advantage of its toolboxes (MathWorks, 2018), optimization runs were developed on an Intel® Xeon® CPU E3-1241 v3 @ 3.5GHz. The FCCU model is presented with detail in Rodriguez (2019). A sensitivity analysis allowed selecting the optimization variables, as well as their limits and an approximate value for the objective function. The objective function in Eq.(1) is the FCCU profit (US$/s), it includes prices for products, feed and utilities used in the unit for the gas oil conversion. Additionally, a penalty factor considering the Lyapunov stability criterion (Torgashov et al., 2003) was added to punish unstable points, allowing to discard them, as it is shown in Eq.(2).

$$\begin{aligned} max \quad & FCCU_{prof} = f(F_{air}, F_{cat}, F_{feed,} T_{feed}, T_{air}) \\ s.t. \quad & \begin{cases} FCCU\ Model\ equations \\ upper\ and\ lower\ bounds \end{cases} \end{aligned} \tag{1}$$

$$\text{where} \quad FCCU_{prof} = \left(\sum P_{products} - \sum P_{feedstock} - \sum P_{utilities}\right) * penalty \tag{2}$$

As candidates, two deterministic algorithms (SQP and Interior Point), a metaheuristic algorithm (Genetic Algorithm) and a Hybrid strategy were analyzed. These algorithms were selected because of their proven capacity to solve NLP problems, and their availability in MATLAB, these and other analysis on the RTO modules implementation are presented in our complementary works (Reyes et al., 2015; Rodríguez, 2019).

3. Results and analysis

Results from a previous work (Rodríguez, 2019) revealed that product value as a function of its quality (octane number) is fundamental for the estimation of the optimal operational conditions, and the significance order of the variables for this purpose is:

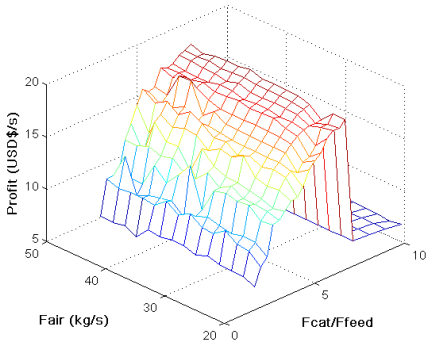


Figure 1. Effect of F_{air} and F_{cat}/F_{feed} ratio on profit.

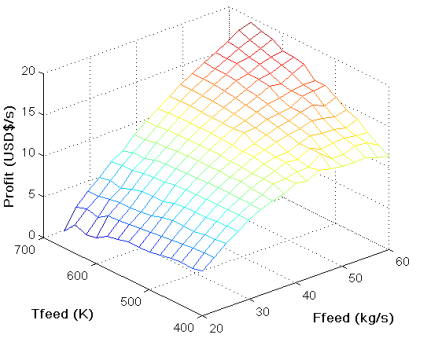


Figure 2. Effect of F_{feed} and T_{feed} on profit.

ratio catalyst flow/feed flow (F_{cat}/F_{feed}) > feed flow (F_{feed}) > feed temperature (T_{feed}) > air flow (F_{air}) > air temperature (T_{air}).

3.1. Sensitivity analysis

From this analysis, optimization variables and their limits were stablished, some key effects of optimization variables on objective function are shown in Figures 1 and 2. The effect of F_{cat}/F_{feed} on FCCU profit is the strongest, when this ratio goes from 2 to 6.8 a high increase on profit is observed; however, beyond 6.8 the profit drastically falls. This is mainly due to the fact that the increment in the F_{cat}/F_{feed} ratio generates a continuous increase in temperature at the riser outlet, a high riser temperature favours cracking but an over cracking diminishes the gasoline fraction (results not shown).

Effects of T_{feed} and F_{feed} on profit are important too and the variables F_{air} and T_{air} have smaller, but not insignificant, effect on profit. These analyses show the high non-linearity of the FCCU system, which generates local optima as observed in next results.

3.2. Assessment of optimization algorithms

In a comparative assessment of some optimization algorithms, it was observed that Genetic Algorithm (GA) allows to obtain satisfactory operational conditions, but it does not generate the solution in the available time for an RTO implementation. In contrast, the assessed deterministic algorithms (Interior Point, and SQP) generated solutions in short calculation time but with worse objective function values (Table 1).

Table 1. Performance of standard optimization algorithms

	Lower bound	Upper bound	Initial Point	Method		
				Int. Point	SQP	G A
F_{cat} (kg/s)	115.0	390.0	294.0	290.7	294.0	385.3
F_{feed} (kg/s)	20.0	60.0	40.6	58.8	40.6	59.6
F_{air} (kg/s)	20.0	46.0	25.4	45.9	25.4	37.6
T_{feed} (K)	400.0	670.0	434.6	448.5	434.7	656.8
T_{air} (K)	310.0	525.0	360.0	373.8	360.0	360.0
T_{ris} (K)	755.0	840.0	779.2	753.8	779.2	833.5
T_{rgn} (K)	810.0	1090.0	972.0	1006.6	972.0	1001.3
Octane Number	85.0	100.0	91.7	89.0	91.7	97.3
Profit (US$/s)	-	-	9.7	12.1	9.9	19.9
Run time (s)	-	-	-	148	185	13,932

Results from each algorithm propose operational conditions with remarkable differences, e.g. conditions obtained by GA include high flow of catalyst, feed and air, and high feed temperature; in other way, IP suggests medium catalyst flow, high feed and air flow and low feed temperature, and SQP suggests medium catalyst and feed flow, low air flow and feed temperature. These results confirm the complexity and high non-linearity of this optimization problem which is a challenge for on-line applications.

Analysing the performance of GA with default parameters, the hypothesis that it is possible to obtain a satisfactory optimization for on-line implementation by adjusting the stopping parameters is formulated. A satisfactory optimization implies to reduce the convergence time without damaging the quality of the solution. Results obtained for several GA runs (Figure 3 shows two of these) were used to analyse the performance and the confidence of the GA. Evolution of Best and Mean fitness shows that the diversity of the population is highly reduced by the 10th generation, and after the 20th generation, no significant change in the profit function value was observed.

From this analysis, *Generation* and *Stall Generations* were selected as adjustable stopping parameters. GA stops when the total number of evaluated generations reaches the value of *Generations*, as on the right side in Figure 3 or when the improvement in the fitness over *Stall Generations* is less than a tolerance (MathWorks, 2018). Analysing these criteria and the evolution of the best and mean fitness, the number of stall generations was adjusted to 10 and the maximum of generations to 20, this generates a reduction from 3.9 to 0.5 hours in the average computation time.

Optimization results with the adjusted value for stopping parameters are presented in Figure 4 and Table 2. This figure shows that improvement in best fitness was very modest since 4th generation. Analysing values for operational variables from GA with default values (Table 1) and adjusted values (Table 2), it can be observed that both solutions propose similar operational conditions: high values (near or on the upper limit) for the flows of catalyst, feed and air, as well as high and medium values for temperatures of feed and air, respectively. Therefore, calculates values for riser and regenerator temperatures are similar (883.5 vs 829.4 and 1001.3 vs 996.3 K). These results suggest that a deterministic search starting on the best point identified by GA algorithm must be considered as an alternative to improve the whole optimization procedure.

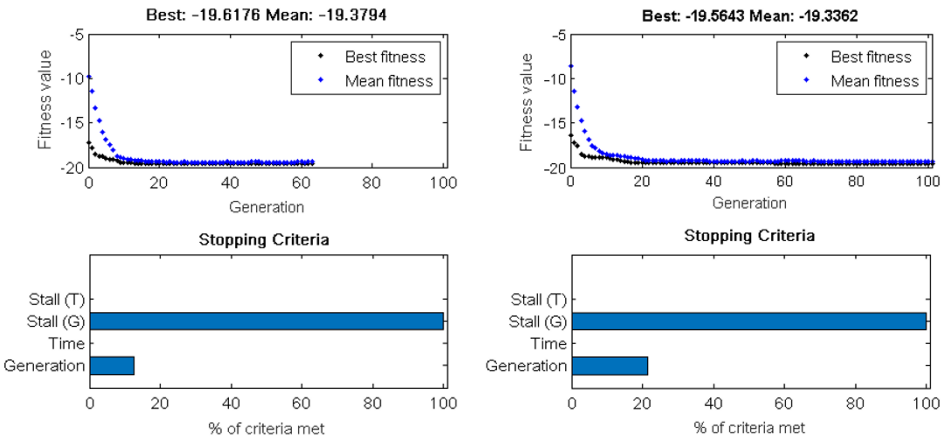


Figure 3. Convergence in two GA runs with default parameters.

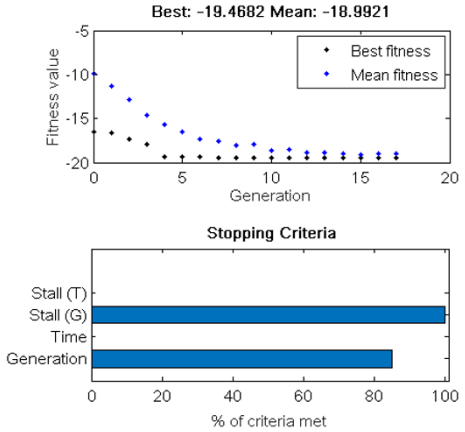


Figure 4. GA convergence with adjusted parameters.

Table 2. Solution from adjusted GA.

Variable	Value
F_{cat} (kg/s)	388.7
F_{feed} (kg/s)	60.0
F_{air} (kg/s)	43.6
T_{feed} (K)	651.7
T_{air} (K)	360.0
T_{ris} (K)	829.4
T_{rgn} (K)	996.3
Octane Number	97.3
Profit (US$/s)	19.8
Run time (s)	1,800

3.3. Hybrid optimization strategy

The hybrid strategy considers a first step with GA and a second one with a deterministic algorithm (Interior Point or SQP). Initial search with GA must allow identifying the region where the global solution is in, and final search with deterministic algorithm must allow to reach the optimal point. To confirm the viability and robustness of this strategy multiple hybrid runs were developed changing the number of total GA generations and the deterministic algorithm.

In general, operational conditions obtained from hybrid runs are similar to the ones previously obtained only by GA (section 3.2). To analyse the ability to identify the global solution, Figure 5 shows the profit obtained after the GA step and after the complete hybrid run, employing 10 GA generations and IP. Some important characteristics of these results are: i) only in 50% of runs deterministic step generates a significant improvement in profit, and ii) the performance of deterministic step does not depend on the results generated from GA: in run 3, instead of the relatively low profit from GA, IP didn´t improve the solution.

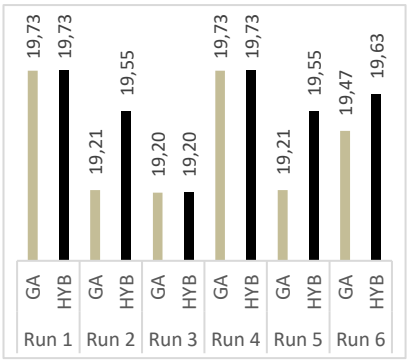


Figure 5. Profit obtained from hybrid strategy: 10 GA Gen + IP.

Table 3. Solutions from hybrid strategy.

Variable	Mean / MAPD (%)		
	10GA+IP	5GA+IP	5GA+SQP
F_{cat} (kg/s)	378.0 / 2.8	381.8 / 1.5	374.4 / 3.7
F_{feed} (kg/s)	59.3 / 0.9	58.4 / 1.2	57.4 / 2.8
F_{air} (kg/s)	40.8 / 5.6	39.8 / 7.2	37.9 / 6.8
T_{feed} (K)	659.5 / 1.2	651.3 / 2.1	651.7 / 1.2
T_{air} (K)	412.7 / 9.1	391.1 / 12.9	386.1 / 11.0
Profit (US$/s)	19.6 / 0.7	19.2 / 1.4	18.8 /2.7
AvImpPro	0.14	0.32	0.19

Table 3 presents average values and the mean absolute percentage deviations (MAPD) for some operational variables and the final profit from hybrid strategy with 10 and 5 GA generation and IP or SQP, as well as the average improvement in profit (AvImpPro) obtained in the deterministic step. These results show that this hybrid strategy doesn´t rigorously guarantee to reach the global optimum because of differences in operational variables values are small but not negligible. The average and the mean absolute percentage deviation (MAPD) show that characteristics of feed (flow and temperature) are the most stable, while characteristics of air have the biggest variation. Improvement in profit obtained with IP and SQP have not significant differences.

Results from hybrid strategy and GA with adjusted parameters show that both strategies could be good alternatives for on-line optimization, and inspirate new studies to analyse their advantages and the challenge to reach, with confidence, the global solution. Additional results, not showed, suggest that stopping criteria for GA must guarantee a minimum number of generation and that iteration time is not a good stopping criterion.

4. Conclusions

Optimization allows to identify optimal operational conditions for a FCCU which improves its profit, and the algorithm choice is key in the success of this search, especially for on-line implementations.

Results from studied algorithms propose operational conditions with remarkable differences and show the complexity and high non-linearity of the FCCU optimization problem, this non-linearity generates local solutions in which deterministic algorithm could stop.

Operational conditions proposed from GA with default and adjusted stopping parameter values confirm that execution time can be reduced in GA, when stopping parameters are adjusted. However, a minimum of generations must be guaranteed to do not reduce the quality of the solution.

Performance of hybrid strategy shows that it isn´t easy to reach the global convergence, and that new studies are necessary to generates reliable strategies.

References

L.T. Biegler, I.E. Grossmann, 2004, Retrospective on optimization, Comput. & Chem. Eng., 28, 8, 1169–1192.

J.L. Fernández, C.I.C. Pinheiro, N. Oliveira, A.I. Neto, F.R. Ribeiro, 2007, Multiplicity of steady states in an UOP FCC unit with high efficiency regenerator, Chemical Eng. Science, 62, 22, 6308–6322.

MathWorks, 2018, How the Genetic Algorithm Works, *Genetic Algorithm*, Available at: https://www.mathworks.com/help/gads/how-the-genetic-algorithm-works.html.

J.D. Reyes, A.L. Rodríguez, A.L., C.A.M. Riascos, 2015, Data Analysis and Modelling of a Fluid Catalytic Cracking Unit (FCCU) for an Implementation of Real Time Optimization, Computer Aided Chem. Eng., 37, 611-616.

A.L. Rodríguez, 2019, Implementation and performance assessment of a real-time optimization system on a virtual fluidized-bed catalytic-cracking plant, MSc Thesis, Universidad Nacional de Colombia.

A.Y.Torgashov, K.C. Park, H.C. Choi, Y.K. Choe, 2003, Stability analysis of distillation control using vector Lyapunov function, Computer Aided Chem. Eng., 15, 1058-1063.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50218-9

Scalable Stochastic Programming with Bayesian Hybrid Models

Jialu Wang[a], Elvis A. Eugene[a], Alexander W. Dowling[a*]

[a]*Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN 46556, USA*
adowling@nd.edu

Abstract

Bayesian hybrid models (BHMs) fuse physics-based insights with machine learning constructs to correct for systematic bias. In this paper, we demonstrate a scalable computational strategy to embed BHMs in an equation-oriented modelling environment. Thus, this paper generalizes stochastic programming, which traditionally focuses on aleatoric uncertainty (as characterized by a probability distribution for uncertainty model parameters) to also consider epistemic uncertainty, i.e., mode-form uncertainty or systematic bias as modelled by the Gaussian process in the BHM. As an illustrative example, we consider ballistic firing using a BHM that includes a simplified glass-box (i.e., equation-oriented) model that neglects air resistance and a Gaussian process model to account for systematic bias (i.e., epistemic or model-form uncertainty) induced from the model simplification. The gravity parameter and the GP hypermeters are inferred from data in a Bayesian framework, yielding a posterior distribution. A novel single-stage stochastic program formulation using the posterior samples and Gaussian quadrature rules is proposed to compute the optimal decisions (e.g., firing angle and velocity) that minimize the expected value of an objective (e.g., distance from a stationary target). PySMO is used to generate expressions for the GP prediction mean and uncertainty in Pyomo, enabling efficient optimization with gradient-based solvers such as Ipopt. A scaling study characterizes the solver time and number of iterations for up to 2,000 samples from the posterior.

Keywords: Hybrid model; Bayesian uncertainty quantification; Optimization; Gaussian process; Pyomo

1. Introduction

Predictive models play a key role in control and decision-making (Adjiman et al., 2021). While the glass-box models are constructed from scientific principles and have a deeper understanding of the underlying processes, they are often complex to form and solve. Many glass-box models contain unknown parameters that are inferred from experimental data. These data are often subject to random phenomena such as variability between experiments or observation noise (Kalyanaraman et al., 2015), which gives rise to aleatory (i.e., parametric) uncertainties. Stochastic programming and robust optimization are routinely used to directly incorporate parametric uncertainty into decision-making frameworks. However, to maintain computational tractability, glass-box models are often simplified or replaced with surrogate models in multiscale engineering frameworks (Biegler et al., 2014). The systematic bias from model inadequacy arising from such simplifications is often referred to as model-form or epistemic uncertainty (McClarren, 2018).

Bayesian hybrid models (BHM) offer a principled framework to quantify, propagate, and mitigate aleatoric and epistemic uncertainties by combining physical glass-box models with black-box surrogate models. In their seminal work, statisticians Kennedy and O'Hagan (2001) proposed a (Bayesian) hybrid modelling framework using Gaussian process models:

$$y = \eta(\boldsymbol{x}|\boldsymbol{\theta}) + \delta(\boldsymbol{x}|\boldsymbol{\phi}, \boldsymbol{D}) + \varepsilon \tag{1}$$

The prediction y consists of three components: the inadequate (simplified or reduced order) glass-box model $\eta(\boldsymbol{x}|\boldsymbol{\theta})$ which depends on the state variables $\boldsymbol{x}$ and model parameters $\boldsymbol{\theta}$; the Gaussian process discrepancy $\delta(\boldsymbol{x}|\boldsymbol{\phi}, \boldsymbol{D})$ which models epistemic uncertainty as a function of the state variables $\boldsymbol{x}$, hyperparameters $\boldsymbol{\phi}$, and data $\boldsymbol{D} = [\boldsymbol{x_{obs}}, \boldsymbol{y_{obs}}]$; and, finally, the observation error ε which is modeled as a random variable with known probability distribution. Unlike other hybrid model architectures, such as a neural differential equation, the probabilistic nature of the GP enables the use of Bayesian calibration (Higdon et al., 2004) to infer the model parameters and hyperparameters and provides readily interpretable uncertainty information. The joint posterior distribution of model parameters resulting from Bayesian model calibration informs the uncertainty in the models; specifically, the distribution of model parameters $\boldsymbol{\theta}$ and observation error ε quantifies aleatory uncertainty while the GP output quantifies epistemic uncertainty. We emphasize that prior applications of the Kennedy-O'Hagan framework in chemical engineering (Mebane et al., 2013, Kalyanaraman et al., 2015, Kalyanaraman et al., 2016, Bhat et al., 2017) predominately considers model calibration and uncertainty propagation and not decision-making under uncertainty.

2. Methods

2.1 Stochastic Programming Formulation

In this work, we develop and implement a single-stage stochastic program formulation in Pyomo (Hart et al., 2017) to optimize decisions using BHMs by minimizing the expected values of an arbitrary objective function $u(y)$ in the form of Eqs. (2a):

$$\min_{\boldsymbol{x}} \mathop{\mathrm{E}}_{\boldsymbol{\theta},\boldsymbol{\phi}}[u(y)] \approx \frac{1}{\sqrt{\pi}} \sum_{s \in S} \sum_{j \in J} w_s w_j u_{s,j} \tag{2a}$$

$$\delta_j = \mu(\boldsymbol{x}|\boldsymbol{\phi}, \boldsymbol{D}) + \sqrt{2}\, z_j\, \sigma(\boldsymbol{x}\,|\boldsymbol{\phi}, \boldsymbol{D}), \qquad \forall j \in J \tag{2b}$$

$$\eta_s = \eta(\boldsymbol{x}|\boldsymbol{\theta}_s), \quad \forall s \in S \tag{2c}$$

$$y_{s,j} = f\left(\eta_s, \delta_j\right), \forall s, j \in S \times J \tag{2d}$$

$$u_{s,j} = u(y_{s,j}), \forall s, j \in S \times J \tag{2e}$$

In Eq. (2a), the expectation E of $u(y)$ is approximated using scenario weights $w_s = 1/|S|$. Set S contains samples from the posterior distribution (trace) of $\boldsymbol{\theta}$. Set J contains Gauss-Hermite quadrature nodes z_j and weights w_j, which are used in Eq. (2b) to approximate the GP output distribution characterized by GP prediction mean μ and standard deviation σ. In Eq. (2c), the glass-box model is evaluated at samples θ_s. In Eqs.

(2d) and (2e), the BHM output y and objective function $u(y)$ are evaluated over the set $S \times J$. This formulation is computationally attractive because the highly nonlinear GP prediction mean and standard deviation are evaluated only once, while the glass-box model, and the objective function are evaluated $|S|$, and $|S| \times |J|$ times, respectively.

2.2 Ballistics Firing Example

We apply the stochastic program to the ballistics example from Eugene et al. (2020):

$$\eta_s = \frac{2v_0^2}{g_s} \cdot \sin\psi \cdot \cos\psi, \forall s \in S \tag{3a}$$

$$y_{s,j} = \eta_s + \delta_j, \forall s, j \in S \times J \tag{3b}$$

$$y_{s,j} - \bar{y} = u_{s,j}^+ - u_{s,j}^-,\ u_{s,j} = u_{s,j}^+ + u_{s,j}^-,\ u_{s,j}^+, u_{s,j}^- \geq 0, \forall s, j \in S \times J \tag{3c}$$

Eqs. (3a, b) describe the BHM. The state variables are $\boldsymbol{x} = [v_0, \psi]$, where v_0 (m/s) is the firing velocity and ψ (°) is the firing angle. The glass-box model has one uncertain parameter, the acceleration due to gravity g. y (m) is the distance to the impact location of the projectile measured horizontally from y=0 which is the firing location of the projectile. We seek to predict the optimum conditions $\boldsymbol{x}$ to hit a target a fixed distance $\bar{y} = 100$ m away, despite neglecting air resistance (epistemic uncertainty) in the glass-box model. The objective function $u(y) = |y - \bar{y}|$ is reformulated using slack variables in Eq. (3c) to provide differentiable constraints for gradient-based optimization. By combining Eqs. (2) and (3), the expectation of u is minimized to find the optimum $\boldsymbol{x} = [v_0, \psi]$ to hit the target.

The observed data $\boldsymbol{D} = [\boldsymbol{x_{obs}}, \boldsymbol{y_{obs}}]$ is generated from the true physical model which includes the effects of air-resistance on the projectile resulting in its non-parabolic trajectory as described by Eugene et al. (2020). Six data points corresponding to observations from experiments 1 to 5 and 6c in Table 1 of Eugene et al., 2020 were used for the *sequential Bayesian calibration* of the hybrid model. First, the glass-box model is calibrated using the data $\boldsymbol{D}$, a likelihood function, and priors (see Eugene et al. 2020 for details) via Hamiltonian Monte Carlo in PyMC3 (Salvatier et al., 2016) which returns a trace of 2,000 samples from the posterior distribution of the glass-box model parameter g. Next, using the mean value of g from the trace, $\bar{g}$, we compute the residuals $\boldsymbol{y_{obs}} - \eta(\boldsymbol{x_{obs}}, \bar{g})$ which represents the systematic bias in the model due to epistemic uncertainty. These residuals are used to train a discrepancy function for which we assume a Gaussian process with kriging kernel:

$$\boldsymbol{y_{obs}} - \eta(\boldsymbol{x_{obs}}, \bar{g}) = \delta(\boldsymbol{x}|\boldsymbol{\phi}, \boldsymbol{D}) \sim \mathcal{GP}(\mu, k(\cdot,\cdot)) \tag{4a}$$

$$k(x_j, x_k) = \sigma_f^2 \exp\left(-\sum_{i=1}^{n} \beta_i \left|x_{ij} - x_{ik}\right|^2\right),\ j, k \in 1, \ldots, m,\ \beta_i \geq 0 \tag{4b}$$

where μ and σ_f^2 are the prediction mean and the variance of the GP model, respectively and β_i is the kriging weight. n=2 denotes the number of input dimensions for $\boldsymbol{x} = [v_0, \psi]$, and m = 6 is the number of observations in the training data set for the GP. The GP model hyperparameters $\boldsymbol{\phi} = [\mu, \sigma_f^2, \beta_1, \beta_2] =$ [-0.68, 2.28, 1.53, 0.13] are trained using maximum likelihood estimation (MLE) (Forrester et al., 2008) implemented in the

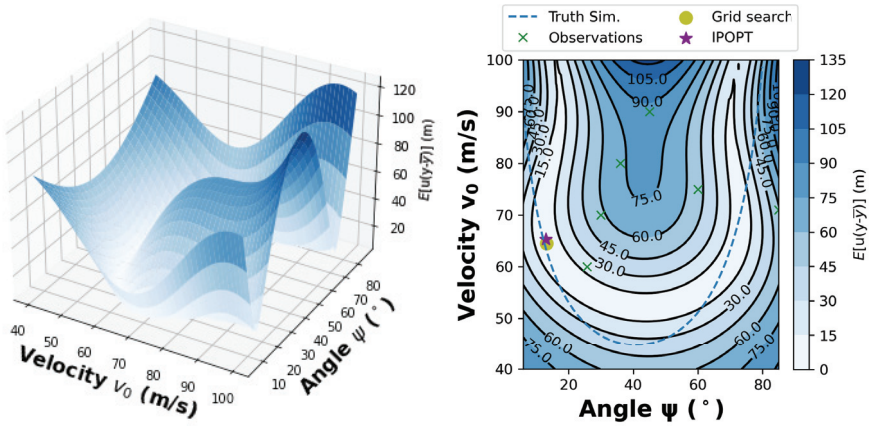


Fig.1. The expectation of the utility function changes with the velocity and the angle visualized in a 3D surface (left) and contour plot (right).

PySMO toolbox of IDAES (Lee et al., 2021). We do not explicitly model aleatory uncertainty due to observation noise in this study.

3. Results

The optimization problem in Eqs. (2, 3) is constructed in Pyomo and solved using IPOPT and the linear solver MA57 (HSL, 2007) using $|S|$ =1000 and $|J|$=10. To visualize the objective function surface and contours in Fig. 1, square instances of Eqs. (2,3), i.e., $\boldsymbol{x}$ is fixed such that there are no degrees of freedom, are evaluated using a grid of 2750 uniformly space samples over $v_0 \in [40, 100]$ (m/s) and $\psi \in [5.7, 85.5]$ (°). Fig. 1 (left) shows that firing at large angles (near vertical orientation) with large velocities results in undesirably high objective values with expected miss distances over 120 m. In contrast, firing at shallower angles (near horizontal orientation) and modest velocities results in expected miss distances less than 15 m. The dashed blue line in Fig. 1 (right) shows the combinations of angles and velocities that result in a direct hit based on simulations of the true physical model. The optimum calculated by the gradient-based solver Ipopt (purple star), and the grid search optimum (yellow dot), are nearly identical and both close to the direct hit line, indicating Bayesian hybrid model and stochastic programming formulation accurately account for epistemic uncertainty. The optimization problem is reliably solved by Ipopt in approximately 4 s on a MacBook with a 2.6 GHz Intel Core i7 CPU, finding an optimal solution as $v_0 = 65.34$ m/s, $\psi = 12.99$ ° and an objective function value of 5.07 m. The optimum found by the grid search is v_0 =64.49 m/s, ψ =13.1 °, and the optimized expectation value is 5.14 m. As expected, the gradient-based solver outperforms the grid search.

Next, we demonstrate the scalability of the proposed stochastic programming formulation. Fig. 2 shows the variation in CPU time and the number of Ipopt iterations as the number of scenarios $|S|$ increases. For each value of $|S|$, Eqs. (2, 3) are resolved 50 times using samples randomly drawn from the posterior distribution trace. For $|S| = 2000$, Eqs. (2, 3) are solved once using the entire trace. The dots represent the medians for both metrics, while the error bar represents the 75th and 25th percentiles. Fig. 2 (left) shows that as $|S|$ increases, the median CPU time increases from 0.03 s with 10 scenarios, to 4.8 s with

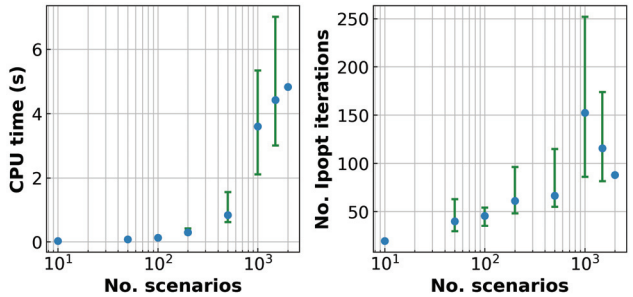


Fig.2. The CPU time and number of IPOPT iterations increase with the number of samples $|S|$. The dots represent the median and the error bars denote the 25%-ile and 75%-ile for 50 random samples except for $|S|$ = 2,000.

2000 scenarios. Likewise, Fig. 2 (right) shows median of the number of Ipopt iterations increases from 19 with 10 scenarios, achieving a peak at 152.5 with 1,000 scenarios, and decreases to 88 with 2,000 scenarios. We hypothesize the large variability in metrics for a given $|S|$ is due to the different scenario data considered for each of the 50 replicates. The problem considering 2,000 scenarios can be solved in less than 4.8 s and 88 iterations with 82,014 variables and 62,012 equality constraints, showing that this stochastic program can be reliably solved by gradient-based solvers with small computational burden. It is also noted that the optimum decisions are practically the same as $|S|$ increases. This highlights the potential to accommodate larger problems with multiple sources of model-form uncertainty, i.e., multiple GP discrepancy functions, by using only a modest number of posterior scenarios, e.g., $|S|$ is 10 or 50. Likewise, we hypothesize this formulation can scale to a large number of glass-box parameters $\boldsymbol{\theta}$ using a modest number of posterior samples and approaches such as sample average approximation. Systematic study of these scaling properties is left as future work.

4. Conclusions

In this paper, we demonstrate a scalable stochastic programming formulation for optimization under both aleatoric (i.e., parametric) and epistemic (i.e., model-form) uncertainties using Bayesian hybrid models (BHMs). By leveraging both Gaussian quadrature rules and PySMO, we demonstrate efficient optimization using the equation-oriented Pyomo modeling environment and gradient-based Ipopt nonlinear programming solver. Through an illustrative ballistics example, we show the Kennedy-O'Hagan inspired BHMs effectively capture epistemic uncertainty; their predictions are consistent with the full-physics true model. Moreover, accounting for epistemic uncertainty in the stochastic programming formulation may be accomplished using as little as 10 samples. As future work, we plan to explore the application of optimization-supported decision-making with hybrid models for diverse applications in process system engineering and adjacent domains. We are especially interested in using BHMs to account for information-loss from model simplified in molecular-to-systems engineering frameworks.

Acknowledgement

This project was funded by NSF CBET Award CBET-1941596. E. A. E. gratefully acknowledges partial support from the Patrick and Jana Eiler's Graduate Fellowship. J.W. gratefully acknowledges partial support from the Carbon Capture Simulation for Industry

Impact ($CCSI^2$), funded through the U.S. DOE office of Fossil Energy by the Lawrence Berkeley National Laboratory through contract #DE-AC02-05CH11231.

References

C.S. Adjiman, N.V. Sahinidis, D.G. Vlachos, B. Bakshi, C.T. Maravelias and C. Georgakis, 2021, Process systems engineering perspective on the design of materials and molecules, Ind. Eng. Chem. Res, 60, 14, 5194-5206, 10.1021/acs.iecr.0c05399

K.S. Bhat, D.S. Mebane, P. Mahapatra and C.B. Storlie, 2017, Upscaling uncertainty with dynamic discrepancy for a multi-scale carbon capture system, Journal of the American Statistical Association, 112, 520, 1453-1467,

L.T. Biegler, Y.-d. Lang and W. Lin, 2014, Multi-scale optimization for process systems engineering, Computers & chemical engineering, 60, 17-30, 10.1016/j.compchemeng.2013.07.009

E.A. Eugene, X. Gao and A.W. Dowling, 2020, Learning and optimization with bayesian hybrid models, 2020 American Control Conference (ACC), 3997-4002, 10.23919/ACC45564.2020.9148007

A. Forrester, A. Sobester and A. Keane, 2008, Engineering design via surrogate modelling: A practical guide, John Wiley & Sons,

W.E. Hart, C.D. Laird, J.-P. Watson, D.L. Woodruff, G.A. Hackebeil, B.L. Nicholson and J.D. Siirola, 2017, Pyomo-optimization modeling in python, Springer, 67,

D. Higdon, M. Kennedy, J.C. Cavendish, J.A. Cafeo and R.D. Ryne, 2004, Combining field data and computer simulations for calibration and prediction, SIAM journal on scientific computing, 26, 2, 448-466, 10.1137/s1064827503426693

HSL, 2007, Collection of fortran codes for large-scale scientific computation, http://www. hsl. rl. ac. uk,

J. Kalyanaraman, Y. Fan, Y. Labreche, R.P. Lively, Y. Kawajiri and M.J. Realff, 2015, Bayesian estimation of parametric uncertainties, quantification and reduction using optimal design of experiments for CO_2 adsorption on amine sorbents, Computers & chemical engineering, 81, 376, 10.1016/j.compchemeng.2015.04.028

J. Kalyanaraman, Y. Kawajiri, R.P. Lively and M.J. Realff, 2016, Uncertainty quantification via bayesian inference using sequential monte carlo methods for CO_2 adsorption process, AIChE Journal, 62, 9, 3352-3368,

M.C. Kennedy and A. O'Hagan, 2001, Bayesian calibration of computer models, Journal of the Royal Statistical Society. Series B, Statistical methodology, 63, 3, 425-464, 10.1111/1467-9868.00294

A. Lee, J.H. Ghouse, J.C. Eslick, C.D. Laird, J.D. Siirola, M.A. Zamarripa, D. Gunter, J.H. Shinn, A.W. Dowling, D. Bhattacharyya, L.T. Biegler, A.P. Burgard and D.C. Miller, 2021, The idaes process modeling framework and model library—flexibility for process simulation and optimization, Journal of advanced manufacturing and processing, 3, 3, 10.1002/amp2.10095

R.G. McClarren, 2018, Uncertainty quantification and predictive computational science : A foundation for physical scientists and engineers, Cham : Springer International Publishing : Imprint: Springer,

D.S. Mebane, K.S. Bhat, J.D. Kress, D.J. Fauth, M.L. Gray, A. Lee and D.C. Miller, 2013, Bayesian calibration of thermodynamic models for the uptake of CO_2 in supported amine sorbents using ab initio priors, Physical Chemistry Chemical Physics, 15, 12, 4355-4366,

J. Salvatier, T.V. Wiecki and C. Fonnesbeck, 2016, Probabilistic programming in python using PyMC3, PeerJ. Computer science, 2, e55, 10.7717/peerj-cs.55

A. Wächter and L.T. Biegler, 2006, On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming, Mathematical programming, 106, 1, 25-57

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50219-0

A Combined Particle Swarm Optimization and Outer Approximation Optimization Strategy for the Optimal Design of Distillation Systems

Fanyi Duanmu, Dian Ning Chia, and Eva Sorensen*

Department of Chemical Engineering, University College London (UCL), Torrington Place, London WC1E 7JE, United Kingdom

**Corresponding author: e.sorensen@ucl.ac.uk*

Abstract

Distillation is by far the most important separation process in the chemical industries, but is also one of the most energy intensive units, thus significant efforts are being placed on optimizing the design and operation of both new and existing distillation systems to achieve potentially significant energy and/or capital savings. This work introduces an optimization strategy based on combining a preliminary stochastic optimization method, either Particle Swarm Optimization (PSO) or Genetic Algorithm (GA), which is used to perform a global search, with a deterministic optimization method, Outer Approximation method (OAERAP), for the local search. A dividing wall distillation column and a hybrid distillation/pervaporation superstructure is used to illustrate the methodology. The performance of stand-alone PSO, stand-alone GA, combined PSO/OAERAP and combined GA/OAERAP, are compared and it is shown that the PSO/OAERAP combination provides the optimal design in a significantly shorter time and can also optimize a superstructure efficiently. This work also presents the use of a parallel computing function when performing stochastic optimization, which significantly speeds up the optimization.

Keywords: Particle Swarm Optimization, Parallel Computing, Distillation, Dividing Wall Column, Hybrid Distillation

1. Introduction

The optimization of a distillation system is a highly non-convex Mixed Integer Nonlinear Programming (MINLP) problem due to the tight coupling between different parts of the mathematical model (thermodynamic equations, MESH equations, etc.) and the existence of discrete (number of stages, feed location, existence of a unit in the superstructure, etc.) and continuous variables (reflux ratio, heat input, flowrates, etc.). The solution is thus is a very challenging task, not only from a computational point of view, particularly for initialization, but also due to the existence of multiple local optima (Javaloyes-Anton et al., 2013). In a previous work (Chia et al., 2021), a combined approach between stochastic (Genetic Algorithm, GA) and deterministic (Outer Approximation / Equality Relaxation / Augmented Penalty, OAERAP) methods was proposed where the preliminary optimal design obtained from the stochastic method acts as the initial values in the deterministic method. It was found that the combined approach can effectively reduce the optimization difficulty, the computational effort, and the overall CPU time.

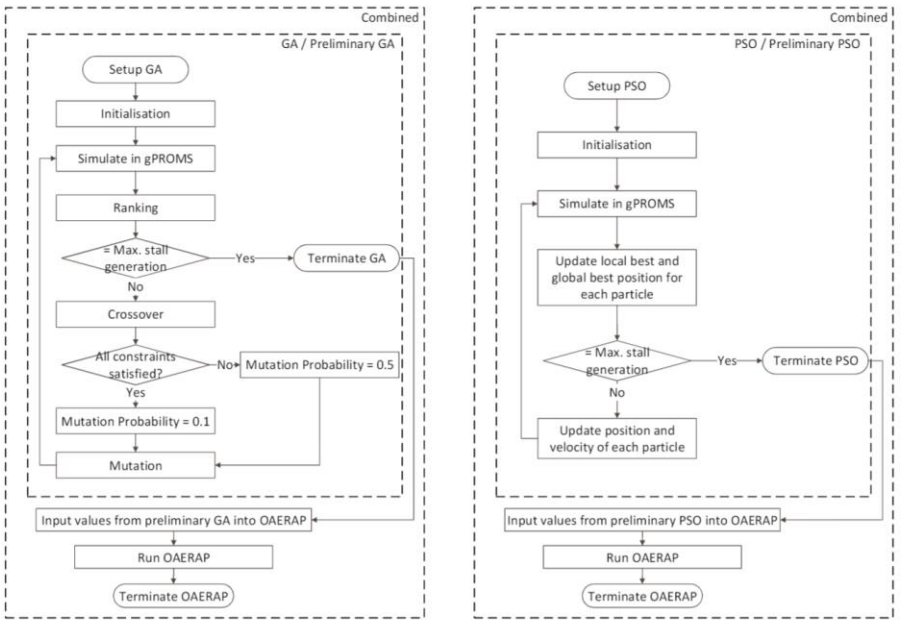


Figure 1: Flowcharts of the optimization procedures for (a) GA/OAERAP strategy (Chia et al., 2021), and (b) PSO/OAERAP strategy.

In this work, the GA/OAERAP approach is extended, replacing the GA with a Particle Swarm Optimization (PSO) which has a faster convergence speed than GA (Mesloub and Mansour, 2009). Two case studies will be used to illustrate the strategy, including a dividing wall column (DWC) and a hybrid distillation/pervaporation process.

2. Methodology

The combined optimization strategy starts with the stochastic optimization method for the global search, followed by the deterministic optimization method for the local search. Figure 1a shows the GA/OAERAP strategy also studied in a previous work (Chia et al., 2021), where the concept of a preliminary stochastic optimization was initially proposed. A looser fitness tolerance (10^{-2}) is utilized in the preliminary stochastic optimization (preliminary GA) while a tighter fitness tolerance (10^{-4}) is used in the deterministic optimization (OAERAP). In this work, another stochastic optimization (Particle Swarm Optimization, PSO) is also considered in order to evaluate the relative optimization performances of the different methods and method combinations.

The flowsheet of the PSO/OAERAP strategy is shown in Figure 1b and the inner box is the procedure for (preliminary) PSO. The PSO is developed in MATLAB (The MathWorks Inc., 2019), and several classical mathematical functions (Schwefel function and constrained Rosenbrock function) and literature (Deb, 2000; Deep et al., 2009) have been used for validation (not shown).

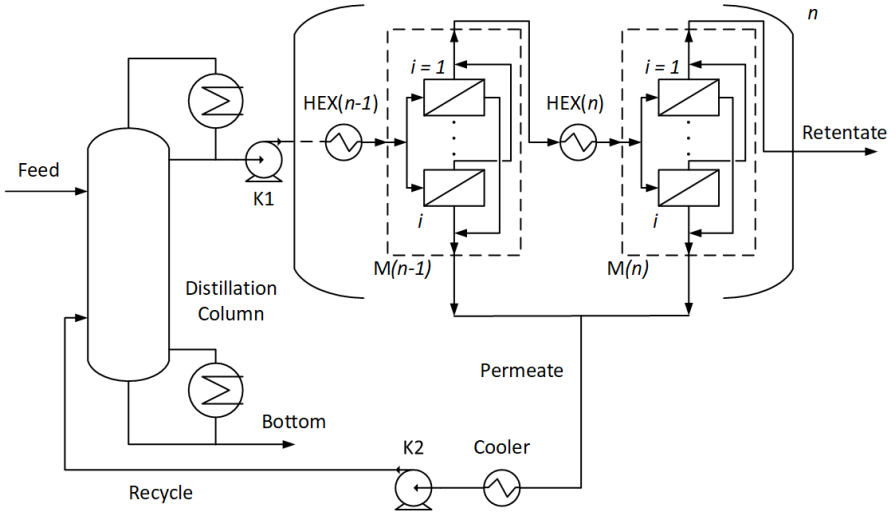


Figure 2: Superstructure of the hybrid distillation/pervaporation process.

The swarm size in PSO and the population size in GA are set to be the same, which is five times the optimization variables, for fair comparison. For the stopping criteria, both PSO and GA are terminated once the fitness value stays constant within the fitness tolerance for 20 consecutive iterations. Other settings of GA can be found in Chia et al. (2021), while the other settings for PSO are given below. The inertia is dynamically varied by randomly choosing a value from a Gaussian distribution of $w \sim N\ (0.72, \sigma)$ at each iteration (the "random adjustments" method by Engelbrecht (2007), where σ should be small enough to prevent w predominantly larger than one) with $\sigma = 1$ used in this work. The two acceleration coefficients, c_1 (cognitive) and c_2 (social), are linearly changed between 0.5 and 2.5, where c_1 decreases while c_2 increases (Engelbrecht, 2007). All particle locations are initialized using uniform randomization and all particle velocities are initially set as zero. The boundary handling method used in this work is the random forth method (Gandomi and Kashani, 2018) and the penalty function utilized is as proposed by Deb (2000). Moreover, the dynamic bound proposed in Chia et al. (2021) is also applied in PSO to reduce the chance of infeasible designs and simulations. For example, the feed location can never exceed the total number of stages in a column, so the upper bound of the feed location is dynamically changed to be equal to the total number of stages. The outer box in Figure 1b shows the OAERAP procedure. The results from the preliminary PSO are used as initial values in OAERAP. The constraint tolerances in both preliminary PSO and OAERAP, and the fitness tolerance in OAERAP, are set as 10^{-4}, although the fitness tolerance in preliminary PSO is set as 10^{-2}.

The GA and PSO are developed in MATLAB, and OAERAP is built-in in gPROMS ProcessBuilder (Process Systems Enterprise, 2020). All rigorous simulations are performed in gPROMS ProcessBuilder, and gO:MATLAB (Process Systems Enterprise, 2019) is used to transfer data between gPROMS ProcessBuilder and MATLAB. Moreover, the parallel computing function is activated in MATLAB to use 18 workers while doing the stochastic optimization to speed up the optimization.

Table 1: Case Study 1, Dividing Wall Column - Comparison between optimization methods

Items	GA	PSO	GA/OAERAP		PSO/OAERAP	
			Prelim.	Final	Prelim.	Final
Prefractionator						
Total stages	23	24	22	23	21	23
Feed stage	12	14	13	14	12	12
Main Column						
Total stages	42	44	41	42	42	43
Feed stages	8/32	8/33	8/31	7/31	9/31	7/31
Liq. sidedraw stages	18/8	18/8	18/6	19/7	18/9	19/7
Vap. sidedraw stage	32	33	31	31	31	31
Liq. side † ($kmol\ h^{-1}$)	332/178	332/176	330/183	330/182	332/186	330/173
Vap. side † ($kmol\ h^{-1}$)	614	627	632	641	645	623
Distillate ($kmol\ h^{-1}$)	333	333	334	336	333	336
Molar reflux ratio	2.29	2.24	2.30	2.22	2.38	2.22
Fitness and Time						
TAC ($M\ \$\ y^{-1}$)	6.0374	5.9924	6.0646	5.9852	6.0206	5.9879
Generation/Iteration	190	97	86	-	61	-
Unparallel CPU time (s)	20346	25704	15764	128	14746	117
Parallel CPU time* (s)	2052	2341	1697	-	1465	-
Total CPU time* (s)	2052	2341	1825		1582	

* Parallel computing used, number of workers/cores = 18

† Liquid/vapor side draw from the main column to the prefractionator

3. Case Studies

The proposed PSO/OAERAP optimization strategy is compared with stand-alone GA, stand-alone PSO, and a GA/OAERAP optimization strategy using two different distillation structures including a dividing wall column (DWC) and a hybrid distillation/pervaporation process. In the stand-alone GA and PSO, the tolerance of both constraints and fitness is set as 10^{-4}. All optimization tasks are repeated several times and the designs with the most common fitness and CPU time are chosen for comparison. All calculations are performed using a desktop with an AMD Ryzen 9 3900X CPU (24 logical processors) with 3.79 GHz and 64 GB memory.

3.1. Description

Case study 1 is the separation of equi-molar benzene/toluene/o-xylene (UNIQUAC as thermodynamic model) in a dividing wall column. The feed flowrate is 1000 $kmol\ h^{-1}$ and the feed condition is saturated liquid at 0.37 *bar*. The column is operated at 0.37 *bar*, and pressure drop is neglected. Since there is no commercially available DWC library in gPROMS ProcessBuilder, a thermodynamically equivalent Petlyuk structure is used instead. Case study 2 is a hybrid distillation/pervaporation process (Figure 2) and the membrane model is a user-defined lumped hollow fiber model (validated against the experimental data by Tsuyumoto et al. (1997), not shown) to separate a feed of 200 $kmol\ h^{-1}$ saturated liquid ethanol/water mixture at 1 *bar* with 10 *mol %* ethanol (UNIQUAC as thermodynamic model). The membrane network allows optimization of number of membrane stages in series, number of membrane modules in parallel at each stage, and the existence of membrane stage feed heaters. Due to computational limitations, a simplified optimization method proposed by Marriott and Sorensen (2003) is used, with the upper bound of the number of membrane stages studied set as eight stages. The

Table 2: Case Study 2, Hybrid Distillation - Comparison between optimization methods

Items	GA	PSO	GA/OAERAP		PSO/OAERAP	
			Prelim.	Final	Prelim.	Final
Column						
Total stages	21	21	20	19	21	21
Feed stages	17/17	17/18	17/17	16/18	17/17	17/20
Distillate ($kmol\ h^{-1}$)	23	23	23	23	23	23
Molar reflux ratio	1.15	1.19	1.45	1.34	1.19	1.16
Membrane Network *						
No. membrane stages	5	5	5	5	5	5
No. modules in stage 1	8	7	8	8	9	7
No. modules in stage 2	11	9	11	10	13	8
No. modules in stage 3	9	11	8	8	12	12
No. modules in stage 4	17	14	14	15	14	15
No. modules in stage 5	18	20	19	20	15	20
Total membrane area (m^2)	378	366	360	366	378	372
Fitness and Time						
TAC ($M\ \$\ y^{-1}$)	0.7575	0.7506	0.7690	0.7606	0.7593	0.7527
Generation/Iteration	93	86	45	-	34	-
Unparallel CPU time (s)	7308	9082	4140	22	4303	9
Parallel CPU time † (s)	813	828	433	-	371	-
Total CPU time † (s)	813	828	455		380	

* Existence of membrane stage feed heater is not shown due to space limit, but all the optimization results show that feed heaters exist for all stages except for stage 1

† Parallel computing used, number of workers/cores = 18

number of membrane stages is first fixed, with all the other variables optimized simultaneously. This procedure is then repeated for each number of membrane stages that are being studied and the optimized total annualized costs (TACs) are recorded. Finally, the structure which gives the lowest TAC is considered as the optimal design, and its optimized variables, CPU time, and TAC are shown in the Results section.

3.2. Results and Discussion

The optimized designs of each case study using the different optimization methods are shown in Tables 1 and 2, respectively. The TAC and CPU time are chosen as the main performance indicators. The optimization is significantly sped up with the use of parallel computing with 18 workers, where it is up to 11 times and 12 times faster for the two case studies, respectively. Comparing stand-alone GA and PSO, and preliminary GA and PSO, it shows that PSO and preliminary PSO usually require fewer iterations but a longer CPU time per iteration than GA. GA requires lesser time per iteration due to its algorithm where a few chromosomes in the current generation will appear in the next generation (elite parents and low mutation probability at the end), leading to a higher rate of feasible simulation and lower CPU time for each generation, owing to the "sequential initialization" algorithm in gPROMS ProcessBuilder (subsequent simulation is initialized using the results from the previous simulation, and the simulation will be easier with similar initial values). However, by considering the parallel CPU time, PSO and preliminary PSO show similar CPU times with GA and preliminary GA, indicating a

better paralleling computing performance than PSO and preliminary PSO. Comparison between stand-alone GA and PSO shows that PSO can yield a better design. However, for GA/OAERAP and PSO/OAERAP, the final TACs are similar. The comparison between combined optimization strategies and their corresponding stand-alone methods shows that the combined optimization strategies can achieve good optimal designs in a shorter CPU time, where the time savings for GA/OAERAP and PSO/OAERAP are up to 44 % and 54 %, respectively. The PSO/OAERAP is even faster than GA/OAERAP with time saving up to 16 % for the combined PSO strategy compared to the GA strategy. In our experience, the optimization of distillation using PSO and preliminary PSO also has a much higher chance to converge into a good optimal design and the results from several repeated optimizations are close.

4. Conclusions

This work considers a combined optimization strategy of a stochastic optimization method (GA or PSO) followed by a deterministic optimization method (OAERAP) for the optimisation of different distillation systems, implemented using parallel computing. The comparison between stand-alone stochastic methods (GA and PSO) and the combined strategy (GA/OAERAP and PSO/OAERAP) shows that PSO/OAERAP can yield an optimal design within a significantly shorter CPU time. With the help of the parallel computing function (18 workers), the optimization speed can be about 11 times faster and the effect is even better for PSO than for GA. Moreover, a simple superstructure optimization of a hybrid distillation/pervaporation process (case study 2) showed that the combined optimization strategy is also capable of handling superstructure optimization of distillation systems.

References

Chia, D. N., Duanmu, F., and Sorensen, E. (2021). Optimal Design of Distillation Columns Using a Combined Optimisation Approach. In Turkay, M. and Gani, R., editors, 31st European Symposium on Computer Aided Process Engineering, pages 153–158. Elsevier B.V.Y.

Deb, K. (2000). An efficient constraint handling method for genetic algorithms. Computer Methods in Applied Mechanics and Engineering, 186(2-4):311–338.

Deep, K., Singh, K. P., Kansal, M., and Mohan, C. (2009). A real coded genetic algorithm for solving integer and mixed integer optimization problems. Applied Mathematics and Computation, 212(2):505–518.

Engelbrecht, A. (2007). Computational intelligence: An introduction. John Wiley&Sons Ltd, 2 ed.

Gandomi, A. H. and Kashani, A. R. (2018). Probabilistic evolutionary bound constraint handling for particle swarm optimization. Operational Research, 18(3):801–823.

Javaloyes-Anton, J., Ruiz-Femenia, R., and Caballero, J. A. (2013). Rigorous Design of Complex Distillation Columns Using Process Simulators and the Particle Swarm Optimization Algorithm. Industrial & Engineering Chemistry Research, 52(44):15621– 15634.

Marriott, J. and Sorensen, E. (2003). The optimal design of membrane systems. Chemical Engineering Science, 58(22):4991–5004.

Mesloub, S. and Mansour, A. (2009). Hybrid PSO and GA for global maximization. International Journal of Open Problems in Computer Science and Mathematics, 2(4):597– 608.

Process Systems Enterprise (2019). gO:MATLAB.

Process Systems Enterprise (2020). gPROMS ProcessBuilder version 1.4.

The MathWorks Inc. (2019). MATLAB R2019b version 9.7.

Tsuyumoto, M., Teramoto, A., and Meares, P. (1997). Dehydration of ethanol on a pilotplant scale, using a new type of hollow-fiber membrane. Journal of Membrane Sci., 133(1):83–94.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50220-7

A semantic based decision support framework to enable model and data integration

Edlira Vakaj*, Linsey Koo[b], Franjo Cecelja[b]

[a]*Department of Computing and Data Science, Brimingham City University, Birmingham, UK*
[b]*Centre for process & Information Systems Engineering, University of Surrey, Guildford GU2 7XH, UK*

Edlira.vakaj@bcu.ac.uk

Abstract

Model reusability and integration with datasets are major contributors towards their interoperability, the concepts that follows process established by computer aided process engineering (CAPE) community (Belaud & Pons 2002). This paper proposes a semantic approach which enables model/data registration, their discovery and concomitantly model their integration. The functionality of the process is fully controlled by a biorefining domain ontology implemented using Ontology Web Language (OWL) and tested using biorefining related scenarios.

Keywords: biorefining, semantic repository, model integration, OWL-S.

1. Introduction

Significant efforts are applied on modelling to support synthesis and design, planning and scheduling, process monitoring and control in biorefining. Modelling addresses the complexity in characteristics of various types of feedstock and associated processing technologies and respective pathways, including thermochemical, chemical, biochemical and hybrid conversion routes, among other aspects. As a result, there are many contributions from modelling and experimental prospective and a large number of custom-made models and data are available to be used and hence reused. Integrating existing models and data developed/generated by different tools and processes, sharing and reusing them towards a common aim is still a challenge. To address this problem, we introduce a semantic approach towards model and data integration, all coordinated by a biorefinery ontology. The concept follows on the principle of service integration which in consequence benefits from respective Ontology Web Service (OWL-S) framework.

2. Theoretical formulation of model/data integration

A superstructure approach is used to represent models by their functionality, inputs, outputs and auxiliary inputs (Figure 1a), whereas datasets are characterised by their functionality in relation to what they describe, outputs representing actual data and auxiliary inputs (Figure 1b). The model inputs and the model and data outputs and auxiliary inputs are characterised by n_I input properties $P_i^{n_I}$, n_O output properties $P_i^{n_O}$ and n_E auxiliary properties $P_i^{n_E}$ as shown in (Figure 1) (Koo at al, 2017).
The whole process of matching and then model and data integration is coordinated by the InterCAPEmodel ontology defined as a 6-tuple structure $O = \langle H^I, H_C, R_i^C, R^C, E^C, S_i^I \rangle$ of i) instances s_j formed of n_M models and n_D datasets, organised into classes S_i^I with N_i^I distinct names representing concepts, ii) a graph $H_C = (S_i^I, is-a)$ forming a

subsumption hierarchy in ontology sense, iii) class relationships R_i^C which form a set of bijective relationships $r_{i,j}$ between all the elements of domain classes, iv) a subsumtion R^C of properties R_i^C , and v) extension E^C of a class S_i^I which is defined by the relationship R_i^C profiling the structural properties of the class by its relations with other classes.

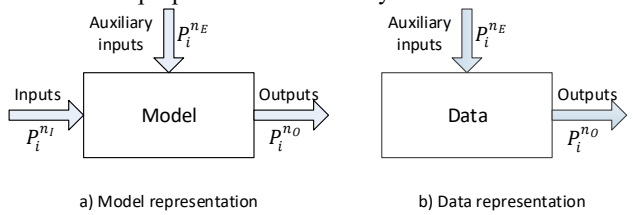


Figure 1 Model and data representation

Model and data matching is performed on the request of the user for model/data discovery or formed by the input properties of requesting model for model/data integration. The matching process is based on i) metrics defined over model/data output properties $P_i^{No} = \{p_{i,j} | p_{i,j} \in \mathbb{R},\ i = 1, \cdots, n_C\}_{j=1}^{No}$ characterising them (only numerical output properties are taken into account) and representing the domain explicit knowledge, and ii) metrics defined over mutual position (distance measured in number of edges) of respective classes in the domain ontology, the InterCAPEmodel ontology, and representing the domain tacit knowledge. For matching the properties $\mathbf{p}_n$, we define h-metric h_k^V as

$$h_k^V = \left\{ \frac{\mathbf{p}_i \cdot \mathbf{p}_j}{\|\mathbf{p}_i\| \|\mathbf{p}_j\|} \right\}_{i,j=1}^{\mathrm{n}},\ k = 1,2,\cdots,2^{\mathrm{n}} \tag{1}$$

representing the measure known as the ***vector similarity***. For metrics measuring the mutual position of respective S_i^I and S_j^I classes in the ontology, we define h-metric h_k^C as

$$h_k^C = \min_{S_C^I \in H^I} \left[\delta(S_i^I, S_C^I) + \delta\left(S_j^I, S_C^I\right) \right] \tag{2}$$

where $\delta(S_i^I, S_C^I)$ ($\delta\left(S_j^I, S_C^I\right)$) is the distance between classes S_i^I (S_j^I) and another class S_C^I measured in number of intermediate edges in graph sense along subsumption H_C and along selected R_i^C relationships. Final match is obtained as an aggregated similarity measure h_k as

$$h_k = \frac{\alpha h_k^V + \beta h_k^C}{\alpha + \beta} \tag{3}$$

where α and β are weighting factors deepening the semantics of the ontology similarity and their values are dictated by the application. The similarity h_k is then used to support decision in model and data integration and/or reuse.

3. Implementation

The semantic based decision support platform was implemented using semantic web technologies and knowledge graphs. The core functionalities include: i) *Model/data Registration* for registering a model with their Inputs, Outputs, Preconditions and Effects, ii) *Model/data Publishing* for sharing models and data, iii) *Model/data Discovery* for creating a set of candidate models and data for integration and/or reuse, iv) *Model/data Selection* towards most appropriate integration, and v) *Model composition* for integration and data exchange between the selected models and datasets. The architecture (Figure 1)

of the platform follows the semantic web application principles in which various and heterogeneous data sources are considered, their mapping to the corresponding schema or ontology for providing a knowledge graph of concepts and associated instances (Kalemi et al. 2017). The platform is organised around the InterCAPEmodel domain ontology and well established OWL-S ontology. The InterCAPEmodel ontology defines a common vocabulary to capture the knowledge of various biorefining models or the supply chain network using taxonomy, attributes, and relations of the models and datasets. OWL-S framework (Martin et al. 2007), is a set of interlinked ontologies used to semantically describe web services. In the approach proposed in this paper, each model or dataset is considered as a web service and hence making use of OWL-S which semantically describes them in terms of their functionality, inputs required for successful execution, outputs generated in the course and after the execution, and preconditions (auxiliary inputs) needed for models to run. The model/data knowledge is acquired through the registration phase and concomitant assigning as ontology instances.

The *Model/data Discovery* module is a core part of the platform based on the match-making process. We have adopted a semantic matching which follows the three-stage input/output matching approach: i) elimination, where all datasets and models which do not satisfy the critical criteria are excluded from the selection list, ii) semantic matching by calculating similarity measures which defines the level of compatibility between the requested model/data and the candidate models/data as defined by eq. (3), and iii) ranking of the candidate models/data by similarity measure. The critical criteria is defined by the user during the formulation of the requesting model. Semantic matching process calculates the semantic relevance between the requesting model and all the models published in the repository. This is calculated as an aggregated value of distance measure (eq. (2)) signifying semantic similarity and property similarity (eq. (1)) (Koo et al., 2017).

JAVA programming language and GraphDB in the form of a native RDF graph database are the key technologies used to implement this platform. The interfaces are designed using Bootstrap and Java Server Pages technologies and can be accessed through web browsers. OWLAPI, a Java library for manipulating OWL ontologies is used to verify the consistency of the knowledge base, perform reasoning and disseminate queries to the knowledge base via SPARQL Protocol. In collaboration with the reasoner (Pellet) and other supporting Java classes, OWLAPI, are checking the domain knowledge for consistency, parsing the inferred ontology for passing it to further elaboration from the front end, expanding the knowledge base with new instances of models or integrated models.

Figure 4 demonstrates the platform interface for model/data discovery and concomitant integration. As the first step, the user defines the request, which contains the functionality of the requesting model, the characteristics and the inputs that it needs. In Step 2 a list of most suitable candidate models/data is provided ranked according to the matching level with justification of the match level.

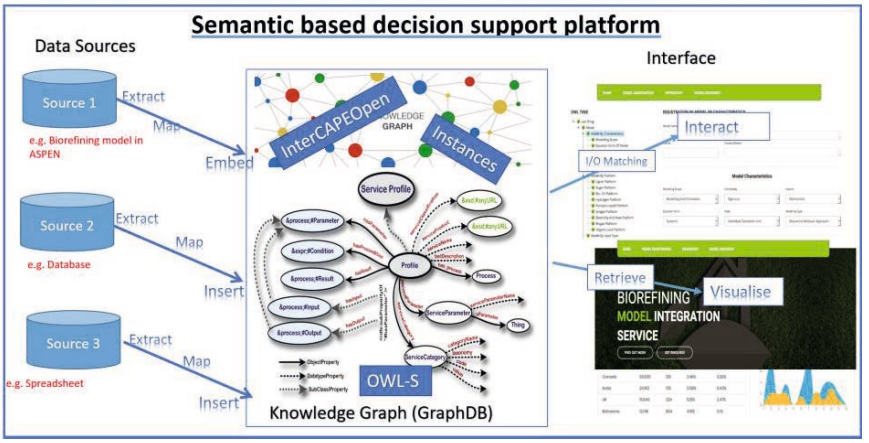


Figure 2 Semantic based decision platform architecture.

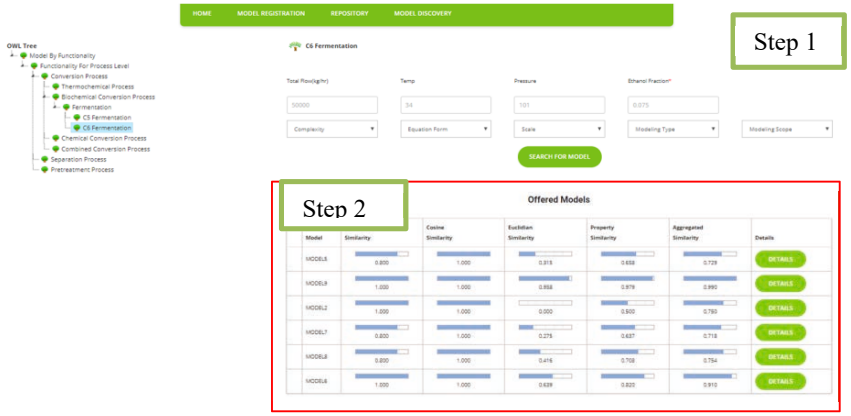


Figure 4 Model/data discovery interface

4. Demonstration

The performance of the proposed model and data integration is demonstrated by a commercial scale process flowsheet for biodiesel production from virgin vegetable oil and waste cooking oil (Zhang *et al.*, 2003; Nguyen and Demirel, 2011). Each stage of the process, such as i) transesterification reaction, ii) methanol recovery, iii) glycerol separation and (iv) fatty acid methyl esters (FAME) purification, are separately registered as an instance in the repository to demonstrate the possible discovery of the models. Here, the model representing FAME purification is the last in chain and therefore identified as a requesting model. The required input component of this model is FAME, hence, the models that have FAME as an output are discovered as a potential match to the FAME

purification model. The discovery process was performed by backward matching process and the profiles of discovered models are listed in Table 1. The restriction to the FAME purification process is the presence of glycerol, therefore, the streams that contain glycerol are eliminated, as illustrated in Figure 5. Therefore, the user is presented with three models that have FAME as an output stream without the presence of glycerol in the stream. The combination of hexane extraction, water washing and glycerol separation processes represents a glycerol separation process. The second stage of backward matching is conducted to run the discovered models, which requires FAME and glycerol as inputs in the stream. As a result, a model representing methanol recovery process that obtains a good separation of methanol from FAME and glycerol was identified. In addition, as a final stage of matching, transesterification models using four different catalysts were discovered as potential matches to the methanol recovery model to form a network of process models representing biodiesel production from feedstock to final product.

Table 1 Profile of processes offering potential match

Process Model	# Input Stream	Input components per stream	# Output Stream	Output components per stream
Hexane Extraction	2	S1. FAME, Glycerol S2. Hexane	2	S1. FAME S2. Glycerol
Water Washing	2	S1. FAME, Glycerol S2. Water		
Glycerol Separation	1	S1. FAME, Glycerol		
Methanol Recovery	1	S1. Methanol, FAME, Glycerol	2	S1. Methanol (Recycle) S2. FAME, Glycerol
Transesterification 1	3	S1. Waste Oil S2. Methanol S3. Catalyst: T1 - H_2SO_4 T2 - H_2SO_4, NaOH	2	S1. FAME, Glycerol, Methanol S2. Catalyst (Recycle)
Transesterification 2				
Transesterification 3	3	S1. Virgin Oil S2. Methanol S3. Catalyst: T1 - NaOH T2 - Ca_3La_1		
Transesterification 4				

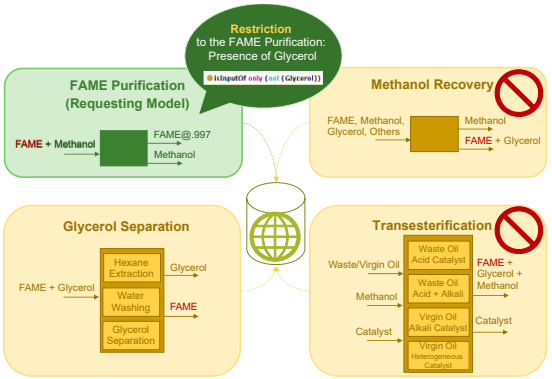


Figure 5 Process of Elimination of the Models with Glycerol

5. Conclusion

The concept of using ontology in model and data integration was introduced to improve upon previous research with particular focus on i) flexibility (partial matching) and reusability (reuse of existing models and data), ii) coordination of every step in the process of model and data integration starting with model and data registration up to their invocation and execution. The semantic algorithm for establishing interoperability between the models and data is presented to reflect the knowledge based on technical compatibility and functional feasibility. The domain ontology embeds both tacit and explicit knowledge in the domain of biorefining modelling. Process models and data are semantically annotated in terms of input(s), output(s), precondition(s), the software environment in which they operate, as well as the functionality they perform. The ma5tching concept is fully justified and its implementation explained.

References

Jean-Pierre Belaud, Michel Pons, Open Software Architecture For Process Simulation: The Current Status of CAPE-OPEN Standard, Computer Aided Chemical Engineering,Volume 10,2002,Pages 847-852,

Kalemi, E., L. Koo, et al. (2017). "A Semantic Repository for Model Integration in Biorefining." Computer Aided Chemical Engineering 40: 2323-2328

Koo, L., N. Trokanas, et al. (2017). "A semantic framework for enabling model integration for biorefining." Computers & Chemical Engineering **100**: 219-231

Nguyen, N. T. and Demirel, Y. (2011) 'A Novel Biodiesel and Glycerol Carbonate Production Plant', International Journal of Chemical Reactor Engineering, 9(1), pp. 5–7.

Zhang, Y. et al. (2003) 'Biodiesel production from waste cooking oil: 1. Process design and technological assessment', Bioresource Technology, 89(1), pp. 1–16

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50221-9

Design of Non-Redundant Sensor Networks for Reliable Estimation in Nonlinear Systems

Garima Patel[a], Mani Bhushan[a*]

[a]*Department of Chemical Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India*
mbhushan@iitb.ac.in

Abstract

In this work, we focus on the problem of designing non-redundant sensor network for steady state nonlinear processes where minimum number of sensors are placed to ensure reliable estimation of all process variables. We propose to use an equation-variable matching approach as the basis for designing sensor networks. This approach is based only on the structural information of the system of equations and hence can be used for nonlinear systems without requiring any linearization or knowledge of operating values of the process variables and process parameters. Based on the obtained equation-variable matching, a signal flow graph which represents the sequence of computation of various variables is constructed. Reliability of estimating each variable is then computed based on information extracted from the signal flow graph. The approach is applied to a nonlinear process case study to design optimal sensor network.

Keywords: Reliability, Equation variable matching, Structural equations, Optimization.

1. Introduction

In a typical chemical process, it is generally not possible to measure each and every variable due to technical and economic issues, as well as due to complexities associated with collecting, and processing a large amount of data in real-time. Choosing appropriate variables to be measured in a process, is the sensor network design problem and is the focus of this work. Sensor networks can be designed to ensure satisfaction of various requirements such as those related to observability, estimability, reliability, etc [Bagajewicz and Sanchez, 1999]. For a well-designed sensor network, it is important to be able to estimate all the variables with high reliability, either based on their direct measurement or relationships with other variables as captured by a process model, given that sensors are prone to failure. The current work focuses on this aspect of reliable estimation. For linear flow processes, many graph-theoretic concepts based on spanning trees, cycles, cutsets, chord sets, etc. have been used to aide in the design of sensor networks for reliable estimation [Ali and Narasimhan, 1993, Prakash et. al., 2020].

Most of the real-world systems are nonlinear in nature. For nonlinear processes, the literature is comparatively scarce. While an approach to design sensor network based on cost and precision is available [DuyQuang and Bagajewicz, 2008], there is no reported work on ensuring reliable estimation for nonlinear processes. In this work, we focus on the problem of designing non-redundant sensor network for steady state nonlinear processes where minimum number of sensors are placed to ensure reliable estimation of all process variables.

We propose to use equation-variable matching approach as the basis for designing sensor networks. Equation-variable matching approach is widely used in flowsheeting (process design) literature [Stadtherr et. al., 1984]. This approach is based only on the structural matrix of the system of equations and hence can be used for nonlinear systems without requiring any linearization or even the knowledge of operating values of the process variables and process parameters. Given a set of variables on which sensors have been placed, we propose to use equation-variable matching approach to match unmeasured variables to equations. A full matching ensures that based on the structural relationships amongst variables, the unmeasured variables can be estimated with the given set of sensors. Based on the obtained matching, a signal flow graph capturing the sequence of computation of various variables is constructed. This in turn allows us to compute the reliability of estimation of each variable given the failure probabilities of sensors on the measured variables. The optimal sensor network design then corresponds to the set of measurements which maximize the minimum reliability of estimation amongst all variables. The proposed approach is generic in the sense that it can be applied to any nonlinear process as it does not make any assumption about the form of the equations.

The rest of the paper is organized as follows. The sensor network design problem for reliable estimation in nonlinear systems is discussed in Section 2. The proposed approach to compute reliability of estimation is discussed in Section 3. An example based on proposed approach is shown in Section 4. Further, case study is presented in Section 5., followed by conclusions and discussions in Section 6.

2. Problem Definition

Consider a nonlinear steady state process modeled as:

$$f(x) = 0 \tag{1}$$

where $x \in R^n$ are the process variables, and $f : R^n \rightarrow R^m$represents the model equations which are nonlinear in nature. Further it is assumed that $n \geq m$. Given the model equations as Eq.(1), the sensor network design problem is to choose $n - m$ variables to be measured so as to maximize the minimum reliability of estimation amongst all variables. This is stated as:

$$\max_{q_1,q_2,\ldots,q_n} \min_{1\leq i\leq n} R_i : s.t. \sum_{i=1}^{n} q_i = n - m, \text{ with } q_i \in \{0,1\},\ i = 1,2,\ldots,n \tag{2}$$

where, R_i is the reliability of estimation of variable x_i and q_i is a binary variable indicating if variable x_i is measured ($q_i = 1$) or unmeasured ($q_i = 0$). For linear mass flow system, computation of R_i is well understood using graph theoretic concepts [Ali and Narasimhan, 1993, Prakash et. al., 2020]. However, for nonlinear systems this issue has not been addressed in literature and is the focus of the current work.

3. Proposed Methodology

Given a set of $n - m$ measurements in a nonlinear process, we now propose an equation variable matching followed by depth-first search of a signal flow graph to compute reliabilities of estimation of all variables. These ideas are discussed next. Let M be the

set of measured variables $m_j \in M, j = 1, 2, \ldots, n-m$, and U be the set of unmeasured variables $u_i \in U, i = 1, 2, \ldots, m$.

3.1. Equation-Variable Matching

As a first step in the proposed approach, each equation is matched to one of its incident unmeasured variable, i.e. to an unmeasured variable which occurs in that equation. This matching is done such that no two equations are matched to the same unmeasured variable, and similarly no two unmeasured variables are matched to the same equations. Matching problem is a well studied problem in literature with several applications and several tools to obtain maximum cardinality matchings [Hopcroft and Karp, 1973, Irving et. al., 2006]. The equation variable matching is based on structural matrix **Z** representation of the system of equations in the form of Eq.(1), such that $Z_{i,j} = 1$ if variable x_j appears in equation f_i, and is 0 otherwise. A full equation-variable matching where all the m equations are matched ensures that the system of equations is structurally solvable to yield the values of the unmeasured variables. Lack of full matching implies that at least one unmeasured variable cannot be estimated. For the given set of measurements, to simplify the notation, assume that equation f_i is matched to unmeasured variable $u_i, i = 1, 2, \ldots, m$. The reliability of estimation of each variable can now be computed as discussed next.

3.2. Reliability Computation

The equation-variable matching information provides a sequence in which the variables can be estimated. In particular, variable u_i can be estimated by solving f_i for u_i. This in turn will need estimates of all other variables appearing in equation f_i and thus reliability of estimation of u_i can be obtained in principle by knowing the reliabilities of estimation of other variables occurring in f_i. However, the other unmeasured variables are also computed from their corresponding matched equations using similar concept and their reliabilities are dependent on the variables appearing in those equations. Only the measured variables are not matched to any equation. This sequential nature of variable computation can be represented by a signal flow graph [Robichaud et. al., 1962] $G = (V, E)$. A signal flow graph G is a digraph, with V being the set of vertices, and E being the set of directed edges. Each vertex corresponds to a variable in the original system. An edge from node v_i (corresponding to variable x_i) to v_j (corresponding to variable x_j) exists if variable x_i appears in the equation which is used to compute variable x_j. Since measured variables are not computed from the equations, they will be the root nodes [Cormen et. al., 2001] (nodes with only output edges) in the signal flow graph.

Depth first search based graph traversal [Reif, 1985] from a root node corresponding to measured variable m_i in the signal flow graph is now performed to obtain the set of nodes which can be reached from this root node. The unmeasured variables belonging to this set of reachable nodes are the ones which require the value of m_i for their computation. This depth first search traversal is performed from each root node. The reliability of estimation of any unmeasured variable u_i is then simply the product of reliabilities of sensors of those measured variables from where u_i is reachable, i.e.

$$R_{u_i} = \prod_{m_j \in S_i} R_{m_j}, \qquad u_i \in U \tag{3}$$

where, S_i is the set of measured variables from which there is a path to variable u_i in the signal flow graph. For the measured variables, the reliability of estimation is simply the reliability of the corresponding sensor being in working condition, i.e.

$$R_{m_j} = (1 - s_j), \qquad m_j \in M \tag{4}$$

with s_j being the failure probability of sensor used to measure variable m_j. For the given set of measurements, the system reliability is then the minimum reliability amongst all measured and unmeasured variables. We now present an example to illustrate the proposed ideas.

Remark 1: In the above discussion, it has been assumed that the sensor failure events across different sensors are independent events.

Remark 2: If full equation-variable matching is not obtained in the equation-variable matching step, then system reliability can be directly assigned a value of 0 without any further processing.

4. Example

Consider ammonia flow process [Ali and Narasimhan, 1993] for illustration here. The process is a mass flow process consisting of eight variables (flow rates) and five process units- each corresponding to a mass balance. The process graph is shown in Figure 1(a). The structural matrix of the system of equations is:

$$\mathbf{Z} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix} \tag{5}$$

This process has eight variables and five equations, and hence three variables need to be measured for a non-redundant sensor placement. For illustration let the measurements be of variables $\{x_1, x_4, x_6\}$. Equation-variable matching is then performed using Dulmage-Mendelsohn decomposition (dmperm function in Matlab) [Dulmage and Mendelsohn, 1958]. The signal flow graph corresponding to this matching is shown in Figure 1(b). In particular, unmeasured variables x_2, x_3, x_5, x_7, x_8 can be matched to equations 2, 3, 4, 5, and 1, respectively. Depth first search traversal is now performed on this graph to compute the set S_i of measured variables from which there is a path to unmeasured variable i in the signal flow graph. For example, set S_8 for unmeasured variable x_8 is $S_8 = \{x_1, x_6\}$ and thus the reliability of estimation of variable x_8 is $R_8 = R_1 R_6 = (1 - s_1)(1 - s_6)$. Similarly, the reliability expressions for all other variables can be computed and are listed as:

$$R_2 = R_1,\ R_3 = R_1,\ R_5 = R_1 R_4,\ R_7 = R_1 R_4,\ R_8 = R_1 R_6 \tag{4}$$

where, $R_1 = (1 - s_1)$, $R_4 = (1 - s_4)$, $R_6 = (1 - s_6)$ with s_1, s_4, s_6 being the failure probabilities of sensors used to measure variables x_1, x_4 and x_6, respectively.

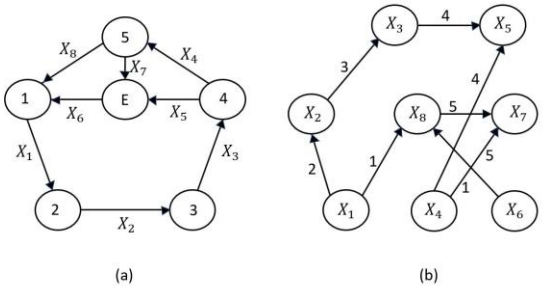
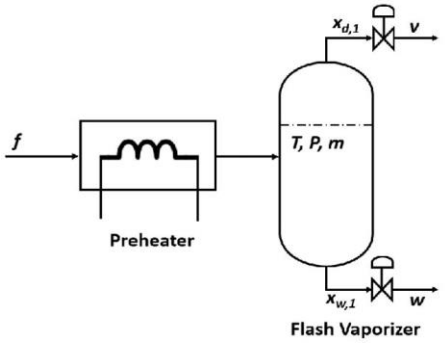


Figure 1: (a) Ammonia process graph [Ali and Narasimhan, 1993] (b) The signal flow graph with sensors on $\{x_1, x_4, x_6\}$

We now present a case study of a nonlinear system to perform optimal sensor placement for reliable estimation.

5. Case Study

Consider a steady state nonlinear process shown in Figure 2(a) explained by [Maurya et. al., 2003]. The variables considered in the process are $f, v, w, T, P, m, x_{d1}, x_{d2}, x_{w1}, x_{w2}$ with the sensor failure probabilities as 0.11, 0.12, 0.14, 0.21, 0.24, 0.32, 0.35, 0.38, 0.30, 0.37, respectively. The mathematical model describing the process is given in Figure 2(b).



$$f_1(f, v, w) = 0$$
$$f_2(f, v, m, x_{d1}, x_{w1}) = 0$$
$$f_3(x_{w1}, x_{w2}) = 0$$
$$f_4(f, v, T, m, x_{d1}, x_{d2}, x_{w1}, x_{w2}) = 0$$
$$f_5(T, P, x_{d1}, x_{w1}) = 0$$
$$f_6(T, P, x_{d2}, x_{w2}) = 0$$
$$f_7(x_{d1}, x_{d2}) = 0$$

(b)

Figure 2: (a) Flash Vaporizer [Maurya et. al., 2003] (b) Mathematical Model

This process has ten variables and seven equations, hence the minimum number of measured variables is three. Thus, there are a total of $\binom{10}{3} = 120$ choices. For each choice, the reliabilities of all the variables are computed using the equation-variable matching and signal flow graph traversal approach discussed in section 3. The minimum reliability across all the variables is the system reliability for that measurement combination. By solving the problem given in Eq.(2), the optimal reliability of system, R_{sys} is 0.266. The variables chosen for optimum sensor placement are $\{m, x_{d2}, x_{w2}\}$.

6. Conclusions

A method to design non-redundant sensor network for reliable estimation for nonlinear processes has been proposed in the current work. Based on the concept of equation variable matching, the approach utilizes only structural information about the nonlinear equations and is hence applicable at the design stage as well. The utility of the approach was demonstrated by applying it to a nonlinear case study. Extension of the approach to optimally design redundant sensor networks is being currently investigated.

References

Y. Ali, S. Narasimhan, 1993, Sensor network design for maximizing reliability of linear processes, *AIChE J., 39*, 820

M. J. Bagajewicz, M. C. Sanchez, 1999, Design and upgrade of nonredundant and redundant linear sensor networks, *AIChE J., 45*, 1927

T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, 2001, Introduction to algorithms, *MIT Electrical Engineering and Computer Science*, The MIT Press, 253

A. Dulmage, N. Mendelsohn, 1958, Coverings of Bipartite Graphs, *Canadian Journal of Mathematics, 10*, 517

N. DuyQuang, M. Bagajewicz, 2008, Design of Nonlinear Sensor Networks for Process Plants, *Ind. Eng. Chem. Res., 47*, 5529

J. E. Hopcroft, R. M. Karp, 1973, An $n^{5/2}$ Algorithm for Maximum Matchings in Bipartite Graphs, *SIAM J. Comput., 2*, 225

R. W. Irving, T. Kavitha, K. Mehlhorn, D. Michail, K. Paluch, 2006, Rank-Maximal Matchings, *ACM Transactions on Algorithms, 2*, 601

M. R. Maurya, R. Rengaswamy, V. Venkatasubramanian, 2003, A Systematic Framework for the Development and Analysis of Signed Digraphs for Chemical Processes. 2. Control Loops and Flowsheet Analysis, *Ind. Eng. Chem. Res., 42*, 4811

O. Prakash, M. Bhushan, S. Narasimhan, R. Rengaswamy, 2020, Sensor Network Design based on System Wide Reliability Criteria: 1. Objetive, *Journal of Process Control, 93*, 66

J. H. Reif, 1985, Depth-First Search is Inherently Sequential, *Inf. Process. Lett., 20*, 229

L. P. A. Robichaud, M. Boisvert, J. Robert, 1962, Signal flow graphs and applications, Englewood Cliffs, N.J., *Prentice Hall*, 214

M. Stadtherr, E. Wood, 1984, Sparse Matrix Methods for Equation-Based Chemical Process Flowsheeting-I, *Comput. Chem. Engng., 8*, 9

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50222-0

A Novel Global-Local Feature Preserving Projection Method Based on Adaptive Linear Local Tangent Space Alignment for Process Monitoring

Yang Li, Cheng Ji, Jingde Wang*, Wei Sun*

College of Chemical Engineering, Beijing University of Chemical Technology, North Third Ring Road 15, Chaoyang District, Beijing, 100029, China
sunwei@mail.buct.edu.cn

Abstract

Due to the rapid increase of complexity in industrial process, feature extraction plays an important role in process monitoring. In conventional process monitoring methods, represented by multivariate statistical process monitoring (MSPM) method, global feature of process data is mainly considered, such as variance information. By contrast, the local feature of process data, obtained by preserving neighbourhood information, is preferred in newly emerging manifold learning method. Process monitoring methods with both global and local features of data were then proposed. Theoretically, better monitoring results can be expected with the consideration of both global and local features, but methods can be significantly different for the extraction of global or local feature, which requires better mathematical description of the global and local features ahead. However, there is no specific discussion on this topic in literature yet. To address this issue, A novel global-local feature preserving projection (NGLFPP) method based on adaptive linear local tangent space alignment is proposed. The neighbourhood of each sample is given adaptively by considering the local distribution of data. The local feature of data is obtained by modified tangent estimation. On the basis of the adaptive neighbourhood, geodesic distance is introduced to represent global feature of data. In order to extract global and local features of data simultaneously, a dual objective optimization function is constructed. Tennessee Eastman (TE) process is employed to validate the proposed method.

Keywords: Global local feature preserving projection; Feature extraction; Local tangent space alignment; Adaptive neighbourhood; Geodesic distance.

1. Introduction

With the increase in the scale and complexity of industrial processes, process safety has become more and more important than ever. Process monitoring technologies, which aim at early detection of process faults, are attracting further attention. Due to the extensive use of distributed control system and advanced measuring instruments, large amounts of operational data were collected, which provides a foundation for data-driven process monitoring technology (Qin and Chiang, 2019). MSPM methods, represented by principal component analysis (PCA) and its variants, are the most intensively and widely studied data-driven process monitoring technique. However, the MSPM methods only focus on the global feature of the data represented by variance information and high-order statistics, while ignoring the local feature of the data expressed by neighbourhood information. In recent years, the emergence of manifold learning makes it possible to

preserve the local feature of data. Many manifold learning methods were first proposed for pattern recognition and their linear variants are further applied for process monitoring, such as local preserving projection (LPP) and neighbourhood preserving embedding (NPE). However, the global feature of data is not explicitly considered by the aforementioned methods.

In view of the above problems, several monitoring methods were proposed, which can preserve the global and local features of data simultaneously. A unified framework namely global-local preserving projection (GLPP) was built based on LPP and it relies on a heat kernel function to provide neighbourhood information based on distance (Luo, 2014). From another perspective, NPE based on neighbourhood reconstruction weights is also extended to preserve both global and local features of the data (Ma et al., 2015). As an effective local feature extraction method, linear local tangent space alignment (LLTSA) has rarely been extended to global-local feature extraction for process monitoring (Zhang et al.,2007). The main reason is that the performance of LLTSA is closely related to the accuracy of local tangent space estimation. PCA-based tangent space estimation in LLTSA assumes that the data are uniformly distributed, and that the neighbourhood space is locally linear. However, the above assumptions are often violated in real data. In addition, LLTSA only relies on the alignment of the neighbourhood tangent space to extract data feature, ignoring global feature of the data.

To further improve the performance of methods in this category, a novel global-local feature preserving projection method is proposed. The adaptive neighborhood based on distribution density is first constructed. Weight based on distance and manifold curvature is introduced to improve PCA-based neighborhood tangent space estimation. With LLTSA, local feature of the data is obtained by aligning the tangent space coordinates. Based on the constructed adaptive neighborhood, geodesic distance is introduced to obtain the global feature of data. A dual-objective optimization function is constructed to fully retain the global and local features of data. Process monitoring method based on NGLFPP is then built for real-time process monitoring.

2. NGLFPP Algorithm

2.1. Neighbourhood Construction

Among monitoring methods that preserve global and local features simultaneously, the k nearest neighbour (KNN) method is most commonly used to construct a neighbourhood for each sample. Since the local density distribution of data is usually different in real process, it is hard to describe the nearest neighbour for any data sample by a universal k. In addition, global feature represented by non-neighbourhood relations also depends on neighbourhood construction. Therefore, a density-based adaptive neighbourhood construction method is introduced to describe neighbourhood information of data (Ji and Yang, 2019). Given a data set $X = [x_1, x_2, \cdots, x_n] \in R^{m \times n}$, initial value k of the nearest neighbor is selected. The local density of sample x_i is defined as follows:

$$\mathrm{D_{ld}}(i) = \frac{1}{d_{i1}+d_{i2}+\cdots+d_{ik}} \tag{5}$$

where d_{ij} denotes the Euclidean distance between data sample x_i and its j^{th} nearest neighbour. The average local density is calculated by the mean of local density as Eq.(6).

$$\mathrm{D_{ald}} = \frac{1}{n}\left(\mathrm{D_{ld}}(1)+\mathrm{D_{ld}}(2)+\cdots+\mathrm{D_{ld}}(n)\right) \tag{6}$$

As shown in Eq.(7), the coefficient derived from the ratio of the local density to the average local density is calculated to adjust the initial number of nearest neighbors. The number of the nearest neighbor for data sample x_i is computed as Eq.(8). The ceil function in Eq.(8) indicates that the number in the parentheses is rounded upward.

$$D_{ef}(i) = exp\left((D_{ld}(i) - D_{ald})/D_{ald}\right) \tag{7}$$

$$k_i = ceil\left(k \times D_{ef}(i)\right) \tag{8}$$

2.2. Local Feature Extraction

Based on the above method, the local distribution of data has been considered. Further considering the inconsistent contribution of samples in the neighbourhood to the tangent space estimation, Zhang *et al*. (2011) proposed an improved local tangent space alignment (ILTSA) method based on distance weights. In addition, the assumption of local linearity for PCA-based local tangent space estimation will be violated when the curvature of the data manifold in the neighbourhood is large. So, weights based on distance and curvature in the neighbourhood are employed simultaneously to improve the accuracy of local tangent space estimation. The weight w is defined as Eq.(9).

$$w_{ij} = w_{ij}^D \times w_{ij}^C = exp\left(-D_E\left(x_i, x_{ij}\right)^2/t\right) \times exp\left(-\frac{D_G(x_i,x_{ij}) - D_E(x_i,x_{ij})}{D_E(x_i,x_{ij})}\right) \tag{9}$$

where w_{ij}^D and w_{ij}^C denote the weight based on distance and manifold curvature. $D_E\left(x_i, x_j\right)$ and $D_G\left(x_i, x_{ij}\right)$ represent the Euclidean distance and geodesic distance between x_i and its j^{th} nearest neighbour x_{ij}. The value of the t is identified as the average value of the Euclidean distance between x_i and its k_i nearest neighbor samples. The optimization objective in tangent space estimation can be reformulated as follows:

$$\begin{cases} \underset{Q}{argmin} \sum_{j=1}^{k_i} W_{ij} \left\| x_{ij} - x_i - Q^T Q\left(x_{ij} - x_i\right) \right\|_2^2 \\ s.t. \quad Q^T Q = I \end{cases} \tag{10}$$

Similar to ILSTA, Eq.(10) can be transformed into Eq.(11), and projection matrix Q can be calculated by eigenvalue decomposition for each neighborhood X_i.

$$\underset{Q^T Q = I}{argmax}\, tr(Q^T(\bar{X}_i W_i)(\bar{X}_i W_i)^T Q) \tag{11}$$

where W_i is a diagonal matrix and $W_i(j,j)=w_{ij}$. $\bar{X}_i$ is defined as $[x_{i1} - x_i, x_{i2} - x_i, \cdots, x_{ik} - x_i]$. Consistent with the LTSA, the alignment of local tangent coordinates is calculated by minimizing global reconstruction error. The objective function of local feature extraction can be further represented as follows:

$$J_{local} = min \sum_{i=1}^{n} \|E_i\|_2^2 = min\ tr(YSWW^T S^T Y) = min\ tr(A^T X B X^T A) \tag{12}$$

where $S = [S_1, S_2, \cdots, S_n]$ and $YS_i=Y_i$. W is a diagonal matrix consisting of W_i and A is a projection matrix which is defined as $Y=A^T X$.

2.3. Global Feature Extraction

Inspired by GLPP, the distance between non-neighbouring samples is retained as a representation of global feature. Compared to Euclidean distance, geodesic distance is introduced due to its more accurate estimation of the non-neighbouring distance on the data manifold. Dijkstra algorithm is used to estimate the geodesic distance by calculating

the shortest path distance based on the adjacency graph. The objective function of global feature extraction is constructed as follows:

$$J_{global} = max \sum_{i,j}^{n} \left\| y_i - y_j \right\|^2 \overline{W}_{ij} = max\ tr(A^T X L X^T A) \tag{13}$$

$$\overline{W}_{ij} = exp\ \left(-D_G\left(x_i, x_j\right)^2 / t \right) \quad if\ x_j \notin \Omega\left(\vec{x}_i\right) \text{ and } x_i \notin \Omega\left(\vec{x}_j\right) \tag{14}$$

where $D_G(x_i, x_j)$ denotes the geodesic distance between x_i and x_j, and the value of t is given by the average value of the geodesic distance between non-neighboring samples. And L denotes a Laplacian matrix, which can be calculated as $L = \overline{D} - \overline{W}$. $\overline{D}$ represents the diagonal matrix, which can be calculated by $\overline{D}(j,j) = \sum_j^n \overline{W}_{ij}$.

2.4. Unified Objective for NGLFPP

In order to preserve both local and global features of the data, a dual-objective optimization function is constructed as Eq.(15):

$$J_{NGLFPP} = \min \eta J_{local} - (1-\eta) J_{global} = \min tr(A^T X M X^T A) \tag{15}$$

where $M = \eta X B X^T - (1-\eta) X L X^T$, η is a tradeoff parameter which can be determined adaptively by the spectral radius of the matrix. The orthogonal constraint $A^TA=I$ is introduced due to its computational advantages in the statistics and residual space. The above-mentioned optimization problem can be transformed into solving of the eigenvalue problem by the Lagrange multiplier method, and projection matrix A consists of the eigenvectors corresponding to the d smallest eigenvalues.

3. NGLFPP-Based Process Monitoring

As training dataset, normalized normal condition data set X is applied to NGLFPP algorithm so as to obtain the projection matrix A. In online monitoring, according to F-distribution and χ^2-distribution, two popular monitoring statistics of new sample x_{new} and their control limits are calculated as follows:

$$T_{new}^2 = x_{new}^T A \Lambda^{-1} A^T x_{new} \le \frac{d(n-1)}{n-d} F_\alpha(d, n-d) \tag{16}$$

$$Q_{new}^2 = \left\| (I - AA^T) x_{new} \right\|^2 \le g \chi_{h,\alpha}^2 \tag{17}$$

where $\Lambda = X^T A A^T X/(n-1)$ denotes the covariance matrix of Y. The complete procedure of NGLFPP-based monitoring method is presented as Figure 1.

4. Case Study

TE process is a well-known benchmark test for evaluating monitoring performance of different methods, and more details on TE process can be found in (Downs and Vogel, 1993). 33 variables, including 22 measurement variables and 11 operation variables, are selected for monitoring purpose. Training dataset with 960 normal samples is used for model construction. The test dataset includes 21 fault types, and each test data contains 960 samples. The performance of the proposed NGLFPP-based method is compared with PCA-based and GLPP-based monitoring methods. For PCA, nine latent variables are selected by cross validation (Lee et al., 2006). For a fair comparison, the same number of latent variables is chosen for GLPP and NGLFPP. The same number of nearest neighbors k=10 is determined for GLPP and NGLFPP. The confidence level for control limits of T^2 and Q statistics are set as 99 % for three monitoring methods.

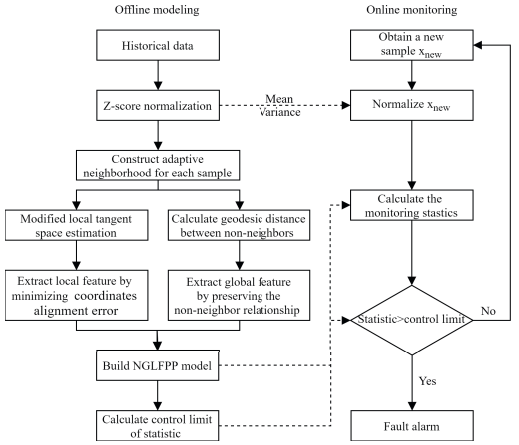


Figure 1 Monitoring diagram based on NGLFPP algorithm

Table 1 FDRs (%) of different method on TE process

	NGLFPP		**GLPP**		**PCA**	
Fault num	T^2	Q	T^2	Q	T^2	Q
1	**99.75**	99.5	**99.75**	99.5	99.125	**99.75**
2	97.25	**98.5**	94	**98.5**	**98.5**	96.125
3	1.75	11.875	1.75	11.5	1.625	1.375
4	**84.25**	96.625	41.375	96.625	6.5	**99.875**
5	**100**	34.375	99.75	**34.5**	24.75	17.5
6	**100**	99.625	**100**	99.5	99.25	**100**
7	**100**	**100**	67.875	**100**	42.125	**100**
8	96.125	**98.625**	79.375	**98.625**	**96.75**	89.25
9	1.25	8.875	0.875	8.875	1.625	1.375
10	72.375	54.375	**80.875**	**54.875**	31.875	16.75
11	**56.625**	71.625	28	71.625	22.125	**72.25**
12	**99.125**	**99.375**	94.75	**99.375**	97.25	89.75
13	**94.5**	94.625	94.125	94.625	93.375	**95.25**
14	**99.875**	**100**	92.5	**100**	81	**100**
15	2.5	14.25	1.5	14	1.125	1.625
16	78	**39.875**	**87.875**	38.75	13.75	16.375
17	**93**	91.75	90.125	91.625	74	**93.125**
18	**90**	90.625	89.875	**90.75**	89.125	89.75
19	66.125	14.25	**89.25**	12.875	0.5	**29.375**
20	**81.75**	**58**	73.875	57.125	32	44.875
21	**36.5**	46.75	8.75	**47.75**	33.625	45.75
Average	**85.847**	**77.139**	78.451	77.035	57.535	71.986

Table2 Average FAR (%) of different method on TE process

	NGLFPP		**GLPP**		**PCA**	
Fault num	T^2	Q	T^2	Q	T^2	Q
Average	0.625	4.028	0.729	4.097	0.486	0.833

The fault detection rates (FDRs) of three monitoring methods for TE process are shown in Table1. Previous studies have shown that faults 3, 9 and 15 are difficult to detect because of their small magnitudes. The monitoring performances of the rest 18 faults is compared and their average values are calculated. The bolded numbers in Table1 represent the best FDR for the corresponding statistic under each fault condition. For T^2 statistics, the proposed method gives the highest detection rate for 13 of 18 faults, and significant improvement for fault 4, 7 and 14. For Q statistics, the proposed method gives similar FDRs to GLPP-based method and higher FDRs compared to the PCA-based method on 10 of 18 faults, especially on faults 8,10,12. The performance of the proposed method can be further seen by the average FDR of 18 faults given in Table 1. False alarm rate (FAR) is also an important indicator for evaluating the monitoring performance, and average FARs of three methods for 18 faults are shown in Table2. Obviously, the proposed method has a lower average FAR for T^2 statistics. Similar to GLPP-based method, the proposed method has a slightly higher average FAR for Q statistics compared with PCA-based method. In general, the proposed method has better performance than PCA-based and GLPP-based methods.

5. Conclusion

In this paper, a novel global-local feature preserving projection method is proposed. Considering the difference of local distribution, the density-based adaptive neighbourhood construction method is applied to determine local and global scopes of data. On this basis, a modified neighbourhood tangent estimation method is integrated into LLTSA to preserve the local feature of the data accurately. Furthermore, global feature of data based on geodesic distance is obtained. The above features of data are extracted simultaneously by constructing a dual objective optimization function. Therefore, the proposed method is more comprehensive in feature extraction of data, and gives better performance in process monitoring on TE process, compared to PCA-based and GLPP-based monitoring methods.

References

J. J. Downs, E. F.Vogel, 1993, A plant-wide industrial process control problem, Computers & chemical engineering, 17, 3, 245-255.

W. Ji, H. Yang, 2019, Soft sensor modeling based on adaptive Isomap algorithm, Journal of Nanjing University of Science and Technology, 43, 3, 269-274.

L. Luo, 2014, Process Monitoring with Global–Local Preserving Projections, Industrial & Engineering Chemistry Research,53, 18, 7696-7705.

J. M. Lee, S. J. Qin, I. B. Lee, 2006, Fault detection and diagnosis based on modified independent component analysis, AIChE Journal, 52, 3501−3514.

Y. Ma, B. Song, H. Shi, Y. Yang, 2015, Fault detection via local and nonlocal embedding, Chemical Engineering Research and Design, 94, 538-548.

S. J. Qin, L. H. Chiang, 2019, Advances and opportunities in machine learning for process data analytics, Computers & Chemical Engineering, 126, 465-473.

P. Zhang, H. Qiao, B. Zhang, 2011, An improved local tangent space alignment method for manifold learning, Pattern Recognition Letters, 32, 2, 181-189.

T. Zhang, J. Yang, D. Zhao, X. Ge, 2007, Linear local tangent space alignment and application to face recognition, Neurocomputing, 70, 7-9, 1547-1553.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50223-2

Prognostics on Noisy and Uncertain Dynamic Systems using Cumulative Sum Chart of Inferential Sensors

Efi Safikou[a], and George M. Bollas[b*]

[a] *Department of Electrical & Computer Engineering, University of Connecticut, 371 Fairfield Way, Unit 4157, Storrs, CT, 06269, USA*
[b] Department of Chemical & Biomolecular Engineering, University of Connecticut, 191 Auditorium Road, Unit 3222, Storrs, CT, 06269, USA, george.bollas@uconn.edu

Abstract

Inferential sensing techniques have been proven very successful in the field of fault detection and isolation, as they yield robust outcomes in a cost-effective and reliable manner. Herein, we propose a framework that deploys inferential sensors in the presence of noise and uncertainty, while focusing on the detection time and early sensitivity to faults. The most informative inferential sensor is derived through symbolic regression with an objective function that uses optimality criteria from information theory, wherein the sensitivity of the inferential sensors with respect to faults and uncertainty is estimated using the system digital twin. For deployment of the inferential sensors, the Cumulative Sum Control Chart method is employed and tuned to monitor deviations from the anticipated system performance. The proposed method is applied for the detection of faults in a crossflow plate-fin heat exchanger, at various levels of measurement noise and uncertainty, under transient operation. When compared to existing (hard) sensors, the inferential sensor provides intelligible deviations from the "fault-free" system response, thus enabling accurate and robust estimates for the initiation and progression of faults.

Keywords: symbolic regression; inferential sensing; prognostics; diagnostics.

1. Introduction

Diagnostics and prognostics are key components of cyber-physical system digital twins to enable effective system monitoring and ensure safety (Vogl et al., 2019). The former encapsulates the process of fault detection and isolation based on prior knowledge of the entire system operation, thus generating an anticipated response that can be attributed to the failure mode. The latter, provides a forecast of future system conditions and potential failure timelines, given only present or historic information (Biggio et al., 2020). While diagnostics is an extensively studied field with a plethora of documented analyses and methods (Palmer & Bollas, 2019), research on prognostics coupled with inferential sensing techniques is limited and more challenging (Adams, 2019). Typically, inferential sensors combine available system inputs and outputs in either analytical expressions, or empirical relationships based on data (Hale & Bollas, 2020). Herein, we employ inferential sensing for the diagnosis and prognosis of faults in the presence of noise and uncertainty, focusing on the time for detection and early sensitivity to faults. The Cumulative Sum Control Chart (CUSUM; Montgomery, 2009) method is used and tuned to monitor present and future deviations from the anticipated system performance to determine potential failure initiation instances.

2. Methods

2.1. System model representation

We assume the existence of an accurate dynamic system model for the system, expressed as a set of differential algebraic equations that are based on physical system knowledge or semi-empirical correlations:

$$\mathbf{f}(\dot{\mathbf{x}}(t), \mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}, t) = 0, \tag{1}$$

where $\mathbf{f}$ is the system governing equations, $\mathbf{x}(t)$ the vector of state variables, $\dot{\mathbf{x}}(t)$ the time derivatives, $\mathbf{u}(t)$ are the admissible inputs, $\boldsymbol{\theta}$ the model parameters, and t denotes the time. All model parameters remain invariant for a timespan, $\boldsymbol{\tau}$, within which Eq.(1) is accurate for steady-state and dynamic operation. Let $\hat{\mathbf{y}}$ be the measured outputs of the system as a function of states, inputs, and parameters, whose N_y components correspond to the available hard sensors: $\hat{\mathbf{y}} = \mathbf{h}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}, t)$. We define the inferential sensors, $\hat{\mathbf{z}}$, as standardized functions of $\hat{\mathbf{y}}$ and $\mathbf{u}(t)$ based on the nominal value, $\hat{\mathbf{z}}_{\text{nom}} = \mathbf{g}(\bar{\mathbf{y}})$, with $\bar{y} = \hat{y}/\hat{y}_{\text{nom}}$. The vector of nominal values for the system outputs, $\hat{\mathbf{y}}_{\text{nom}}$, contains output values corresponding to predetermined values for the admissible system inputs, $\mathbf{u}$. A $(N_y + N_z)$-dimensional binary vector, $\mathbf{a}$, can be introduced to define active and inactive sensors, depending on their use for a particular diagnostics or prognostics test. Note that N_z denotes the number of inferential sensors. For inactive sensors the respective index in $\mathbf{a}$ is equal to 0, while the positions of active sensors correspond to elements equal to 1. The vector of the system inputs $\mathbf{u} = [\mathbf{u}_p(t), \mathbf{u}_q]$ contains the uncertain inputs, $\mathbf{u}_q$, as well as the controllable system inputs, $\mathbf{u}_p(t) = [\mathbf{u}_p^1, \mathbf{u}_p^2, \ldots, \mathbf{u}_p^{N_k}]$, with N_k denoting the number of changes in u_p during the examined period (i.e., the number of tests). Considering $\mathbf{u}_p^{[k]}$ that obtains acceptable values from a continuous test design space, $\mathbf{U}_p$, we can design tests $k \in \{1, \ldots, N_k\}$. The vector of system parameters, $\boldsymbol{\theta}$, can be divided in three components, $[\boldsymbol{\theta}_f, \boldsymbol{\theta}_q, \boldsymbol{\theta}_p]$, where $\boldsymbol{\theta}_f$ are the parameters that represent faults, $\boldsymbol{\theta}_q$ the parameters related to system uncertainty, and $\boldsymbol{\theta}_p$ the known and invariant parameters that symbolize the system design. It follows that the faults, the parameters related to uncertainty, as well as the system inputs, are concatenated into a N_ξ-dimensional vector, $\boldsymbol{\xi} = [\boldsymbol{\theta}_f, \boldsymbol{\theta}_q] \cup [\mathbf{u}_q] = [\boldsymbol{\xi}_f, \boldsymbol{\xi}_q]$, whose elements are separated into faults, $\boldsymbol{\xi}_f$, and uncertainty, $\boldsymbol{\xi}_q$. Note that $\boldsymbol{\xi}$ is later employed during the optimization of the test design. We assume that the anticipated fault parameter values, $\tilde{\boldsymbol{\xi}}_f$, can be acquired from historical data of system components or are simply known a priori from past system operation. Moreover, faults are considered the only indicators of undesirable system performance, with the uncertainty parameters being part of the normal operation that should not lead to false alarms. Let $l \in \{1, \ldots, N_l\}$ denote the faults scenarios, with N_l being the total number, and $\tilde{\boldsymbol{\xi}}_f^{[l]}$ the anticipated fault scenarios. Then, each fault scenario, l, corresponds to a different fault level in $\boldsymbol{\xi}_f$, with the uncertainty vector $\boldsymbol{\xi}_q$ remaining constant and containing the mean values of the uncertain system parameters and inputs.

2.2. Active FDI test design optimization

Consider a vector $\boldsymbol{\phi}^{[k]}$ that contains a set of continuous variables $\mathbf{u}_p^{[k]}$, and a binary vector $\mathbf{a}$ related to sensor selection. Then, for all N_k, this design vector can be formulated as $\boldsymbol{\phi} = [\mathbf{u}_p^1, \ldots, \mathbf{u}_p^{N_k}, \mathbf{a}]$. The optimal FDI test design, $\boldsymbol{\phi}^*$, is the product of a mixed-integer non-

linear optimization program, which is founded on the maximization of the estimability of the faults as functions of the sensitivity of active sensors with respect to faults (Palmer et al., 2016) and, specifically, the minimization of the joint confidence between faults and faults with uncertain parameters. For dynamic tests (in which the system transient is monitored and used for FDI), the optimization problem is then as shown in Eq.(2):

$$\boldsymbol{\phi}^* = \left[\mathbf{u}_p^*(\boldsymbol{\tau}_1), \ldots, \mathbf{u}_p^*(\boldsymbol{\tau}_{N_k}), \mathbf{a}\right] \in \arg\max_{\boldsymbol{\phi}\in\boldsymbol{\Phi}} \ln \Psi_{D_S}$$

s.t.

$$\begin{aligned} &\mathbf{f}\left(\dot{\mathbf{x}}(t), \mathbf{x}(t), \mathbf{u}_p(t), \boldsymbol{\theta}_p, \tilde{\boldsymbol{\xi}}, t\right) = \mathbf{0}, \\ &\hat{\mathbf{y}} = \mathbf{h}\left(\mathbf{x}(t), \mathbf{u}_p(t), \boldsymbol{\theta}_p, \tilde{\boldsymbol{\xi}}, t\right), \\ &\mathbf{x}^{\mathrm{L}} \le \mathbf{x} \le \mathbf{x}^{\mathrm{U}}, \forall\, t \in \boldsymbol{\tau}, \\ &\mathbf{u}_p \subset \mathbf{U}_p, \forall\, t \in \boldsymbol{\tau}, \end{aligned} \tag{2}$$

where $\mathbf{u}_p^*(\boldsymbol{\tau}_k)$ denotes the optimal inputs for each test of timespan $\boldsymbol{\tau}_k$. The objective function selected for the optimization problem is the D_s-optimality criterion, Ψ_{D_s}, that is based on the Fisher Information Matrix (FIM), which encapsulates the existing knowledge about the faults and system uncertainty; see Eq.(3). The FIM, $\mathbf{H}_{\xi}$, can be evaluated as the resultant of three secondary matrices that correspond to $\boldsymbol{\xi}_f$ and $\boldsymbol{\xi}_q$; see Eq.(4). The foregoing submatrices $\mathbf{H}_{ff}$, $\mathbf{H}_{fq}$, and $\mathbf{H}_{qq}$, represent the covariance between (i) faults, (ii) faults and system uncertainty, and (iii) system uncertainty, respectively.

$$\mathbf{H}_{\xi}^{[l]}\left(\boldsymbol{\phi}, \tilde{\boldsymbol{\xi}}\right) = \begin{bmatrix} \mathbf{H}_{ff}^{[l]} & \mathbf{H}_{fq}^{[l]} \\ \mathbf{H}_{qf}^{[l]} & \mathbf{H}_{qq}^{[l]} \end{bmatrix} = \frac{\sum_{k=1}^{N_k} \sum_{i=1}^{N_y+N_z} a_i \sigma_i^{-2} Q_i^{[k,l]^T} Q_i^{[k,l]}}{\sum_{i=1}^{N_y+N_z} a_i} \tag{3}$$

$$\mathbf{H}_{ff}^{[l]} \in \mathrm{R}^{N_f \times N_f},\ \mathbf{H}_{fq}^{[l]} \in \mathrm{R}^{N_f \times \left(N_\xi - N_f\right)},\ \mathbf{H}_{qf}^{[l]} = \left(\mathbf{H}^{[l]}\right)^{\mathrm{T}}_{fq},\ \mathbf{H}_{qq}^{[l]} \in \mathrm{R}^{\left(N_\xi - N_f\right)\times\left(N_\xi - N_f\right)} \tag{4}$$

where N_f denotes the number of fault parameters, and σ_i^2 is the measurement variance of the i^{th} output. The sensitivity matrix, $\mathbf{Q}$, consists of the partial derivatives of the active hardware or/and inferential sensors, with respect to $\boldsymbol{\xi}$ (Palmer & Bollas, 2019). In more detail, the D_s-optimality criterion of Eq.(5), minimizes the covariance between the entire set of fault parameters, uncertain parameters, and inputs, towards increased fault isolation. Enhanced detection is achieved through the neglection of the covariance between the elements of $\boldsymbol{\xi}_q$. Note that summation over all possible fault scenarios N_l enables designs for different faults or fault severity levels, including the healthy state.

$$\Psi_{D_s}\left(\mathbf{H}_{\xi}\right) = N_l^{-1} \sum_{l=1}^{N_l} \log\left|\mathbf{H}_{ff}^{[l]} - \mathbf{H}_{fq}^{[l]}\left(\mathbf{H}_{qq}^{[l]}\right)^{-1}\mathbf{H}_{qf}^{[l]}\right|^{1/N_\xi}. \tag{5}$$

2.3. Cumulative Sum control chart

The CUSUM is a type of control chart, which constitutes a particularly effective tool when small process shifts are of interest. It calculates the cumulative sums of the sample values deviations, by taking into consideration all the information in the respective sequence (Montgomery, 2009). The cumulative sums of deviations from a target are plotted in the CUSUM chart, which can be represented in a tabular form that provides an ease of interpretation and a robust control limit setting. Two statistics, namely the upper, C_i^+, and the lower, C_i^-, CUSUM charts, are constructed in the tabular CUSUM; see Eq.(6). Deviations above the target are accumulated in the C_i^+ statistic, while sub-target inconsistencies are collected in the C_i^- statistic. Along these lines, we have:

$$C_i^- = \min\{0, C_{i-1}^- + x_i - \mu + K\}, \text{ and } C_i^+ = \max\{0, C_{i-1}^+ + x_i - \mu - K\}, \quad (6)$$

where μ denotes the target value, while the starting values C_0^- and C_0^+ are equal to 0. In case one of the C_i^+ and C_i^- violates the control limits set by the decision interval, H, the process is assumed out-of-control. Additionally, the reference value, K, must be selected properly to achieve an adequately quick detection, based on the respective out-of-control value (i.e., fault). It is worth mentioning that the critical tuning parameters H and K regulate the sensitivity of the CUSUM chart. Herein, the target value μ corresponds to the fault-free condition of the system, and K is set equal to σ, where σ is the standard deviation of the incorporated sensors. For H we assume a typical value equal to 5σ.

2.4. Genetic Programming

Genetic Programming (GP) is employed to discover functional relationships between the incorporated variables (i.e., symbolic regression). In each generation, the most informative inferential sensor is pursued and then maintained until a better one is derived. The D_s-optimality criterion incorporating solely inferential sensors was selected as the objective function in this study, although fusion of hard and inferential sensors is also feasible. In that sense, the program of Eq. (3) is updated to include the inferential sensors as the optimization variables, $\boldsymbol{\phi}_{\mathbf{GP}} = \mathbf{z}$, which is solved for the FDI test design calculated from Eq. (3). After the most informative inferential sensor is obtained, it is employed in Eq. (3) as a new sensor and the FDI test design $\mathbf{u}_p$ is re-optimized. This operation iterated between FDI test design optimization and inferential sensor evolution (with D_s-optimality being the consistent objective) until the test design and functional form of the inferential sensor remain the same in consecutive iterations, when the process is terminated. Otherwise, the GP algorithm is trained with the new optimal system input, until convergence is achieved. For the calculation of the sensitivities of the inferential sensors, the chain rule of partial derivatives of automatic differentiation is applied, making the calculation of D_s-optimality in the symbolic regression step computationally inexpensive.

3. Case Study

The proposed framework was applied on a crossflow plate-fin heat exchanger (PFHE) system (Palmer et al., 2016), where the measured outputs correspond to temperatures and pressures of the outlet streams, $\mathbf{y} = [T_c, T_h, P_c, P_h]$, and the system admissible input is the mass flow rate of the hot stream, $u_p = \dot{m}_{h,i}$ (kg/s). The PFHE fault studied is thermal fouling resistance in the cold stream side of the PFHE, $\theta_{f=1} = R_f$, subject to uncertainty that includes the cold air inlet stream moisture content, $\omega_{H2O} \sim N(7, 4)$, and the cold air inlet temperature, $T_{in}^{cold} \sim N(35, 25)$. The FDI test design and inferential sensors were optimized for three fault scenarios; namely: (i) a fault free case (R_f = 0.4), (ii) 20% blocked fouling (R_f = 1.6), (iii) 50% blocked fouling without error (R_f = 4). The FDI design is aware of the model parameters that represent uncertainty and the standard deviation of the hard sensors, but not the exact instantiations of uncertainty and noise that the system may exhibit. In non-exact terms, the goals of FDI test design and inferential sensor optimization are to minimize the covariance between fault parameters, and fault parameters and uncertainty. Then, to investigate system performance in various operating scenarios, we examine two cases of PFHE performance, in which we injected uncertain parameters $[\omega_{H2O}, T_{in}^{cold}]$= [5.6, 46], and $[\omega_{H2O}, T_{in}^{cold}]$= [12, 35], respectively. In both, we also inject a linear increase in the PFHE fouling that starts at t = 100 s (R_f = 0.4) and lasts for Δt = 360 s (R_f = 4).

4. Results – Discussion

For brevity, in Fig. 1 we present the outcomes of the CUSUM chart for a single hard sensor (i.e., T_c). As illustrated in Figs. 1a and 1c, for the first case described in Section 3, a fault is erroneously detected by CUSUM at time $t = 3$ s, due to the instantiation of uncertainty and noise. Figs. 1b and 1d (i.e., for the second case) show that in a different uncertainty scenario, the system is out-of-control at $t = 288$ s, when the fouling level is close to 20% - blocked. The required time for fault detection is 188 s. The foregoing observations manifest that CUSUM is unpredictable when employed solely with hard sensors to detect system faults in the presence of noise and uncertainty.

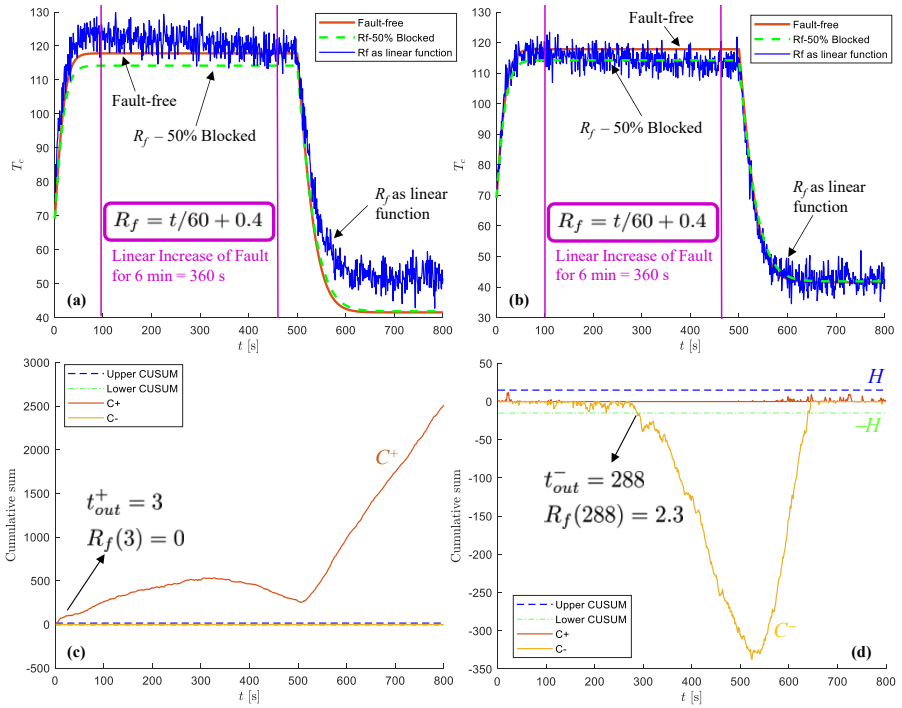


Figure 1. Hard sensor timeseries for linearly increasing fault, in the cases of (a) $[\omega_{H2O}, T_{in}^{cold}]$= [5.6, 46], and (b$[\omega_{H2O}, T_{in}^{cold}]$=) [12, 35], with (c, d) the respective results of the CUSUM chart.

The first examined case seems more challenging for CUSUM, thus is adopted for the implementation of the optimal inferential sensor, obtained by using the same system outputs as the independent variables of symbolic regression. The functional form of the D_s-optimal inferential sensor is shown in Eq.(7) and it is notably simple and explainable. The inferential sensor uses the ratio of outlet temperatures as the key driver to decrease the evidence of uncertainty and improve the estimability of the fault:

$$Z_{Ds} = [(T_c - 4.7440)/T_c + (T_c/2)/T_h]/1.6556 \quad (7)$$

In Fig. 2, we can observe that the inferential sensor detects the existence of a fault at time $t = 362$ s (i.e., after 262 s). The value of the fault at that time is $R_f = 3$ (see Fig.2b), which is slightly larger than 20% blocking of the PFHE (i.e., $R_f = 1.6$) , but smaller than the value of 50% blocked that constitutes the maximum assessed fault level (i.e., $R_f = 4$). While there is noise in its sequence (see Fig. 2a), the inferential sensor reduces the

deviations induced by epistemic uncertainty, thus enabling the robust and effective detection of faults in the dynamic test. It is important to note that the CUSUM chart method is more effective when small shifts in the sensor sequence take place. However, the fluctuations when using solely hard sensors are erratic; see also Fig. 1a. In contrast, the incorporation of an inferential sensor provides much smoother deviations (see Fig. 2a), thus making them particularly suited for use along with CUSUM charts. This was also consistent for the fault and uncertainty scenario studied in Figs. 1b and 1d, where the inferential sensor produced no false alarms and detected the PFHE fouling at $t = 200$ s, (fouling level under 20%), earlier than the most informative hard sensor, T_c.

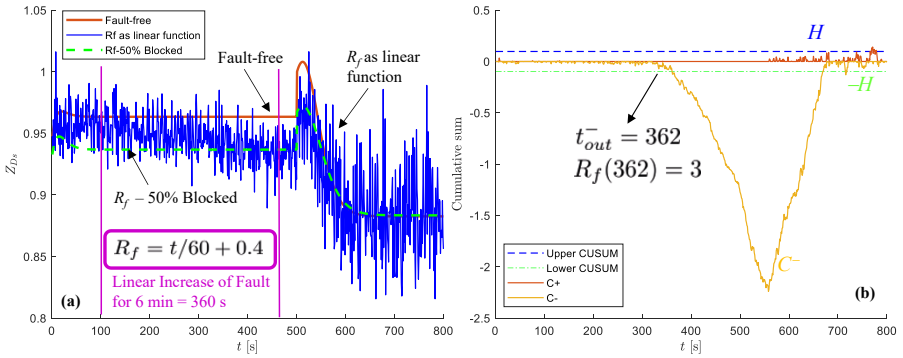


Figure 2. (a) The timeseries of the inferential sensor, with a linearly increasing fault. (b) The respective tabular form of the CUSUM chart.

5. Conclusions

We showed that the CUSUM chart method is effective for FDI when small deviations from the target sequence (i.e., fault-free scenario) occur. The use of CUSUM, however, solely with hard sensors induces erratic sequence fluctuations, thus hindering accurate fault detection. On the other hand, the employment of inferential sensors via the proposed framework, was shown to mitigate such issues. Therefore, we argue that the combination of CUSUM charts and inferential sensing techniques can constitute a robust tool for fault diagnosis and prognosis.

References

Biggio, L., & Kastanis, I. (2020) Prognostics and Health Management of Industrial Assets: Current Progress and Road Ahead, *Front Artif Intell*, **3**, 10.3389/frai.2020.578613.

Vogl, G.W., Weiss, B.A., & Helu, M. (2019) A review of diagnostic and prognostic capabilities and best practices for manufacturing, *J Intell Manuf*, **30**, 79-95, 10.1007/s10845-016-1228-8.

Hale, W.T., & Bollas, G.M., (2020) Least-squares- and information-theory-based inferential sensor design for fault diagnostics, *ACC 2020*, 3182–3187.

Adams, Van H. (2019) Data Prognostics Using Symbolic Regression, engrXiv, 10.31224/osf.io/fq8ze.

Palmer, K.A., Hale, W.T., Such, K.D., Shea, B.R., & Bollas, G.M. (2016) Optimal design of tests for heat exchanger fouling identification, *Appl Therm Eng*, **95**, 382-393, 10.1016/j.applthermaleng.2015.11.043.

Montgomery, D.C. (2009) *Introduction to Statistical Quality Control*, 6th edition, Wiley & Sons, 9780470169926.

Palmer, K.A., & Bollas, G.M. (2019) Sensor selection embedded in active fault diagnosis algorithms, 505 IEEE Trans Control Syst Technol, 1–14.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50224-4

Quantifying Subsea Gas Leakages using Machine Learning: a CFD-based study

Gustavo L.R. Caldas[a*], Thiago F.B. Bento[b], Roger M. Moreira[b] , Maurício B. de Souza Jr.[a]

[a] *Escola de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro 21941-909, Brazil*
[b] *Escola de Engenharia, Universidade Federal Fluminense, Niterói 24210-240, Brazil*
gustavo.caldas@eq.ufrj.br

Abstract

Subsea images captured on-site can be used to quantify gas leakage in the subsea environment. In this work, gas leakage in reduced conditions was simulated by Computational Fluid Dynamics (CFD). The aim is to develop a computational vision tool to quantify the leakage. The images generated from CFD simulations were processed by a convolutional neural network (CNN) structure, the U-Net. A class is attributed to each image pixel, and a post-processing algorithm computes the corresponding bubble area. Two cases were carried out: image segmentation into two (water and bubble) and three classes (bubble interface included). The multi-class U-Net shows a good agreement with CFD results compared to the binary one because separating the pixels into just two categories leads to bubble diameter overestimation. Hence, this method is of potential use in fault detection and diagnosis and could support the decision-making process on deepwater leakage remediation.

Keywords: gas leakage; machine learning; convolutional neural network; process safety.

1. Introduction

Subsea oil and gas activities demand safety procedures and constant monitoring to prevent impact on marine ecosystems and financial losses for the operating companies (Figueredo et al., 2022). Several resources might take hold for this purpose. For instance, real-time leakage filming is possible with the Remotely Operated Vehicles (ROV) equipped with a camera onboard. These images, however, provide information only on whether the leak is occurring. For a better assessment, it is of great interest to develop a quantitative tool to support the decision-making process of intervention.

A possible parameter for the leak estimation is the bubble diameter (Jamialahmadi et al., 2001), which could be computed using image processing techniques. More recently, convolutional neural networks (CNN) - a type of Machine Learning (ML) algorithm – became part of these techniques (Goodfellow et al., 2016). CNNs are sparsely connected neural networks, i.e., not all neurons are connected to the ones of the subsequent layers. As a result, it saves plenty of computational resources when dealing with tensor data such as images and sounds (Krizhevsky et al., 2012). In a CNN structure, the first argument is the input, and the second one, the kernel (filter). Typically, the input is a tensor containing the image height, width, and input channels (colors). The output is called the feature map, which stores the characteristics of the

input data and simultaneously reduces its size by using a kernel smaller than the matrix – this is the reason for the sparse connectivity. The kernels' number, shape, and activation function are hyper-parameters defined by the user (Goodfellow et al., 2016).

Convolutional neural networks have already been applied to fault detection and diagnosis problems. Wu and Zhao (2018) verified its usefulness on the Tennessee Eastman process. The relation between different process variables and sampling time is concatenated into two-dimensional matrices, adequate for CNN computing. The fault diagnosis rate scored 88.2 %. Li et al. (2018) proposed a CNN to detect chemical leakage in hydrocarbon tanks based on image recognition. They obtained 85.82 % accuracy. Bai et al. (2021) developed a real-time classifier of gas dispersion state in a bubble column using a novel CNN architecture named BubbleNet. It differentiated flow conditions according to bubbles' size and shape after being trained to a labeled dataset. It scored 97.8 % and 97.5 % of the performance for the training and test, respectively.

In some chemical engineering applications with multi-phase flows, e.g., liquid-liquid extraction, it is fundamental to know the particle size distribution, a variable of interest for the transport phenomena control. Schäfer et al. (2019) investigated it using a particular convolutional neural network, the U-Net. This network was designed for image segmentation (Ronneberger et al., 2015), an application interested in localizing objects and boundaries by partitioning the image pixels into various segments. Thus, the U-Net permits phase fractions distinction. Another advantage is that post-processing enables the calculation of the droplet size distributions from the U-Net output. Therefore, the present study aims to develop a system capable of quantifying leakages in subsea processes employing the U-Net convolutional neural network.

2. Methodology

We carried out reduced model simulations of gas leakages employing Computational Fluid Dynamics (CFD). Reduced model is a technique that is used to save computational costs by downscaling the original phenomenon. For instance, it reproduces an event from the subsea scale to the laboratory. Gas leakages are released with different velocities (v) and from different orifice diameters (d). The initial value problem is solved via a finite volume method. The Volume of Fluid (VoF) method is employed to model the two-phase gas-liquid flow. Continuity and the unsteady RANS (Reynolds-Averaged Navier-Stokes) equations are satisfied in the fluid domain, with the classical κ-ε turbulence model being used. The CFD results are being validated with experiments and semi-empirical models. They agree on the trend found in the literature (Jamialahmadi et al., 2001). The simulation was carried out in ANSYS Fluent software, producing videos that represent the leakage. Each video frame generated an image set, totalizing 3159 images from the different conditions.

In a second step, the images are forwarded to a CNN model, called the U-Net structure. The main goal of this architecture is to classify each pixel individually as belonging to some class. The images are the input for training this network, and the targets are the masks created by a segmentation method. The CNN was developed in Keras environment employing Python with Tensorflow as backend. The segmentation was carried out using the unsupervised Otsu's methodology (Otsu, 1979) in the Scikit Image library written in Python. It is an algorithm whose aim is to find a threshold that can divide the pixels of a grayscale image into two clusters (classes): foreground f and background b. A threshold t is searched, such that the intra-class variance, represented

in Eq. 1, is minimized (and the inter-class is maximized as well). The weights ω calculated contain the probabilities of a pixel to belong to one of the classes. In this case, classes are water (label zero) and bubble (label one). Given the importance of phase fraction when accounting bubble diameter, the problem was extended to multi-segmentation, in which the interface is labeled as number two. Multi-level thresholding can be performed as described by Otsu (1979).

$$\sigma^2_{w(t)} = \omega_b(t)\sigma^2_b(t) + \omega_f(t)\sigma^2_f(t) \tag{1}$$

The U-Net structure is shown in Figure 1. It is composed of a down-sampling part: successive blocks of convolutional 2D layers with filters of window dimension 3x3 and initialization "He" followed by 20 % dropout; a second convolutional layer; and a max pooling layer, which takes the maximum value over the window 2x2. In the next block, the number of filters is doubled (starting with 32). The second part comprises the up-sampling operations: transposed convolution (deconvolution) layers with filters 2x2 and stride 2x2. Information is concatenated from the corresponding feature maps of convolutional and deconvolutional layers. Another two convolutional layers are present on each block with half of the filters from the previous up-sampling block. The batch size is 128.

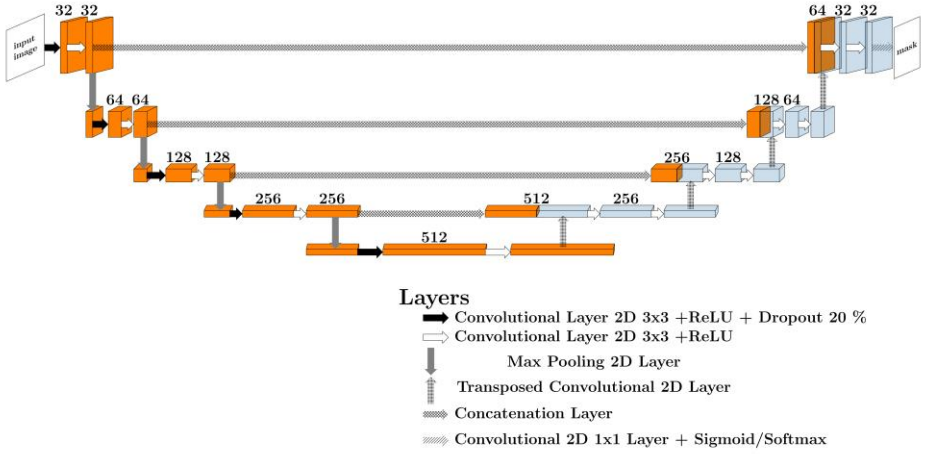


Figure 1: The U-Net architecture.

The total number of parameters for the binary class and the multi-label segmentation problems are 7,759,521 and 7,759,587, respectively. The metric used in this case was the Dice-Sørensen coefficient (Eq. 2a), which computes the similarity between the actual and predicted samples in relation to the group. It is important to use one-hot encoding format for the multi-class problem. Thus, the dice coefficient is extended for each class C (Eq. 2b). For the one-hot encoding format, the categorical cross-entropy (Eq. 3) was employed as a loss function to be minimized.

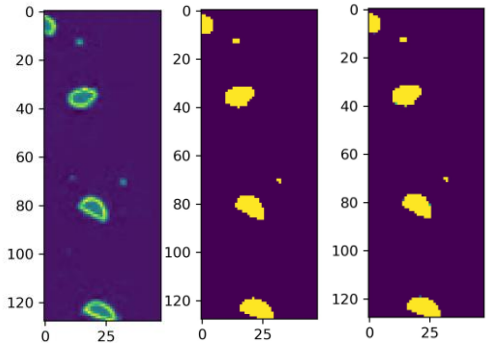
$$DSC = \sum_{n=1}^{N} \frac{2\sum_{j=0}^{128}\sum_{i=0}^{48} \hat{y}_{ij,n} * y_{ij,n}}{\sum_{j=0}^{128}\sum_{i=0}^{48} \hat{y}_{ij,n} + \sum_{j=0}^{128}\sum_{i=0}^{48} y_{ij,n}} \tag{2a}$$

$$DSC = \sum_{c=1}^{C} \frac{\mathrm{DSC_c}}{\mathrm{C}} \tag{2b}$$

$$CCE = -\frac{1}{N}\sum_{n=1}^{N}\sum_{j=0}^{128}\sum_{i=0}^{48}\sum_{c=1}^{C} y_{c,ij,n}\, log\, \hat{y}_{c,ij,n} \tag{3}$$

3. Results and Discussion

Figure 2 shows an image sample (U-Net input), the corresponding binary segmentation mask, and the mask predicted. Bubble statistics are presented in Table 1.



(a) (b) (c)

Figure 2: Snapshot of a sample: (a) U-Net input. (b) Mask generated by the binary Otsu's thresholding (U-Net target). (c) Mask predicted by the binary U-Net.

Table 1: Binary U-Net: predicted $\hat{d}_b$ against expected numerical diameter d_b.

d (mm)	v (m/s)	d_b (mm)	count	$\hat{d}_b$ (mm) mean	$\hat{d}_b$ (mm) std. dev.
0.5	0.25	6.45	350	7.82	0.36
0.5	0.625	7.22	369	9.79	1.16
0.5	1.0	7.69	374	10.19	1.95
1.0	0.24	6.37	317	9.44	0.94
1.0	0.37	6.46	311	9.95	1.72
1.0	0.5	7.28	377	11.74	2.01
5.0	0.02	6.70	373	8.67	0.37
5.0	0.055	8.05	308	10.51	0.96
5.0	0.09	8.11	380	12.14	2.30

The U-Net output is very similar to the target as the Dice-Sørensen coefficients for training and validation imply: 0.9915 and 0.9888, respectively. An overestimation is reported when comparing the expected numerical diameter d_b with the predicted one ($\hat{d}_b$). The reason is that the binary Otsu's thresholding does not set apart the interface and the bubble. This factor influences the area for calculation.

Due to the overestimation, it was decided to investigate further and add a phase fraction, turning the problem into a multi-class one. Figure 3 shows the analog result to Figure 2. The resulting mask resembles much more to the original image when compared to the previous case. The training was also successful. The multi-dice coefficient for training and validation reported 0.9507 and 0.9573, respectively. Similarly, the categorical cross-entropy loss found was 7.14 x 10^{-3} (training) and 5.34 x 10^{-3} (test). The predicted diameter by the multi-class U-Net shows a good agreement with the expected numerical diameter, as statistics shown in Table 2. Low standard deviations suggest that the biggest bubbles are relatively uniform for each dataset. Deviations from the actual values do not exceed 10 %, except for the 1.0 mm diameter crack cases.

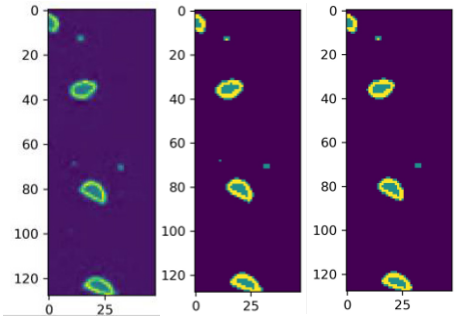


Figure 3: Snapshot of a sample: (a) U-Net input. (b) Mask generated by the multi-label Otsu's thresholding (U-Net target). (c) Mask predicted by the multi-class U-Net.

Table 2: Multi-class U-Net: predicted $\hat{d}_b$ against expected numerical diameter d_b.

d (mm)	v (m/s)	d_b (mm)	count	$\hat{d}_b$ (mm) mean	$\hat{d}_b$ (mm) std. dev.
0.5	0.25	6.45	350	6.13	0.34
0.5	0.625	7.22	369	7.23	0.90
0.5	1.0	7.69	374	7.41	1.23
1.0	0.24	6.37	317	7.13	0.73
1.0	0.37	6.46	311	7.34	1.22
1.0	0.5	7.28	377	8.39	1.37
5.0	0.02	6.70	373	6.67	0.32
5.0	0.055	8.05	308	7.68	0.69
5.0	0.09	8.11	380	8.66	1.74

Regarding the model convergence, each model's total training time lasted around 1h30 min (2.5 - 3 min/epoch) in an Intel Core i5-10210. It has achieved less than 0.10 of loss in the fifth epoch, and after 15 epochs, more than 0.90 of Dice similarity coefficient.

4. Conclusions

A novel methodology was presented to quantify gas leakages that can be applied in a subsea environment, combining convolutional neural networks and a segmentation tool. The U-Net enabled the multi-segmentation post-processing to reach good predictability of the bubble diameter (less than 10 % deviation in general, the worst case was 15.24 % deviation). It is noteworthy that this performance was achieved with a relatively low amount of data (3159). For future works, the validation of the methodology is in progress by exposing the CNN to experimental data.

Acknowledgments

[a]Financial support from the Human Resources Program of the National Agency of Petroleum, Natural Gas, and Biofuels – PRH-ANP, funded by oil companies qualified in the R&D and Innovation clause from ANP Resolution 50/2015 (process no. 041319). [b]Financial support from PETROBRAS (process no. PETROBRAS/UFF/FEC 5900.0113248.19.9).

References

C. Bai, H. Park, Yong, C.Y. Ng, L. Wang, 2021, Classification of gas dispersion states via deep learning based on images obtained from a bubble sampler, Chemical Engineering Journal Advances, 5, 100064

A. K. M. Figueredo, M. B.de Souza, P. F. F. Frutuoso e Melo, C. A. Vaz Junior, J. Di Domenico, 2022, How does subsea processing safety policy impact safety? Lessons learned from the Brazilian experience, Journal of Loss Prevention in the Process Industries, 74, 104668

I. Goodfellow, Y. Bengio, A. Courville, 2016, Deep Learning, MIT Press, 775p.

M. Jamialahmadi, M.R. Zehtaban, H. Müller-Steinhagen, A. Sarrafi, J.M. Smith, 2001, Study of Bubble Formation Under Constant Flow Conditions, Chemical Engineering Research and Design, 79, 5, 523-532

A. Krizhevsky, I. Sutskever, G.E. Hinton, 2012, Image net classification with deep convolutional neural networks, Advances in Neural Information Processing Systems, 25, 1097–1105

L. Li, X. Jia, W. Tian, S. Sun, W. Cao, 2018, Convolution Neural Network based Chemical Leakage Identification, Computer Aided Chemical Engineering, 44, 2329 – 2334.

N. Otsu, 1979, A Threshold Selection Method from Gray-Level Histograms. In: IEEE Transactions on Systems, Man, and Cybernetics, 62, 2168-2909.

O. Ronneberger, P. Fischer, T. Brox, 2015, U-Net: Convolutional networks for biomedical image segmentation, In: Medical Image Computing and Computer-Assisted Intervention – Lecture Notes in Computer Science, 9351, 234–241, Springer

J. Schäfer, P. Schmitt, M.W. Hlawitschka, H. Bart, 2016, Measuring particle size distributions in multiphase flows using a convolutional neural network, Chemie Ingenieur Technik, 91, 11, 1688–1695

H. Wu, J. Zhao, 2018, Deep convolutional neural network model based chemical process fault diagnosis, Computers & Chemical Engineering, 115, 185-197

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50225-6

Dynamic Risk Analysis for Process Safety

Yuchen Wang[a,b], Zuzhen Ji[a], Shuanghua Yang[a,b], and Yi Cao[a,b*]

[a] *Department of Chemical Engineering, Zhejiang University, Hangzhou, 310000, China*
[b] *Institute of Zhejiang University-Quzhou, 78 Jiuhua Boulevard North, Quzhou, China 32400*
caoyi2018@zju.edu.cn

Abstract

Successful implementation of chemical production systems requires an effective mechanism to assess dynamic risk quantitatively. Existing works of dynamic risk analysis are based on mechanism examination and probability estimation. Complexity arises due to the difficulty while collecting historical accident data. Also utilizing the indicators obtained from the combination of multivariate variables cannot clearly represent the relationship between physical components and safety. To address the shortcomings of the existing methods, this paper proposed a dynamic risk analysis method for process safety. This is achieved by firstly using canonical variable analysis (CVA) to monitor the system, and then computing the remaining useful life (RUL) from simulation as the indicator representing the dynamic risk. A coal slurry gasification process is applied to illustrate the effectiveness of the proposed method. The method has the potential for the industry to understand how a failure occurred and at what time to prevent.

Keywords: Fault detection; Failure prognosis; Safety prediction; Dynamic risk analysis

1. Introduction

Modern chemical processes are becoming complex, which further arises the challenges to the safety of chemical production. The traditional chemical safety assessment is mostly subjective associated with risk analysis, while many risk methods are based on the knowledge and experience of the experts who conduct the analysis. No doubt, this would cause deviations among different experts. Although the ISO31000 has addressed a standard way to analyse risk by including experts from different perspectives to compromise the deviations, there are still weakness, for example, time-consuming.

Risk is used to represent the safety of a process. Recent studies were focused to develop quantitative or dynamic methods to analyse the risk for the chemical process. An efficient process monitoring scheme is able to capture the complex relationships among the interacting process variables and also to estimate the risk of the process continually (Amin et al.,2020). The current dynamic risk methods are often developed based on historical data or operation data (Zio, 2018). Historical statistics data is also known as Accident Sequence Precursor (ASP) data, which refers to the data recorded via historical accidents. Amin et al. (2020) used a combination of the naïve Bayes classifier, Bayesian network, and event tree analysis to detect and diagnose the fault, and obtain the probability that indicates the dynamic failure prognosis. Alternatively, to ensure a reliable result, a large amount of ASP data is often required. Some other methods are using operational data to examine the safety risks. Zadakbar et al. (2013[a]) used a combination of the multivariate statistic method and probability model to assess the system dynamic risk. Other risk analysis works are using prognostic methods; however, these works are often bespoke.

Some other research has utilized the model-based method to demonstrate the development of the system to get the dynamic risk (Zadakbar et al., 2013[b]). However, it's difficult to attain the precise model for the complicated process, or the modelling cost is not acceptable. In addition, recent process monitoring methods are usually based on the indicators obtained by multivariable combinations, which cannot clearly reveal the relationship between specific physical meaning and safety indicators.

The purpose of the work is to propose a method independent of the mechanism model and probability model that analyses the evolution from fault to failure using dynamic risk analysis. The dynamic risk in this paper is the relative risk which is compared with the risk within the time series of the system itself. Canonical variables analysis (CVA) is utilized to detect the fault and identify the subspace model, then simulate the safety critical variables to get the system remaining useful life (RUL), which represents the time from fault to failure. The RUL is utilized as the metric of system dynamic risk.

The rest of the paper is organized as follows. Section 2 introduces a specific method framework, and then, the proposed approach is demonstrated to be effective through a gasification case study in Section 3. Finally, the paper is concluded in Section 4.

2. Methodology

The overall methodology has been concluded in Figure 1. We need to analyse the safety critical variables (SCVs) and their threshold values in advance. The definition of SCVs can be seen in previous work (Ji et al., 2021). And the SCVs are considered as the directed standard that assesses whether system failure occurs. Once any SCVs is out of the threshold, the system will be regarded as having a failure. Next is to monitor and identify the system using CVA method. If a fault is detected at a point in time, the identified model is used to simulate the SCVs at each point of time until any SCVs exceeds threshold. We can obtain the RUL for each point of time, which is regarded as an indicator that reveal the system relatively dynamic risk, and can provide in-time information with reference value to the operators that shows current dynamic risk and the degree of urgency to solve the fault.

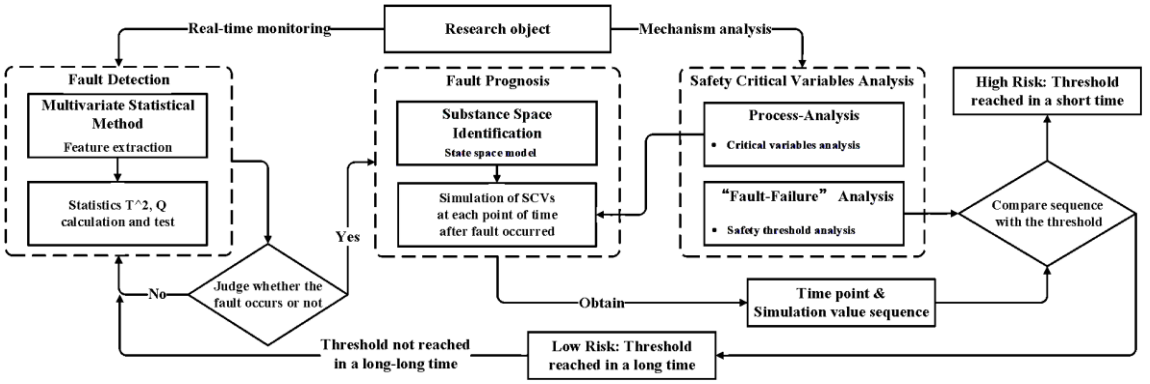


Figure 1 Research Approach

2.1. Safety critical variable analysis

There are adequate traditional methods to analyse system safety, such as event tree analysis, fault tree analysis, bow-tie analysis, and HAZOP. The above methods can be used for SCV selection.

2.2. System monitoring

The proposed process monitoring method is based on the CVA (Ruiz-Cárcel et al, 2016). It can be divided into two parts: Off-line modelling and On-line monitoring.

1) Off-line modelling

Assume y and u are the outputs and the inputs of the system after pre-processing of normal operation dataset. The past and future vectors p_t and f_t of current time point can be defined as Eq. (1), where p and f are the hysteresis.

$$p_t = [y_{t-1}^T, \dots, y_{t-p}^T, u_{t-1}^T, \dots, u_{t-p}^T]^T \qquad f_t = [y_t^T, y_{t+1}^T, \dots, y_{t+f}^T]^T \tag{1}$$

Then we get the past and future Hankel matrix P_t and F_t. And the number of columns of Hankel matrix can be calculated as $\mathrm{N} = n - p - f + 1$, where n means the sample number.

$$P_t = [p_t, p_{t+1}, \dots, p_{t+N-1}] \qquad F_t = [f_t, f_{t+1}, \dots, f_{t+N-1}] \tag{2}$$

The solution for the optimization problem to find the linear combination that best correlates both data sets can be obtained by using the Singular Value Decomposition (SVD) of the Hankel matrix H, where $\sum_{A,B}$ represents the sample covariance matrix of two matrices A and B.

$$H = \sum_{F_t,F_t}^{-1/2} \sum_{F_t,P_t} \sum_{P_t,P_t}^{-1/2} = UDV^T \tag{3}$$

The reduced matrix V_r can be taken by selecting the columns of V related with the r highest eigenvalues from D. And the rest q columns of matrix V is recorded as V_q. Calculate state vector x_t, change vector e_t outside state space, residual vector r_t based on mapping matrix J, L, R according to Eq. (4). And $x_t = Jp_t$, $e_t = Lp_t$, $r_t = Rp_t$. Then the Hotelling and SPE statistics can be calculated as $T_s^2 = x_t^T x_t$, $T_r^2 = e_t^T e_t$, $Q = r_t^T r_t$.

$$J = V_r^T \sum_{P_t,P_t}^{-1/2} \qquad L = V_q^T \sum_{P_t,P_t}^{-1/2} \qquad R = I - J^T J \tag{4}$$

And for identification of the state space model, given a set of inputs u and outputs y, the model that represents the linear state-space can be described as Eq. (5). The system outputs y_k and inputs u_k are known and x_k can be obtained from CVA analysis, the only unknowns of the system are the matrices A, B, C, D. The multivariate regression can be used for the calculation of these matrices.

$$\begin{aligned} &x_{k+1} = Ax_k + Bu_k + w \quad y_k = Cx_k + Du_k + v \\ &\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \sum \left[\binom{x_{k+1}}{y_k}, \binom{x_k}{u_k} \right] \cdot \sum \left[\binom{x_k}{y_k}, \binom{x_k}{u_k} \right]^{-1} \end{aligned} \tag{5}$$

2) On-line monitoring

The new process data can also be organized as the form of past vector $p_{t,fault}$, and according to Eq. (4), we can obtain the state vector $x_{t,fault}$. Then we can get the SPE and Hotelling statistics of the new process data. Then the process can be monitored by comparing the statistics with the corresponding threshold. If the threshold is continuously exceeded, the system is considered being in faulty.

2.3. Dynamic risk analysis

For the new process, if a fault is detected, the state space model is updated in real time after 30 sampling intervals. In order to improve the accuracy of multi-step simulation, the model coefficient is retrained. The specific method is shown as Eq. (6-7). According to the identified model, SCV is simulated over time until one of them exceeds the threshold, and the number of steps is recorded as RUL. Take the RUL as the indicator for the relative dynamic risk and its evolution trend.

$$\begin{aligned} &\hat{y}_k = Cx_k + Du_k + v \\ &\hat{x}_{k+1} = Ax_k + Bu_k + w \qquad \hat{y}_{k+1} = C\hat{x}_{k+1} + Du_{k+1} + v \quad \ldots \\ &\hat{x}_{k+n} = A\hat{x}_{k+n-1} + Bu_{k+n-1} + w \qquad \hat{y}_{k+n} = C\hat{x}_{k+n} + Du_{k+n} + v \end{aligned} \tag{6}$$

$$L1(A,B,C,D) = argmin(\| [\hat{y}_k \quad \hat{y}_{k+1} \quad \cdots \quad \hat{y}_{k+n}] - [y_k \quad y_{k+1} \quad \cdots \quad y_{k+n}] \|_2^2) \tag{7}$$

3. Application of the proposed methodology

3.1. Introduction of the coal slurry gasification

Coal water slurry gasification is an entrained flow gasification process for generating syngas under high temperature and high pressure with coal water slurry as raw material and oxygen or air as gasification agent. The main component of syngas is carbon monoxide, carbon dioxide and hydrogen. The gasification process applied as a case study is based on a dynamic simulation model which contains coal slurry storage tank, gasifier combustion chamber and quench chamber. Figure 2 illustrated the gasification process.

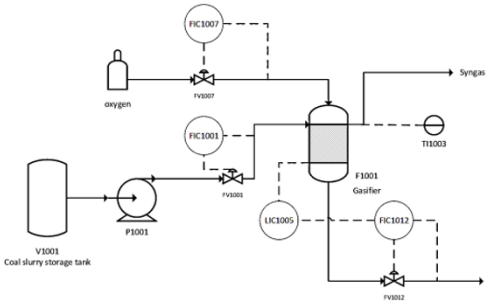


Figure 2 Coal slurry gasification process

3.2. Determine safety critical variables

There are 14 measured variables and 10 manipulated variables in the gasification process. And the SCVs are selected as gasifier combustion chamber temperature R1.T and R2.T using the method proposed previously (Ji et al., 2021). The threshold of the upper and lower temperature of combustion chamber are $[920, 1019]$ ℃ and $[1405, 1505]$ ℃.

3.3. Fault detection

Considering the large fluctuation of coal composition in the actual process, the component of coal is selected as the variable of fault. The fault is introduced after 3.5 hours of simulation, and the total simulation time is 13.5 hours. The response curves of oxygen flow, gasifier temperature, CO and CO_2 molar fractions are illustrated in Figure 3.

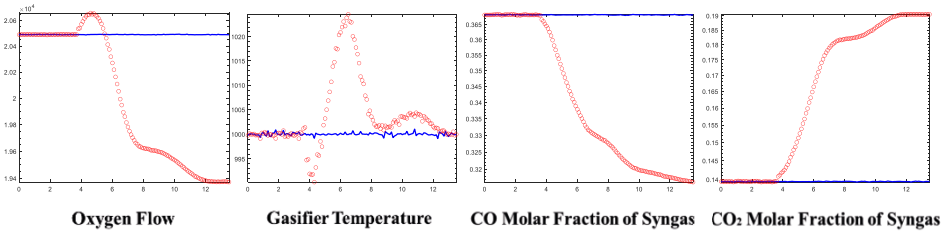


Figure 3 Response curve (Circle line: fault condition; Solid line: normal condition)

Then the CVA algorithm is used to realize the safety monitoring of gasification process. The results are shown as Figure 4. The dotted line refers to the threshold of statistics under the condition of confidence of 0.999.

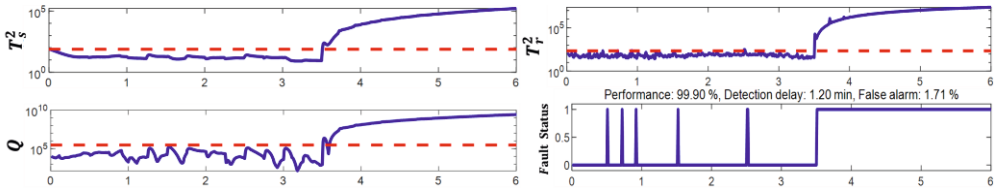


Figure 4 Fault monitoring of gasifier

From Figure 4, it can be seen that the process fault detection rate based on three statistics is 99.90%, the false alarm rate is 1.71%, and the fault detection delay time is 1.20min. According to the sampling interval of 0.01h, it is the delay of two sampling intervals.

3.4. Dynamic risk analysis for the gasification

Based on Sections 2.2 and 2.3, we obtained the identified model of the normal system, and then verify the accuracy of the model on the testing dataset sampled in normal operation. The validation results are illustrated in Figure 5, where the circle line represents simulation value, and solid line represents real value. And based on the coefficient of determination R^2, it can be seen that the accuracy is acceptable and satisfactory.

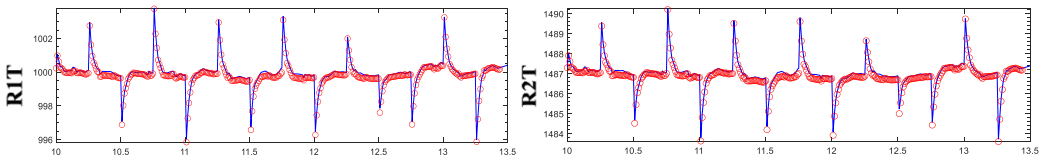


Figure 5 Verification of the model ($R_{R1T}^2 = 0.9886, R_{R2T}^2 = 0.9864$)

When the fault is detected, the identified model trained from the normal mode is not applicable. According to the section 2.3, the state space model is retrained using fault data and the results will be used as the initial value of multi-step regression training. Finally, the simulation of SCVs is executed for each time point. Figure 6 is an example of the 400th point of time, where the circle line represents simulation value, solid line represents real value, and black line means the threshold. Their accuracy is acceptable.

For the dynamic risk analysis, the simulation of each point has been executed until the values of SCVs are out of the threshold, and the needed time length is recoded. The variation of RUL with sampling points is shown in Figure 7. The asterisk represents RUL obtained from simulation, while circle represents real value from the dynamic model. The RUL obtained from simulation is compared with the real value of remining useful life, and their correlation coefficient is 0.999, which means that the simulation traces the real

circumstance well. From Figure 7, RUL decreases continuously, hence the risk is increasing correspondingly. And with the continuous development of the fault, the RUL has been reduced to an unacceptable level, i.e., corresponding measures must be applied to the system immediately.

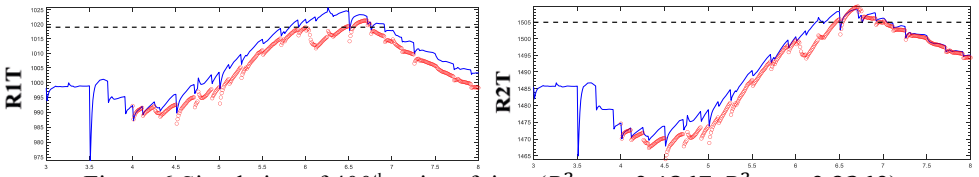


Figure 6 Simulation of 400th point of time ($R^2_{R1T} = 0.4867, R^2_{R2T} = 0.9360$)

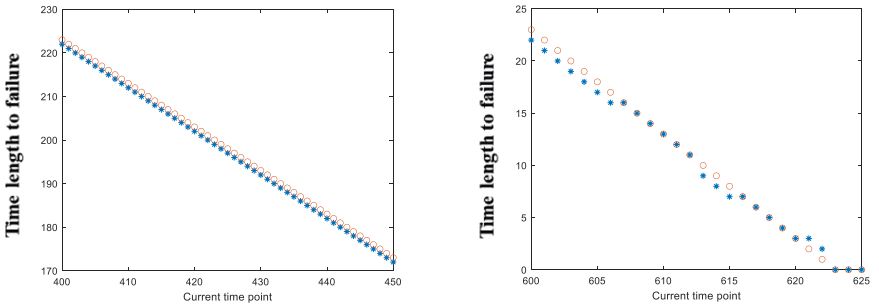


Figure 7 Development of RUL over time

4. Conclusions

This work proposed a novel dynamic risk analysis approach that could better demonstrate the dynamicity of risk based on RUL. This would be useful for managing process risk. In this approach, the process is monitored by the CVA, and once the fault is detected, then CVA would identify a model for the simulation of SCVs analyzed from the system mechanism, and then the simulation would obtain the time RUL which indicated the dynamic risk of the system. The proposed approach was applied to a coal slurry gasification for its application and validation. The result proved the method to be effective to reveal the system dynamic risk and its evolution trend.

References

T. Amin, F. Khan, S. Ahmed, S. Imtiaz, 2020. A novel data-driven methodology for fault detection and dynamic risk assessment. The Canadian Journal of Chemical Engineering 98, 2397 – 2416.

E. Zio, 2018. The future of risk assessment. Reliability Engineering & System Safety 177, 176-190.

O. Zadakbar, S. Imtiaz, F. Khan , 2013. Dynamic risk assessment and fault detection using principal component analysis. Industrial & Engineering Chemistry Research, 52(2), 809 – 816.

O. Zadakbar, S. Imtiaz, F. Khan, 2013. Dynamic risk assessment and fault detection using a multivariate technique. Process Safety Progress 32, 365 – 375.

C. Ruiz-Cárcel, L. Lao, Y. Cao, D. Mba, 2016. Canonical variate analysis for performance degradation under faulty conditions. Control Engineering Practice, 54, 70 – 80.

Z. Ji, S.Yang, Y. Cao, 2021. Harmonizing safety and security risk analysis and prevention in cyber-physical systems. Process Safety and Environmental Protection, 148(5)

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50226-8

Monitoring of smart chemical processes: A Sixth Sense approach

Bogdan Dorneanu[a], Mohamed Heshmat[b], Abdelrahim Mohamed[c], Harvey Arellano-Garcia[a,*]

[a]*LS Prozess- und Anlagentechnik, Brandenburgische Technische Universität Cottbus-Senftenberg, Cottbus, D-03046, Germany*
[b]*Faculty of Environment and Technology, University of the West of England, Bristol, BS16 1QY, United Kingdom*
[c]*Institute for Communication Systems, Home of 5GIC, University of Surrey, Guildford, GU2 7XH, United Kingdom*
arellano@b-tu.de

Abstract

This paper introduces the development of an intelligent monitoring and control framework for chemical processes, integrating the advantages of technologies such as Industry 4.0, cooperative control or fault detection via wireless sensor networks. The system described is able to detect faults using information on the process' structure and behaviour, information on the equipment and expert knowledge. Its integration with the monitoring system facilitates the detection and optimisation of controller actions. The results indicate that the proposed approach achieves high fault detection accuracy based on plant measurements, while the cooperative controller improves the operation of the process.

Keywords: fault detection, wireless sensor networks, cooperative control, Industry 4.0.

1. Introduction

Industry 4.0 is transforming chemical processes into complex, smart cyber-physical systems, by the addition of elements such as smart sensors, Internet of Things (IoT), big data analytics, or cloud computing. Modern engineering systems and manufacturing processes are operating in highly dynamic environments, and exhibiting scale, structure and behavior complexity. Under these conditions, plant operators find it extremely difficult to manage all the information available, infer the desired conditions of the plant and take timely decisions to handle abnormal operation (Natarajan & Srinivasan, 2014). Human beings acquire information from the surroundings through sensory receptors for vision, sound, smell, touch, and taste, the Five Senses. The sensory stimulus is converted to electrical signals as nerve impulse data is communicated with the brain. When one or more senses fail, the humans are able to re-establish communication and improve the other senses to protect from incoming danger. Furthermore, the mechanism of reasoning has been developed during evolution, which enables analysis of present data and generation of a vision of the future, which might be called the Sixth Sense.

As industrial processes are already equipped with the five senses: hearing from acoustic sensors, smelling from gas and liquid sensors, seeing from cameras, touching from vibration sensors and tasting from compositions monitors, the Sixth Sense could be achieved by forming a sensing network which is self-adaptive and self-repairing, carrying

out deep-thinking analysis with even limited data, and predicting the sequence of events via integrated system modelling.

In this paper, an intelligent monitoring and control framework for chemical processes is proposed, which takes advantage of recent technological developments such as wireless sensor networks (WSNs), 5G communication, or cooperative control to ensure stable process operation.

2. System architecture

The framework consists of five main components, designed in a modular manner (Fig.1). The first one is a wireless sensor network (WSN), transmitting over a 5G communication network, that facilitates data management for improved fault detection. The second component is an efficient fault detection algorithm that can analyse the data and classify it in faulty or normal. The third component is a knowledge-based and model-based fault detection monitoring system. For the fault-detection, a two-stage method based on a hybrid learning approach is applied, which utilizes supervised and unsupervised learning. The fourth component is a cooperative model predictive control (MPC) system that takes the required measures to ensure stable process operation. Finally, the fifth component is an intelligent adaptive decision-making framework.

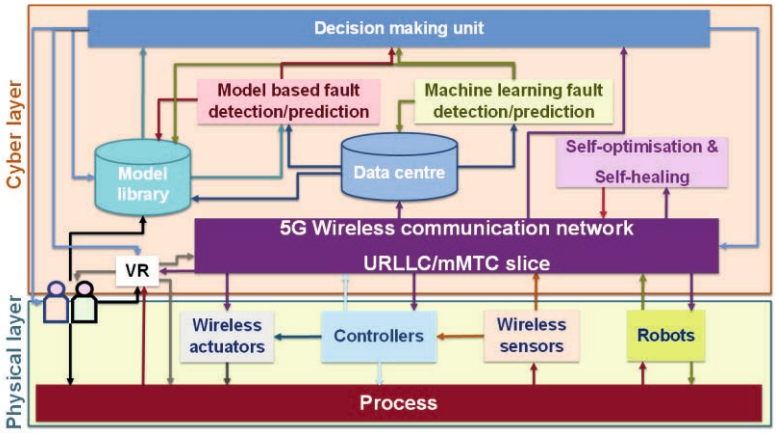


Figure 1: System structure - High level overview

The proposed architecture divides the system into a physical layer, which includes the industrial process, the wireless sensors and actuators, the controllers, and inspector robots, and a cyber layer, formed of the wireless communication network, the fault detection algorithms and the decision-making framework.

2.1. Industrial process

The application of the proposed architecture is investigated on a mini plant (Fig.2) available at the Department of Chemical and Process Engineering of the University of Surrey (Guildford, United Kingdom). The plant produces sodium ion solution for sale to fine chemical, pharmaceutical and food industry. The raw material (sodium chloride contaminated with calcium chloride) is pre-mixed with sodium bicarbonate. This feedstock is then fed into a reactor vessel charged with pre-heated water, and reacted at 65 ^{0}C. The suspension resulted from the reaction is then pumped by a positive displacement pump through a plate filter, which removes a high proportion of the calcium

carbonate. The output stream splits into a product and a recycle line. The recycle line is fed back into the reactor, while the product is passed through a polishing filter, to remove any remaining solids, and then to the lot tank. The carbon dioxide resulted in the reaction is absorbed in an alkaline solution. The plant is provided with wireless sensors connected to the cellular network, that monitor process variables (e.g., temperature, pressure, processing unit level, etc.) and a distributed control system. The sensors are transmitting data every second.

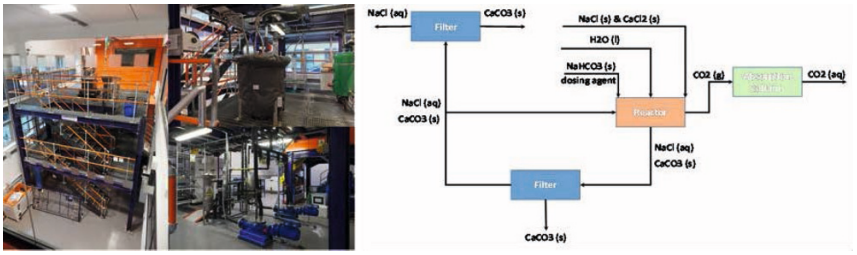


Figure 2: Mini plant and simplified process diagram

Robots are also deployed in the plant for periodic maintenance and surveillance. In the proposed framework, the mobile robots will autonomously patrol around the plant. They are equipped with sensors for recording the necessary measurements, performing the periodic inspection and surveillance, as well as acting as relays in deep fading areas. For the mobile robots exploited in this system, the navigation is achieved using standard robot operating system navigation stack.

2.2. Communication network

Legacy cellular systems have been designed primarily for human initiated mobile broadband communications, making them highly suboptimal for narrow band, short-bust, sporadic traffic (e.g., sensor measurement data) generated by sensors in chemical plants. It is envisioned that a new design paradigm is needed to support large numbers of heterogeneous sensing devices with diverse requirements and unique traffic characteristics. Comparing to sensors in traditional IoT network, those deployed in extreme environments need to operate in harsh, sometimes hazardous conditions, and are, thus, prone to wear and tear, and cannot be easily replaced, posing major challenges in designing resilient networks for reliable communication.

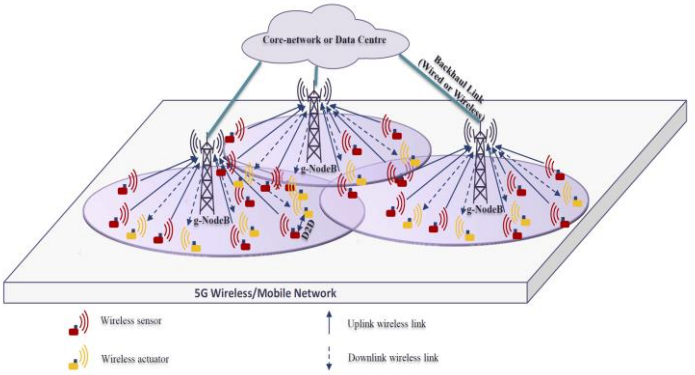


Figure 3: Communication network

For the system presented in this paper, a centralized control mechanism of the communication network is considered, where the sensors are connected to a fusion node via wireless links, as shown in Fig.3. The wireless links can also be used to send commands to the actuators within the mini plant. The resulting network consists of a heterogeneous set of periodic and event triggered sensors with mixed requirements, characteristics and traffic models.

Considering the plant and associated sensors' heterogeneity, a statistical model rather than a deterministic model is chosen for the sensor transmission events. The number of incoming packets (or events when each event generates a single packet) per unit time follows the Poisson distribution, while the packet interval is modelled as an exponential distribution. This results in probability-based transmissions that can be controlled by the arrival rate and the inter-arrival time.

2.3. Machine learning

In the machine learning model, both fault detection and prediction applications are embedded for sophisticated fault handling, as illustrated in Fig.4, and takes in online data streams and inputs to feed them to both (or either one of) the fault detection and fault prediction sub-models.

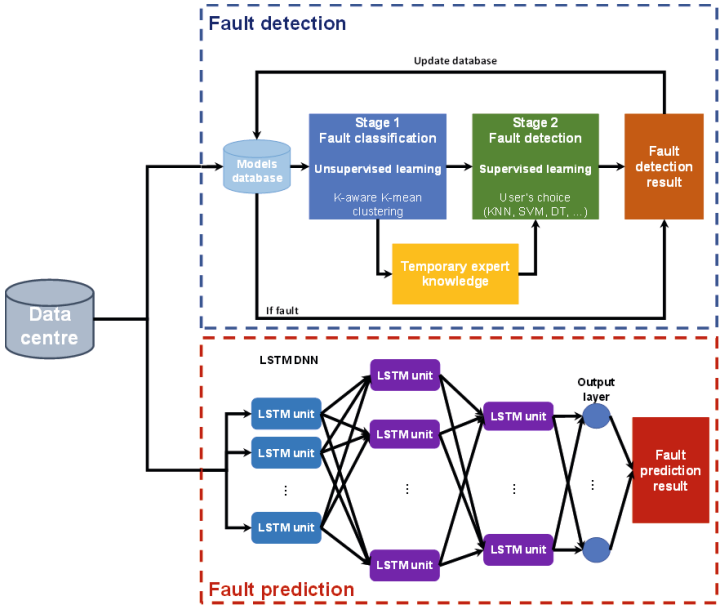


Figure 4: Fault detection and fault prediction model

In the fault detection, a batch of online data streams are checked for expert knowledge recognized types, recorded in the databases. If the streams pass this initial check, a combination of unsupervised and supervised learning algorithms are applied on the data for fault classification. The unsupervised learning algorithms, namely K-aware K-mean, are an extension of the conventional K-mean clustering algorithms with additional capacity to self-optimize the K-value ($K \leq 1$). This phase aims to acquire a temporary expert knowledge on what the minority of the current data (the smallest cluster) is like and labels them as outliers. Subsequently, a choice of the classical supervised learning

algorithms is conducted based on the temporary expert knowledge to classify the new types of faults and comparing them with the expert knowledge database. Once a new type of fault is validated by the system, the database is updated.
In the fault detection sub-model, a long short-term memory (LSTM) deep neural network (DNN) is proposed to perform online fault prediction. The relationship between the elements of a single LSTM unit is described by the following relationships.

$$i_t = \sigma_g \cdot (W_i \cdot x_t + U_i \cdot h_t + b_i) \quad (1)$$
$$o_t = \sigma_g \cdot (W_o \cdot x_t + U_o \cdot h_t + b_o) \quad (2)$$
$$f_t = \sigma_g \cdot \left(W_f \cdot x_t + U_i \cdot h_f + b_f\right) \quad (3)$$
$$c_t = f_t \circ c_{t-1} + i \circ \sigma_c \cdot W_c \cdot x_t + U_c \cdot h_t + b_c \quad (4)$$
$$h_t = o_t \circ \sigma_h(c_t) \quad (5)$$

With $x \in \mathcal{R}^n$ and $h \in \mathcal{R}^h$ the input and the first hidden layer output of the LSTM-DNN, respectively; t the time step; $i, o,$ and f the input gate, the output gate and the forget gate vectors, respectively; $c \in \mathcal{R}^h$ the cell state vector; $W_i,\ U_i,\ W_o,\ U_o, W_f, U_f, W_c$ and U_c the weight matrices; b_i, b_o, b_f and b_c the biases; σ_g, σ_h and σ_c the activation functions. In the equations above ° denotes the Hadamard product.

2.4. Cooperative control

A cooperative distributed MPC approach is considered to achieve highly flexible dynamic optimal control. A cooperative protocol is defined using a simple algorithm to reach an agreement regarding the state of a number of N process units (agents). The MPC controller receives information from the process unit in the form of the output process variable, y, as well as from the neighbouring units in the form of a cooperation variable, v. Based on this exchange of information, the controller will correct the input variables to the system, u).

2.5. Decision-making system

To add intelligent and adaptive decision-making capabilities, a multi-agent system (MAS) is defined to be integrated with the cooperative MPC. The MAS has two main tasks: to decide the optimal connectivity between the distributed MPCs for safer and better operation, and to monitor the system and detect any deviation in the behaviour, which is then transmitted to the controllers. The proposed MAS consists of follower agents, which keep track of equipment behaviour and its relationship with other control units, a coordinator agent, which decides the optimal connectivity between controllers, and a monitor agent, which analyses the information gathered from the coordinator agent and reports the expected failures and recommendations to the operator.

3. Results

The framework presented in the previous sections has been implemented for fault detection and prediction applications in the mini plant. A dataset was obtained from the plant during the year 2017-2018 and consists of over 10 million samples, each containing measurements for 43 variables, collected at a frequency of 1 Hz. Data types include both floating point and Boolean value data. The batch size for both fault detection, and prediction algorithms is set to 10,000. All simulations are conducted in Python and Keras application programming interface (API) with Tensorflow backend on an Intel i7-8700 CPU (6 Core/12-Thread, 12 MB cache, up to 4.6 GHz with Intel Turbo Boost

Technology) and dual Nvidia GeForce GTX 1080 Ti GPUs (with 11GB GDDR5X each). For all supervised learning algorithms, the train-test ratio is chosen at 0.75:0.25. During fault detection, the result of classifying the original high dimensional data is projected into a lower dimension (2D plane) using principal component analysis (PCA) as shown in Figure 5. Table 1 compares different supervised learning algorithms regarding fault detection accuracy.

Table 1: Complexity comparison

Supervised learning classifier	Detection accuracy, %
Logistic regression	97.5
K-nearest neighbors	100.0
Decision tree	99.4
Linear support vector machines	99.5

For fault prediction, a three hidden layer LSTM-DNN is configured to perform one time step ahead prediction for 40 dynamic variables, with 30, 50, and 30 units used, respectively. The mini-batch method is used to accelerate the stochastic gradient optimiser. Each LSTM hidden layer uses a linear activation function and the same applies for the output layer. The proposed LSTM-DNN is able to achieve a prediction accuracy of 99.7%. The train-test learning curve comparison is illustrated in Figure 6.

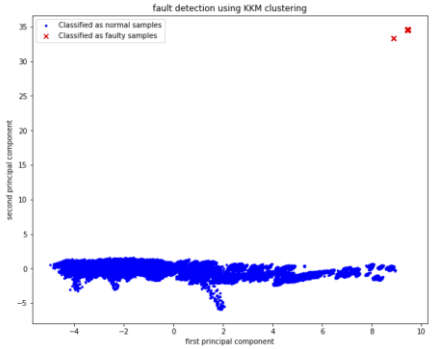


Figure 5. Fault detection

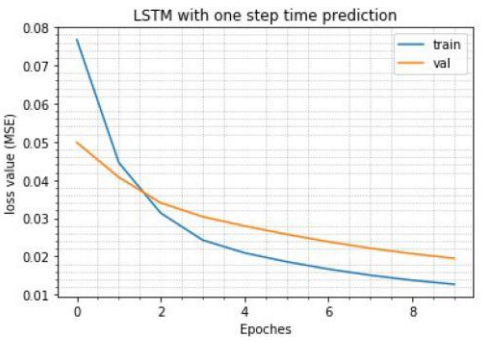


Figure 6. Train-test learning curve comparison

4. Conclusions

In this paper, a high-level interdisciplinary framework is developed for leveraging capabilities of 5G WSNs, machine learning algorithms and cooperative control to step towards the industrial Sixth Sense. Since different applications have different requirements, the proposed system is built in a modular manner. A higher decision layer is proposed for fault detection and prediction to optimise the plant operation. Focusing on a chemical process as a case study, the proposed framework has been used to control the plant. The results indicate that the proposed approach achieves high fault detection/prediction accuracies based on real plant measurements. Future work will consider the impact of mobile wireless sensors deployed in robots, as well as integrating the machine learning fault detection/prediction schemes network-specific models to achieve both network and process fault prediction.

References

Natarajan, S., & Srinivasan, R., 2014, Computers and Chemical Engineering 60, pp. 182-196

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50227-X

Fault detection in a benchmark simulation model for wastewater treatment plants

Pedram Ramin[a]*, Xavier Flores-Alsina[a], Sebastian Olivier Nymann Topalian[a], Ulf Jeppsson[b], Krist Gernaey[a]

[a] *Process and Systems Engineering Center (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Building 228A, DK-2800, Kgs. Lyngby, Denmark*
[b] *Division of Industrial Electrical Engineering and Automation, Department of Biomedical Engineering, Lund University, Box 118, SE-221 00, Lund, Sweden*
*Corresponding Author's E-mail: pear@kt.dtu.dk

Abstract

The International Water Association (IWA) Benchmark Simulation Models (BSM1 and BSM2) have been successfully used in both industry and academia to test and verify control strategies in wastewater treatment plants (WWTPs). In this study, a new (plant-wide) benchmark simulation model, the BSM2-LT, is developed to evaluate monitoring algorithms. This platform provides opportunities to generate various sensor/actuator and process faults. To make this realistically, different Markov-chain models are used to re-create the alternation of sensor/actuator states based on predefined occurrence probability. The same principle is used to describe the occurrence of toxic/inhibitory compounds. Using this platform, one can test the performance of a monitoring algorithm such as a fault detection method. To demonstrate this in an example, a multivariate method based on adaptive dynamic principal component analysis (dPCA) was used to detect faulty events. The performance of the monitoring algorithm is evaluated with a penalization index, scoring from 0 to 100. While the tested method had a good false alarm score, it resulted in a low false acceptance. While the results could be certainly improved, the main focus of this study is the benchmark simulation model and not presenting a well optimized monitoring algorithm. The software which will be produced and freely distributed in the near future, will allow an objective evaluation of monitoring algorithms for WWTPs for any user.

Keywords: Benchmark simulation, Fault detection, Markov chains, Monitoring algorithms, Wastewater treatment

1. Introduction

Instrumentation, control and automation for industrial chemical and biochemical processes to attain cost-effective and safe process operation are highly dependent on reliability of the real-time measurements. Despite considerable development of online sensors during the past decades, their dependability is still impaired due to various fouling and failing issues. Occasionally unsatisfactory measurement performance can prevent full instrumentation of plant-wide control systems. This is especially important for wastewater treatment plants (WWTPs) where often fault-tolerant control systems needs to be implemented. Small WWTPs generate up to 500 signals (including on-line and off-line signals), whereas larger ones typically register over 30,000 (Olsson 2012). Despite a

large number of available signals, data reconciliation and validation for online instrumentation has remained a largely unexplored field with a lack of standardized approaches. Most data are stored unstructured, with lots of gaps, repetition, ambiguity and uncertainty. This has led to "data-rich, information-poor" situations in which data sets are often too large and complex for processing and analysis to be used for decision-making. To turn raw data into useful and actionable information, data need to be validated. This can be achieved through a fault detection procedure. While several advanced data validation tools have already been developed for various chemical industries, their application to wastewater treatment facilities is not straight forward. Compared to chemical industrial processes, municipal WWTPs have unique features which complicates their monitoring and control strategies: (i) Wastewater influent is characterized by short- and long-term high temporal variability in both quantity and quality. WWTPs are impacted by growing urban areas, and variable rain events; (ii) Input materials (wastewater, storm water) cannot be stored in large quantities if the supply exceeds the process capacity; (iii) Low effluent water quality cannot be discarded or ignored; (iv) Wastewater treatment processes exhibit nonlinear, nonstationary, autocorrelated, cross-correlated behavior and are characterized by changes at many different time-scales; (v) Challenging environmental conditions impose sensor impairment due to solids deposition, precipitate formation and biofilm growth.

1.1. Benchmark simulation models

Benchmark simulation models are great tools to test and evaluate different monitoring methods under various fault conditions (sensors or process faults). The Benchmark simulation for Tennessee Eastman process (TEP) (Downs and Vogel 1993) is perhaps the best known platform in chemical engineering, resembling a real-world example, to test different fault detection strategies in an industrial reaction process (Gravanis et al. 2022). Another example is the IWA Benchmark Simulation Model No. 1 (BSM1) and No. 2 (BSM2) (Gernaey et al. 2005; Jeppsson et al. 2006) describing typical biological and physico-chemical processes in WWTPs (Gernaey et al. 2014). BSM1 includes a five-reactor activated sludge plant configuration with a (non-reactive) secondary clarifier, whereas BSM2 is a plant-wide description of a typical WWTP, extending BSM1 with pretreatment and sludge treatment processes. These benchmark platforms allow simulation-based performance evaluation of process control and monitoring strategies (Gernaey et al. 2014). An extension to BSM1, named BSM1-LT, was also developed (Rosen et al. 2004) to enable long term evaluation of control strategies (e.g. sludge retention time control based on wastewater flowrate) and process monitoring performance (e.g. detecting equipment failure such as nitrate and ammonium sensors). The BSM1-LT has also been used to test different univariate fault detection methods (Corominas et al. 2011). A preliminary attempt was made to simulate different faults in BSM2 followed by detection and diagnosis using incremental principal component analysis (IPCA) (Kazemi et al. 2020).

The aim of this study is to extend BSM2 by including "realistically" different sensor/actuator and process faults which are compatible and unified with the previous developments (influent generator, process models, sensor and actuator models, simulation procedure, evaluation criteria). The final output, the BSM2-LT, will include different scenarios which will be suitable to test univariate/multivariate statistical monitoring methods as well as fault-tolerant control strategies. The final software will be freely available for any user.

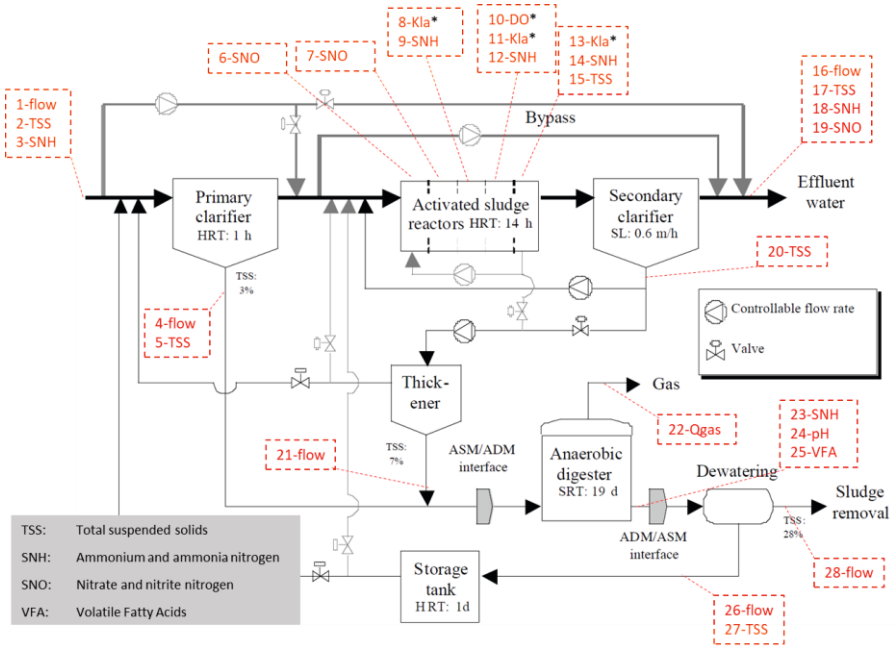


Figure 1. Layout of BSM2/ BSM2-LT. Process data for objective evaluations in this study are indicated with numbers (definition are given in the box). Controlled variable (DO) and manipulated variable (K_La), indicated with asterisks, are subject to faults in this study.

2. Methods

Two different types of faults, sensor/actuator fault and process fault were investigated.

2.1. Sensor/actuator fault generation

Faults were created only for the oxygen (DO) sensor in reactor 4 (used to control oxygen in aerated reactors 3 to 5) and for the actuator, K_La (oxygen mass transfer coefficient, equivalent to airflow in aerated reactors). Faults were modelled using a Markov-chain approach, where 7 different states were considered with a predefined occurrence probability, reported here as % for DO and K_La, respectively: (i) fully functional (93.8% and 88.3%); (ii) excessive drift (0.5% and 9.2%); (ii) shift (0.9% and 0.1%); (iv) fixed value, (1.6% and 1.6%); (v) complete failure (0.2% and 0.7%); (vi) wrong gain (2.7% and 0%); (vii) calibration (0.4% and 0%). The model was regenerated from earlier studies (Rosen et al. 2008; Corominas et al. 2011) following the same recommendations and parameters.

2.2. Process fault generation

Process disturbances were created by introducing two inhibitory or toxic substances, represented by new state variables S_{tox} and X_{tox}, in soluble and particulate form (absorbed to other particulate matters), respectively. The inhibition here is defined as the reduction in growth rate and toxicity is defined as both reduced growth rate and increased decay rate. The biological impact is defined by different inhibition/toxicity parameters using combined concentration of S_{tox} and X_{tox}. These parameters impact growth and decay of heterotrophs and autotrophs in activated sludge reactors (Gernaey et al. 2014). To

generate S_{tox} and X_{tox} at the influent of BSM2-LT, two separate Markov-chain models were used to create a basic 'seed' file. Three states were used, no discharge, discharge with low amount (inhibitory discharge), discharge with high amount (toxic discharge). For S_{tox}, toxicity is assumed to occur every six months and inhibition every two weeks, with a duration of 3 hours for both. For X_{tox}, toxicity is assumed to occur every six months and inhibition once every month, with a duration of 1 hour for both. The 'seed' file was then used in the sewer model of the influent generator in a similar approach as described elsewhere (Gernaey et al. 2014).

2.3. Fault detection

An adaptive-dynamic fault detection was tested (Odom et al. 2018) using dynamic principal component analysis (dPCA) with a moving window (8 days). To take into account the correlation between sequential observations, all the data were lagged two-time steps (two samples). The first 60 days were used for training and were assumed to be a 'normal' period (all faults and toxicity were removed in this period). Two multivariate monitoring statistics, i.e. Hotelling's T^2 and the square prediction error (SPE) were used with a threshold of 0.5%. A flag was issued if this threshold was violated. Five consecutive flags would then trigger an alarm, indicating fault in the data. In total 28 variables were selected for the fault detection as presented in Fig. 1.

2.4. Performance evaluation

To evaluate the performance of the previously described fault detection method, an evaluation index was used as a measure of reliability (Corominas et al. 2011). This index is ranging from 0 (not reliable) to 100 (reliable). It is calculated by penalization points when the fault detection does not succeed. Penalizations are importantly based on: (i) late detection in which penalty points exponentially increases and reach a maximum level, (ii) extra penalizing intermittent detection for unfavorable switching from correct detection to non-detection during a fault event. Indices are defined for false alarm (J_{FAL}), and false acceptance (J_{FAC}) as well as a total index (J). Moreover, precision and accuracy were also calculated according to a calculated confusion matrix (error matrix).

3. Results

The results of fault generation, model simulation and fault detection are presented in Fig. 2. The results are only shown for reactor 4 for the tested scenarios, i.e. sensor/actuator fault (S1A to S1G) and process fault (S2A to S2G). Table 1 summarises the performance of the fault detection using different evaluation indices. This assessment is done considering T^2 alarm, SPE alarm, and the combination of both.

3.1 Sensor/actuator fault

The DO sensor in reactor 4, is connected to a feedback PI controller with a set point tracking objective. The DO signal oscillated around the set point of 2 g/m^3 (Fig. 2 S1D). A combination of drift in the K_La signal and a wrong gain for the DO sensor, caused the oxygen in reactor 4 to ramp down to almost zero. This fault was captured by T^2 and SPE with delays (Fig 2. S1C). In three occasions of fixed value faults for the DO sensor, the signal from this sensor is reduced to nearly zero, causing the oxygen in reactor 4, momentarily to reach 6 g/m^3. This was due to an overreaction of the control system, compensating for an apparent lack of oxygen. Fixed fault in the K_La signal, significantly brought down the K_La value to around 30, causing the oxygen to drop close to zero. Continuation of this scenario for 10 days, reduced the population of autotrophs by 75%

(Fig. 2 S1F). Consequently, it took nearly 100 days until the situation was brought back to normal in the absence of any other major failure. In this period, ammonium concentrations reached above 40 g/m^3 (S1G). This fault period was well detected by both T^2 and SPE. Overall, the tested fault detection showed relatively good false alarm performance but low false acceptance score (Table 1).

3.2 Process fault

While S_{tox} (Fig. 2 S2A) follows the transport dynamics of the soluble species, X_{tox} (Fig. 2 S2B) resembles those of particulates. This implies that S_{tox} is impacted by dilution with relatively short retention time whereas the retention time for X_{tox} depends on the sludge retention time. This creates different effects on the plant performance. In BSM2-LT the activated sludge system has a hydraulic retention time (HRT) of 14 hours, and the sludge retention time (SRT) in the anaerobic digester is 19 days. In case of appearance of toxic/inhibitory chemicals in reactor 4, autotrophs were in general more impacted than heterotrophs (Fig. 2, S2D and S2F). This in turn caused ammonium concentrations to rise. However, loss of heterotrophs due to high S_{tox} concentration, triggered readily degradable substrate concentration to rise sharply on three occasions. T^2 statistics could capture these faulty moments. In other periods, due to frequent occurrence of toxicity, especially S_{tox}, both T^2 and SPE showed a high level of alarm. In general, SPE was found to be more sensitive compared to T^2 for the tested faults and generated nearly 60% more flags (Fig. 2 S1C and S2C). Better optimization of the hyper parameter, e.g. threshold for SPE and T^2 statistics and the number of flags considered before initiating an alarm, is expected to improve the results. Variable selection for fault detection is also another challenge. It was for example found that only considering the activated sludge reactors (variables 6 to 15) can improve J from 60 to 72 for the sensor/actuator faults scenario.

4. Conclusion

The BSM2-LT presented in this study provides opportunities to create realistic scenarios for both sensor/actuator and process faults. Using this platform, one can test the performance of a fault detection method. The method should ideally distinguish various faults and isolate highly consequential deviating instruments, and initiate an alarm before the system has undergone a severe disturbance. The tested fault detection method here showed good precision and false alarm scores but resulted in low accuracy and false acceptance. Further studies will focus on expanding this platform to simulate more realistic fault scenarios (e.g. for the anaerobic digester), which can occur simultaneously. A combination of both univariate and multivariate monitoring methods will be also used which is expected to improve the results.

Table 1. Evaluation indices for fault detection method tested in this study. TP: true positive, TN: true negative, FP: false positive, FN: false negative.

	Indicator	**Precision (%)**	**Accuracy (%)**	**J_{FAC} (0-100)**	**J_{FAL} (0-100)**	**J (0-100)**
	Formula	$\frac{TP}{TP+FP}$	$\frac{TP+TN}{TP+FP+FN+TN}$			
Sensor fault scenario	T2 alarm	95.4	78.2	2.3	95.4	61.9
	SPE alarm	68.2	59.4	20.8	68.2	51.0
	T2 + SPE alarm	70.4	65.1	41.8	70.4	60.0
Process fault scenario	T2 alarm	65.6	10.6	2.1	65.6	5.8
	SPE alarm	99.4	45.7	37.3	99.4	41.0
	T2 + SPE alarm	88.8	43.2	36.1	88.8	39.2

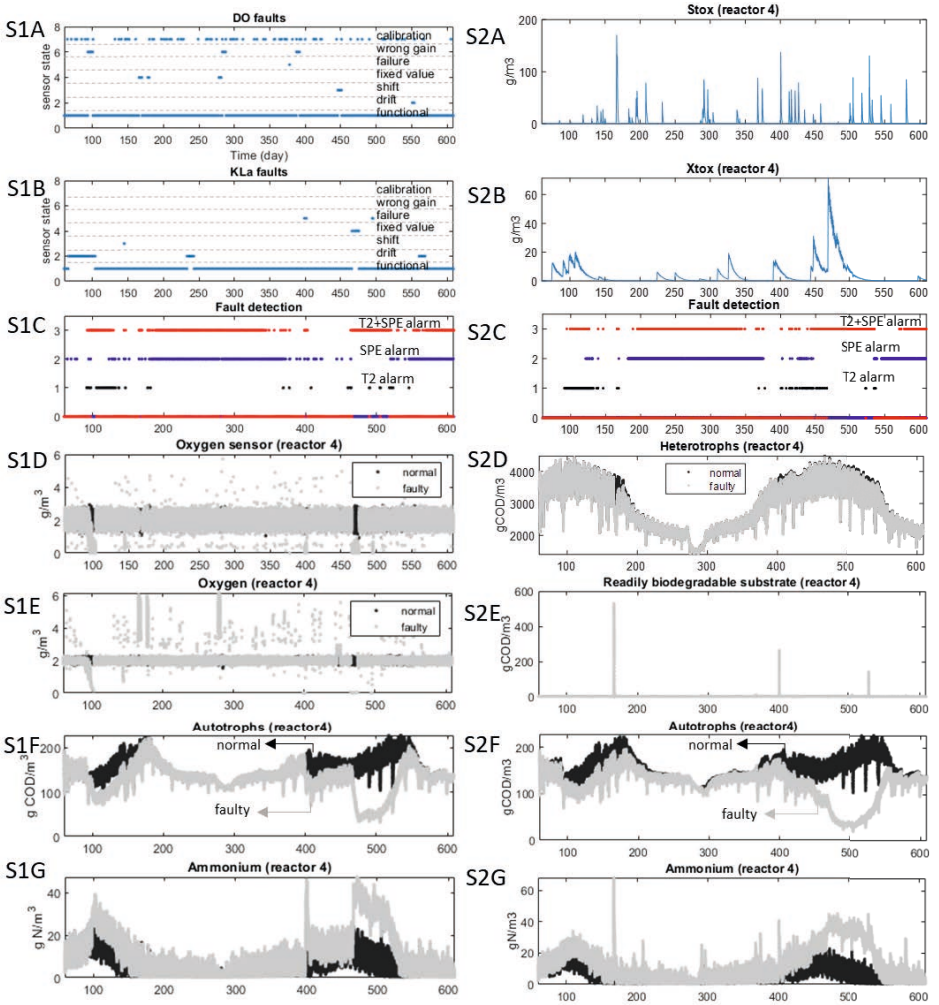


Figure 2. The fault scenarios in this study for sensor faults (S1A to S1G) and process faults (S2A to S2G). The results are shown only for reactor 4.

References

Corominas L, et al., 2011, Biotechnology & Bioengineering 108:333–344.
Downs JJ, Vogel EF, 1993, Computers & Chemical Engineering 17:245–255.
Gernaey KV, et al., 2014, IWA Scientific & Technical Report No. 23. IWA Publishing, London, UK.
Gernaey K V, et al., 2005, Industrial Electrical Engineering and Automation, Lund University, Sweden.
Gravanis G, et al., 2022, Computers & Chemical Engineering 156:107531.
Jeppsson U, et al., 2006, Water Science & Technology 53:287–295.
Kazemi P, et al., 2020, Water Science & Technology 82:2711–2724.
Odom GJ, et al., 2018, Applied Stochastic Models in Business and Industry 34:880–892.
Olsson G, 2012, Water Reseach 46:1585–1624.
Rosen C, et al., 2004, Water Science & Technology 50:41–49.
Rosen C, et al., 2008, Water Science & Technology 57:337–344.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50228-1

Formulation of integrated key performance indicator dashboard for chemical plants

Yasunori Kobayashi[a]*, Yoshiyuki Yamashita[a]

[a]*Department of Chemical Engineering, Tokyo University of Agriculture and Technology, Tokyo 184-8588, JAPAN*
yasunori.5ba8shi@gmail.com

Abstract

A process control system (PCS) exhibits monitoring functions that help board operators achieve safe and stable manufacturing in real time. Because of the rapid changes in the business environment, management objectives, such as carbon neutrality, energy conservation, and profitability, are required in addition to conventional objectives of safety, availability, and productivity. As additional management objectives conflict with conventional objectives, board operators need to monitor the manufacturing process from an optimization standpoint while understanding the new management strategy. New monitoring functions for the next generation of PCS were formulated in this study to help operators achieve energy-saving and profitable manufacturing.

Keywords: KPI, Dashboard, Optimization, Profit, Energy savings

1. Introduction

The monitoring functions in a general process control system (PCS) consist of process flow graphs, process trend graphs, and process alarms. Critical process variables were selected as subjects for monitoring based on a hazard and operability (HAZOP) study (from a safety standpoint) and control narratives (from a stability standpoint). Monitoring functions display these variables as key performance indicators (KPIs) for operators (operational KPIs). Process flow graphs display real-time process variables on a simplified piping and instrument diagram (P&ID) to highlight the material balance and heat balance of relevant manufacturing processes reflecting the impact of disturbance. Process trend graphs display the real-time trends of the process variables to highlight mutual correlation, propagation, and forecasting. Process alarms activate alarms when the process variable exceeds defined limits, such as high, low, deviation, and velocity, requiring manual interventions. These functions are specialized to achieve the independent management objectives of safety, availability, and productivity.

Because of management objectives, operational KPIs need to be reselected based on carbon neutrality, energy conservation, and profitability. New monitoring functions are required to comprehensively display an increased number of operational KPIs from a higher perspective. As additional management objectives conflict with conventional objectives (for example, energy savings impact productivity, and safety margin impacts profitability), new monitoring functions are required to indicate the balance in KPI achievement across multiple management objectives. In addition, because energy-saving, profitable manufacturing is still new to operators owing to a lack of practical training, new monitoring functions are required to establish the relationship between operational KPIs and management KPIs (cause and effect). Based on these considerations, we

formulated new monitoring functions, including an integrated KPI, KPI tree graph, KPI balance graph, KPI trend graph, and KPI alarm.

2. Integrated KPI

The integrated KPI consists of five management objectives in three organizational roles. **Operational KPIs** are mainly used by operators. Only controllable process variables using PCS were selected. **Technical KPIs** are mainly used by the process and mechanical engineers to bridge the gap between plant management and operations. The KPIs calculated using the operational KPIs were selected from the viewpoints of chemical and mechanical engineering. **Management KPIs** are primarily used in plant management. Aggregated KPIs connected to technical KPIs were selected.

Safety-related KPIs are categorized as **safety KPIs**. Although safety KPIs have already been selected based on a HAZOP study, safety allowance can be defined as a new KPI that shows how the safety margin is sufficiently small. Asset-related KPIs are categorized as **availability KPIs**, including those related to the availability and reliability of plant assets. Productivity-related KPIs are categorized as **production KPIs**, including KPIs related to capacity utilization and throughput. KPIs related to carbon neutrality, energy savings, and product quality are categorized as **energy KPIs** because carbon emissions are linked to energy consumption, and quality giveaway increases energy consumption. Profitability-related KPIs are categorized as **profit KPIs**, including the yield of high-value products, material loss, and quality giveaways. An example of an integrated KPI for the depropanizer process is presented in Table 1. The P&ID of the target depropanizer process are shown in Figure 1.

The fractionator was controlled by two operating variables: material balance and heat balance. The change in material balance indicated a change in the cut point, and a change in the heat balance indicated a change in fractionation with variation in the reflux ratio. The material balance was controlled using the C3 flow rate. The C4 flow rate was automatically adjusted using a bottom-level controller. The heat balance was controlled using a reboiler low-pressure steam (LPS) flow rate. A condenser duty was automatically adjusted because the condenser outlet temperature was controlled by the cooling tower water flow rate. The reboiler return temperature was correlated with the C4 content in C3. The C3 flow rate was changed as needed to satisfy the specifications for C3 (that is, stabilize the reboiler return temperature). Board operators manually change the setpoints of the reboiler return temperature and reboiler LPS flow rate to prevent quality giveaway or off specifications.

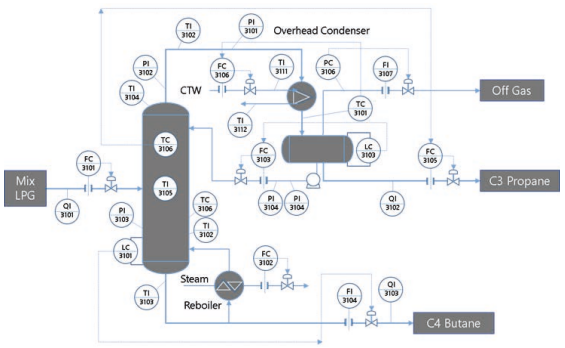


Figure 1. Depropanizer process

Table 1. Example of integrated KPI for depropanizer process

Management Objective	Management KPIs	Technical KPIs	Operation KPIs
Safety KPIs	Safety allowance	Safety allowance	Feed rate, Bottom level, Reflux rate, Reboiler return temp, Overhead drum level, Overhead drum pressure, CTW flow rate, Reflux pump inle: and outlet pressure
Availability KPIs	Availability rate Reliability rate	Availability rate Overhead dew point Dew point allowance Vapor Velocity	Feed rate Overhead dew point Overhead pressure, External reflux temp, Overhead dew point Overhead pressure, Overhead temp, C3 flow rate, Reflux rate, Off gas flow rate
Production KPIs	Capacity utilization rate Feed supply rate Total product rate	Capacity utilization rate Feed supply rate Total product rate C3 product rate C4 product rate	Demonstrated max feed rate, feed rate Planned feed rate, feed rate Planned C3 flow rate, C3 flow rate, Planned C4 flow rate, C4 flow rate Planned C3 flow rate, C3 flow rate Planned C4 flow rate, C4 flow rate
Energy KPIs	CO2 emission Energy consumption Energy efficiency Energy loss	CO2 emission Steam consumption Reboiler duty Reboiler duty ratio (Feed) Reboiler duty ratio (C3) Reboiler duty ratio (C4) Reflux ratio (Feed) Reflux ratio (C3) Reflux ratio (C4) Overhead condenser duty Feed enthalpy C3 enthalpy C4 enthalpy Off gas enthalpy	Low pressure steam flow rate, C02 specific consumption Low pressure steam flow rate Low pressure steam flow rate, Low pressure steam condensing heat Reboiler duty, Feed rate Reboiler duty, C3 flow rate Reboiler duty, C4 flow rate Reflux rate, Feed rate Reflux rate, C3 flow rate Reflux rate, C4 flow rate Off gas latent heat, CTW flow rate, CTW inlet temp, CTW outlet temp Feed rate, Feed specific heat, Feed temp C3 flow rate, C3 specific heat, C3 temp C4 flow rate, C4 specific heat, C4 temp Off gas flow rate, Off gas specific heat, Off gas temp
Profit KPIs	Marginal profit Yield rate Material loss Quality giveaway	Marginal profit C3 yield rate C4 yield rate Material loss C3 in C4 rate C4 in C3 rate	Feed rate, C3 flow rate, C4 flow rate, Off gas flow rate, each unit price Feed rate, C3 flow rate Feed rate, C4 flow rate Feed rate, C3 flow rate, C4 flow rate, Off gas flow rate C3 content in C4 C4 content in C3

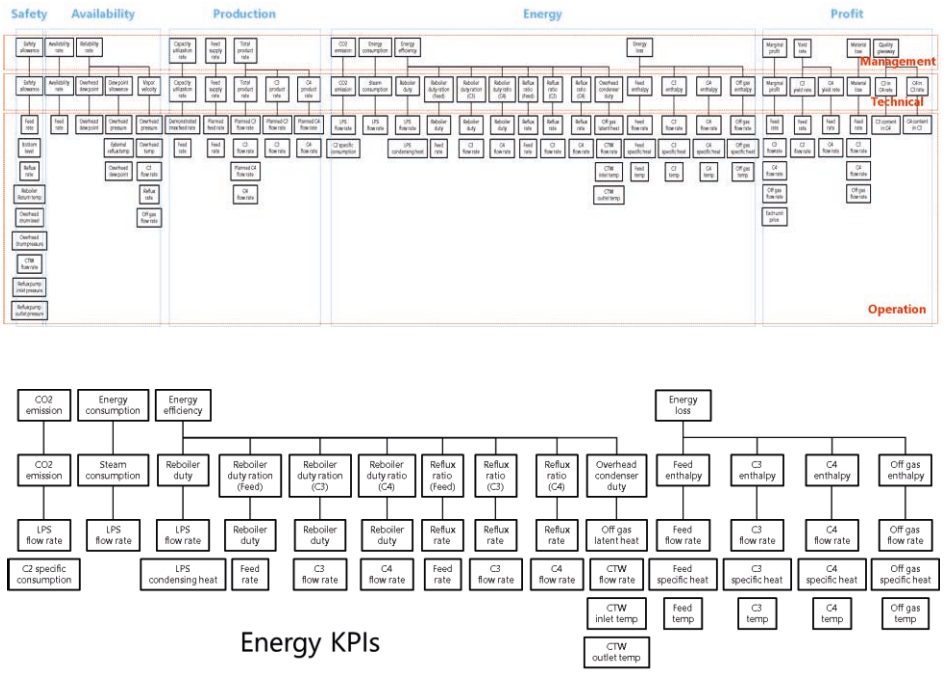


Figure 2. Design example of KPI tree graph

3. KPI tree graph

An example of the design of the KPI tree graph for the depropanizer process is shown in Figure 2. The KPI tree graph displays all the KPI information with mutual connections in a hierarchical manner. This information can help board operators understand which operational KPI impacts which management KPI and encourage board operators to take prompt action against operational KPI alarms before impacting management KPI. If the KPI alarm function is embedded in this graph (for example, the color of the box is changed during alarming), the KPI tree graph can assist board operators in solving all the KPI alarms equivalently toward total optimization.

4. KPI balance graph

An example of the design of the KPI balance graph is shown in Figure 3. This graph displays the KPI achievement scores in a pentagon for each management objective. The shape of the pentagon starts from the largest regular pentagon at the beginning of the day (the KPI score of each management objective starts at 100). The shape is deformed and small in real time when the relevant KPIs are alarming. If all the KPIs are alarming during the day, the shape disappears (the KPI score of each management objective reaches zero). The KPI balance graph helps board operators optimize KPIs across conflicting management objectives, and the scoring function motivates board operators as though it is a game. An example of the score calculation for each management objective in each organizational role (for example, profit operation) is expressed in Eq. (1):

$$\text{KPI Score (\%)} = \frac{1}{x_1}\sum_{n_1=1}^{x_1}\frac{8640-\sum_{n_2=1}^{8640}x_2}{86.4} \tag{1}$$

where x_1 is the number of KPIs (for example, 15 in profit operation). The alarm state of each KPI was determined every 10 s periodically from the beginning to the end of the day (8640 times per day). The value of x_2 is 1(alarming) or 0 (normal).

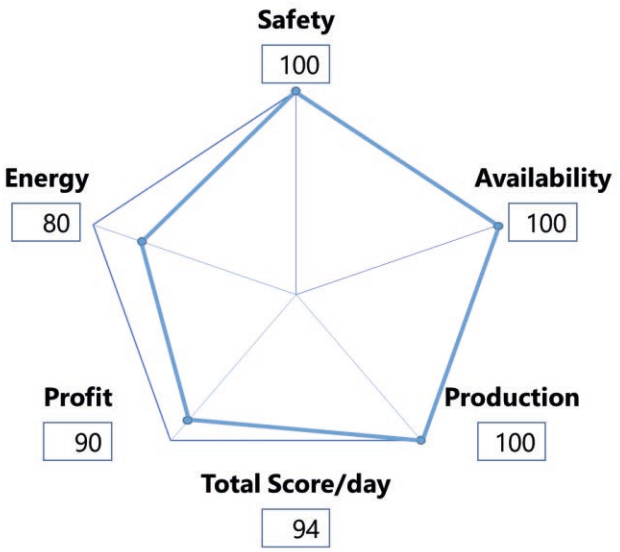


Figure 3. Design example of KPI balance graph

5. KPI trend graph

An example of the design of the KPI trend graph for the depropanizer process is shown in Figure 4. The KPI trend graph shows multiple KPI trend graphs. This graph can help board operators understand the exact propagative order and time lag among the relevant KPIs to identify the root cause. The acknowledgement function can be allocated in each graph because it is crucial to clarify human responsibility for KPI alarm recovery. The event log of the KPI alarm notification, acknowledgement, and recovery should be stored for future operation analysis. The guidance function will be valuable for junior operators to identify the root cause of the KPI alarm and adopt adequate countermeasures until they master plant operations.

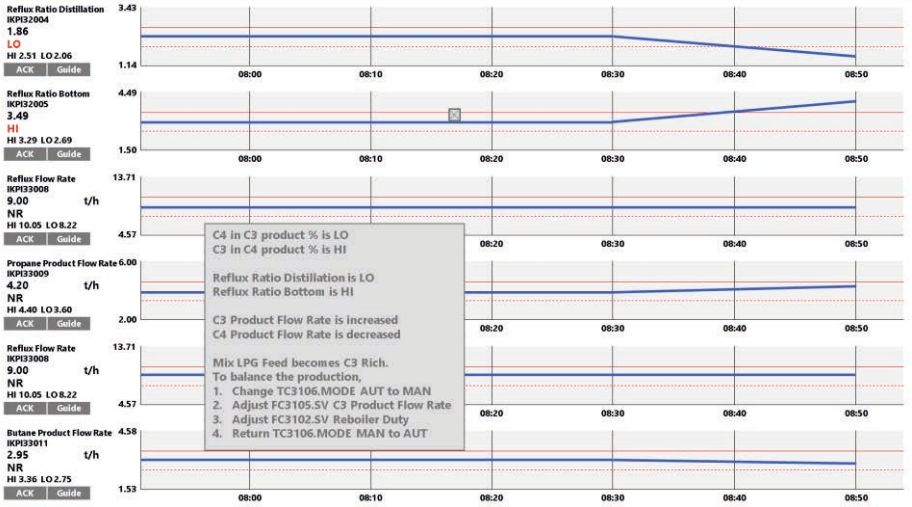


Figure 4. Design example of KPI trend graph

6. Case Study

New monitoring functions for the next generation of PCS worked satisfactorily in the following case study in the depropanizer process, while the fractionator feed (mixed LPG) became C3 rich.

- The operator identified C4 in C3 product in the **LO** alarm (this indicates that quality giveaway was losing profit), and C3 in C4 product is in the **HI** alarm (this indicates that the C4 product was going to be off the specification) on the KPI tree graph.
- Subsequently, the column temperature controller increased the C3 product flow rate to compensate for the lower column temperature. At this moment, the operator identified that the C3 product flow rate increased, the reflux flow rate decreased owing to the constant reboiler duty, and the C4 product flow rate decreased on the KPI trend graph.
- During the adjustment of the column temperature, the operator identified that the reflux ratio distillation (R/C3) was in the **LO** alarm, and the reflux ratio bottom (R/C4) was in the **HI** alarm (this indicated that the vapor and liquid balance in the fractionator shifted) on the KPI tree graph. In the relevant KPI balance graph, the shapes of the profit and energy became slightly small.

Based on the above information, the senior operator identified that the fractionator feed (mixed LPG) became C3 rich. The junior operator reached the same conclusion by checking the guidance message on the relevant KPI trend graph. The following are performed to adjust the fractionator to the optimum operation point (minimum quality giveaway within its specification for both products).

- The operator first turns off the column temperature controller TC3106 (by changing the control mode from AUT to MAN) to accelerate the adjustment.
- Next, the operator adjusts the set point of the C3 flow controller FC3105 (that is, adjusting the material balance) and the reboiler duty–steam flow controller FC3102 (this means adjusting the heat balance).
- After adjusting the column operation based on the economical guidance message, the operator turned on the column temperature controller TC3106 (changing the control mode from MAN to AUT) for stable operation.

In this case, new monitoring functions based on the integrated KPI help board operators realize energy-saving and profitable operations before impacting the relevant management KPIs.

7. Conclusion

This paper presents novel monitoring functions for an integrated KPI framework. An example formulation of the functions was demonstrated using a case study plant. We believe that these functions help process manufacturers achieve energy-saving and profitable manufacturing. Further investigations on other case study scenarios, such as the optimization of conflicting KPIs, are planned to analyze the effectiveness of the proposed functions.

References

ISO, 2014, KPIs for Manufacturing Operations Management, ISO22400.

L. Zhu, C. Johnsson, M. Varisco, M. M. Schiraldi, 2018, Key performance indicators for manufacturing operations management – gap analysis between process industrial needs and ISO 22400 standard, Procedia Manufacturing, 25, 82-88.

W. Hu, A. Almansoori, P.K. Kannan, S. Azarm, Z. Wang, 2012, Corporate dashboards for integrated business and engineering decisions in oil refineries: An agent-based approach, Decision Support Systems, 52, 729-741.

C. F. Lindberg, S.T. Tan, J. Y. Yan, F. Starfelt, 2015, Key performance indicators improve industrial performance, Energy Procedia, 75, 1785-1790.

C. Groger, M. Hillmann, F. Hahn, B. Mitschang, E. Westkamper, 2013, The operational process dashboard for manufacturing, Procedia CIRP, 205-210.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50229-3

Evaluation of risk in the biodiesel production process with supercritical ethanol

Diana Berenice Vega-Guerrero[a], Fernando I. Gómez-Castro [a,*], Antioco López-Molina [b]

[a]*Departamento de Ingeniería Química, División de Ciencias Naturales y Exactas, Campus Guanajuato, Universidad de Guanajuato, Noria Alta s/n Col. Noria Alta, Guanajuato, Gto. 36050, Mexico*
[b]*División Académica Multidisciplinaria de Jalpa de Méndez, Universidad Juárez Autónoma de Tabasco, Carr. Villahermosa-Comalcalco km. 27, Jalpa de Méndez, Tabasco 86205, Mexico*
fgomez@ugto.mx

Abstract

Production of biodiesel can be performed by using supercritical alcohols as reactants, avoiding the use of catalysts. Such processes show high reaction rates and have the advantage of treating raw materials with high contents of free fatty acids or water. Thus, supercritical treatments are appropriate strategies to transform low-cost raw materials. Nevertheless, one of the main concerns on these processes is related to safety issues due to the high-pressure levels required to obtain the desired yields. In this work, a process to produce biodiesel from a waste cooking oil treated with supercritical ethanol is designed using simulation tools. A strategy to assess the risk of this production scheme is developed in terms of a set of indicators. Through this strategy, the most hazardous streams in the process are detected, and the potential consequences of an accident are assessed. This information will be helpful for future work to establish strategies to reduce the risk in these high-pressure production processes and to design processes under inherent safety principles.

Keywords: biodiesel, supercritical ethanol, risk assessment, Hazard Process Stream Index.

1. Introduction

Production of biofuels has taken importance on the last years due to the continuous search for cleaner alternatives to provide energy in the transport and industrial sectors. Biodiesel is one of the most known biofuels, consisting on a mixture of alkyl esters. It can be used blended with fossil diesel to run engines, reducing emissions of CO, particulate matter, and unburned hydrocarbons (Teixeira et al., 2012). Such biofuel is traditionally produced by treating vegetable oils with a short-chain alcohol in the presence of a catalyst, commonly sodium hydroxide. Nevertheless, the main contribution to the product's final cost is the price of the raw material, representing 60-80% of the production cost (Lee et al., 2011). Thus, low-cost oils must be used as raw materials, e.g. waste cooking oil. This represents a disadvantage for the homogeneous basic catalyst route since waste oils have high concentrations of free fatty acids, which promotes saponification when using basic catalysts. To overpass such issue, acid pretreatments have been developed, followed by basic transesterification.

Nevertheless, this implies the need for additional neutralizing and recovery steps. An alternative approach implies the use of supercritical fluids to treat the oils, as supercritical methanol (Saka and Kusdiana, 2001) and supercritical methyl acetate (Saka and Isayama, 2009). In terms of costs and emissions of carbon dioxide, the one-step supercritical methanol approach is the most promissory (Gómez-Castro et al., 2015). Ethanol can also be used as a supercritical reactant, opening the possibility of using renewable ethanol. Among the advantages of using supercritical alcohols, it can be mentioned that the formation of two liquid phases in the reaction system is avoided, increasing the reaction rate.

Moreover, transesterification and esterification occur simultaneously. Therefore, the presence of free fatty acids does not affect the performance of the reaction. Additionally, water in the feedstock has no adverse effects on the yield (Kusdiana and Saka, 2004). This turns the supercritical treatments into an appropriate strategy to transform low-cost raw materials. Nevertheless, one of the main concerns on these processes is given by safety issues due to the high-pressure levels required to obtain the desired yields.

Sometimes, process safety has been considered as a verification criterion or as a governmental requirement. However, it is important to include process safety as a design and decision criterion from the early design stages in future projects, especially for processes and technologies under development, such as high-pressure biodiesel production. There are several features in the high-pressure biodiesel technologies that make it hazardous. These processes work at high pressure (up to 20 MPa); such condition propitiates many flammable substances in case of the loss of containment scenario. Furthermore, the temperature used in these processes (270 – 380°C) increases the formation rate of the vapor cloud in case of a flammable substance's release scenario. At the same time, the hot surfaces can act as an ignition source (Gómez-Castro and López-Molina, 2020). Therefore, the identification of dangerous process streams from an early stage allows generating inherently safe designs. Currently, published indices are focused on identifying explosion scenarios. However, these indices cannot accurately identify hazards on biofuel and bioproduct production processes, where combustible or aqueous mixtures are involved. The Hazardous Process Streams Index (HPSI) is one of the alternatives to identify hazards in the early stages of process design, improving inherent safety (López-Molina et al., 2020). The information needed to estimate process hazards is obtained from process simulation software.

2. Case study

Waste cooking oil is used as raw material, which is modeled with a composition of 94% triglycerides and 6% free fatty acids, with a distribution of 12.3 wt% of palmitic acid, 24 wt% of oleic acid, 56.6 wt% of linoleic acid and 7.1 wt% of linolenic acid for each fraction (Bulla Pereira, 2014). On the other hand, ethanol is used as a supercritical reagent. This alcohol can be obtained from biomass, which contributes to reducing the environmental impact of the process (Gómez-Castro et al., 2017).

A small-scale biodiesel production plant falls into a production range of less than 10,000 t/y, while a large-scale plant is in a range of more than 50,000 t/y (Skarlis et. al., 2012). The above has been taken as a basis for defining the case studies in this work to compare the production of biodiesel at small, medium, and large scales. Thus, the process design will be carried out with three production levels: 10,000, 30,000, and 50,000 tons per year.

3. Process simulation

The process is designed and analyzed with the support of the Aspen Plus® V.8 software. First, the oil and ethanol must be conditioned to the reaction conditions, 20MPa and 300°C. Subsequently, both streams enter the reactor, where triglycerides and free fatty acids are converted to ethyl esters. After the reaction, the pressure is reduced to ambient conditions. Next, the excess ethanol is recovered. The bottoms stream has two phases: phase 1 (biodiesel mainly) and phase 2 (glycerol with a small amount of water), which are separated in a decanter. Figure 1 shows the process diagram.

The reactor is modeled using the RPlug module, using the RK-Aspen model to represent the thermodynamics in the system. The reactor dimensions are set to achieve the highest possible biodiesel yield, verifying that the BIOD stream complies with the corresponding ASTM specifications.

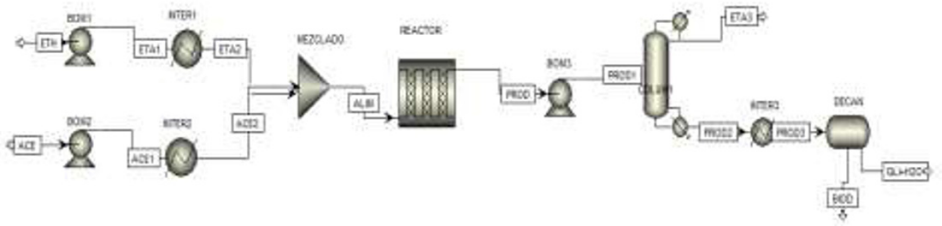


Figure 1. Biodiesel production process with supercritical ethanol in one step.

4. Risk assessment

Once the simulation has been carried out, information on the heat of combustion, molar fluxes, density, and operating pressure is extracted. Together with the flash point of the pure components, these parameters are used to determine the hazardous equipment employing the Hazardous Process Stream Index (HPSI). Once the dangerous streams are identified, it is necessary to define the consequences and process risk. The relative risk approach is used to compare and rank the risks of each alternative (López-Molina et al., 2020). The entire process relative risk (R_T) is the average of the individual risks. As such, an increase in capacity implies an increase in flow and inherently a more significant mass release in the case of loss of containment and its consequences. Therefore, the damage diameter for explosion and fire is estimated through ALOHA software.

5. Results

The results from the HPSI index calculation show the first process stages as the most dangerous due to the use of flammable substances; the same occurs in the three production scales. The analysis indicates that the hazardous streams are ETA1, ETA2, ALIM, PROD, and ETA3. This is explained in terms of the substances present in the stream; for the streams, ETA1, ETA2, and ETA3, the substance present is ethanol, while the ALIM and PROD streams carry glycerol. Table 1 shows the results of HPSI.

Table 1. HPSI results for the 10,000 ton/y scale.

Stream/index	$I_{pressure}$	$I_{density}$	I_{flow}	I_{flash}	$I_{combustion}$	HPSI
ETA	0.012	1.32	1.91	2.33	0.92	0.04
ACE	9.7E-06	0.57	0.04	0.47	1.12	3.8E-07
ETA1	2.31	1.31	1.91	2.33	0.92	7.26
ACE1	2.31	0.59	0.04	0.47	1.12	0.09
ETA2	2.31	0.64	1.91	2.33	0.92	14.88
ACE2	2.31	0.57	0.04	4.67	1.12	0.09
ALIM	2.31	0.59	1.95	4.67	0.97	3.46
PROD	2.31	0.67	1.95	4.67	0.99	3.10
PROD1	0.02	0.29	1.95	4.67	0.99	0.05
ETA3	0.02	0.003	1.79	2.37	0.92	13.78
PROD2	0.02	1.45	0.16	4.67	1.13	0.0011
PROD3	0.02	1.72	0.16	4.67	1.13	9.5E-04
BIOD	0.02	1.75	0.12	4.67	1.19	7.4E-04
GLIC-WATER	0.02	2.53	0.04	4.67	0.57	8.3E-05

The HPSI values for each scale are the same since the substances and process conditions are identical for each stream. It is worth mentioning that the HPSI does not consider the effect of production scaling. Therefore, no change in the values of this index is reflected. The pressurized reactor (REAC) and the distillation tower (COLUM) are considered as the hazardous equipment, defining an explosion as a probable accident due to the flammability of ethanol (class 3, according to NFPA). On the other hand, glycerol has low flammability (class 1, according to NFPA), so this stream has a low probability of fire.

The effect of increased production on the risk was evaluated by the relative risk of each process. Table 2 shows that the largest production scale has a high-risk level, confirming that the consequences will be more significant.

Table 2. Relative risk calculated using the HPSI index.

Production scale (t/y)	Relative total risk (R_T)	Risk level
10,000	0.2	Low risk
30,000	0.6	Medium risk
50,000	1.0	High risk

The scenario of explosion and fire for the two dangerous equipment (REAC and COLUM) shows the effect of process capacity on the consequences of an accident. Table 3 presents the affectation distance in case of explosion and fire. These distances represent values between 3.1 and 1.1 psi for explosion and the flammable area for a fire with 10 minutes of duration.

Table 3, Consequences of explosion and fire for each scale production

	10,000		30,000		50,000	
	REAC	COLUM	REAC	COLUM	REAC	COLUM
Explosion (m)	20	23	25	27	30	32
Fire length (m)	46	49	70	75	90	95
Fire width (m)	14	15	17	18	20	22

6. Conclusions

This paper presents a comparative study on the magnitude of hazard between supercritical processes and their scale of biodiesel production, applying an improved inherent safety index. This index presents a measurement-oriented approach that quantifies the safety of the process at the design stage and considers safety scores based on individual streams, providing a more intuitive and easier way to identify the hazard of each stream. Based on the results obtained, it has been observed that the most significant risk that can occur is a fire or explosion due to the inherent flammability of ethanol. The relative risk is high in the case of a production scale of 50,000 t/y. Therefore, it should be considered of vital importance to improve the safety of the process. Finally, the risk associated with the production scale is validated by the analysis of the consequences.

Acknowledgments

The authors acknowledge the support given by Universidad de Guanajuato and Universidad Juárez Autónoma de Tabasco for the development of this project.

References

F.I. Gómez-Castro, A. López-Molina, 2020, Non-catalytic production of biodiesel: energy and safety considerations, Biofuels: Advances in Research and Applications, 1-29

F.I. Gómez-Castro, J.G. Segovia-Hernández, S. Hernández-Castro, V. Rico-Ramírez, C. Gutiérrez-Antonio, A. Briones-Ramírez, I. Cano-Rodríguez, Z. Gamiño-Arroyo, 2015, Analysis of alternative non-catalytic processes for the production of biodiesel fuel, Clean Technologies and Environmental Policy, 17, 2041-2054

D. Kusdiana, S. Saka, 2004, Effects of water on biodiesel fuel production by supercritical methanol treatment, Bioresource Technology, 91, 3, 289-295

S. Lee, D. Posarac, N. Ellis, 2011, Process simulation and economic analysis of biodiesel production processes using fresh and waste vegetable oil and supercritical methanol, Chemical Engineering Research and Design, 89, 12, 2626-2642

A. López-Molina, A.A. Huerta-Martínez, F.I. Gómez-Castro, C. Conde-Mejía, 2020, A novel approach to identify hazards in non-conventional/intensified biofuels production processes, Chemical Engineering and Processing – Process Intensification, 157, 108139

S. Saka, D. Kusdiana, 2001, Biodiesel fuel from rapeseed oil as prepared in supercritical methanol, Fuel, 80, 2, 225-231

S. Saka, Y. Isayama, 2009, A new process for catalyst-free production of biodiesel using supercritical methyl acetate, Fuel, 88, 7, 1307-1313

E.C. Teixeira, C.D.P. Mattiuzi, S. Feltes, F. Wiegand, E.R.R. Santana, 2012, Estimated atmospheric emissions from biodiesel and characterization of pollutants in the metropolitan area of Porto Alegre-RS, Anais da Academia Brasileira de Ciências, 84, 3, 655-667

A.E. Bulla Pereira, 2014, "Design of the biodiesel production process from frying oils". Master's Thesis in Engineering - Mechanical Engineering, Universidad Nacional de Colombia.

F.I. Gómez-Castro, M.G. Aldana-González, C. Conde-Mejía, C. Gutiérrez-Antonio, A.G. Romero-Izquierdo, R. Morales-Rodríguez, 2017, "Process integration for the supercritical production of biodiesel and the production of lignocellulosic bioethanol", *Computer Aided Chemical Engineering*, Vol. 40, p. 931-936.

Str. Skarlis, E. Kondili, J.K. Kaldellis, 2012, "Small-scale biodiesel production economics: a case study focus on Crete Island", *Journal of Cleaner Production*, Vol. 20, No. 1, p. 20-26.

A. López-Molina, D. Sengupta, C. Shi, E. Aldamigh, M. Alandejani, M. Mahmoud El-Halwagi, 2020, "An Integrated approach to the design of centralized and decentralized biorefineries with environmental, safety, and economic objectives", *Processes*, Vol. 8, No. 12, 1682.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50230-X

Process Monitoring Based on Deep Neural Networks with Continuous Wavelet Transform

Chinatsu Ukawa[a*], Yoshiyuki Yamashita[a], and Seiji Hotta[a]

[a]*Tokyo University of Agriculture and Technology, 2-24-16 Naka-cho, Koganei, Tokyo 184-8588, JAPAN*
s193040z@st.go.tuat.ac.jp

Abstract

In this paper, we proposed a novel fault detection method using a three-dimensional convolutional neural network (3DCNN), and continuous wavelet transform (CWT). Multivariate time series data of a chemical plant was transformed into time-frequency scalogram using CWT, and the 3DCNN was trained with the CWT scalogram. The output obtained was a binary unit representing whether there was a fault or not. The proposed method was applied to the Tennessee Eastman process datasets for fault detection, since they are widely used for verifying the performance of fault detection methods. Our method yielded accurate results for fault detection.

Keywords: Fault Detection; Deep Neural Network; Continuous Wavelet Transform; Time Series Data

1. Introduction

Fault detection and diagnosis play an important role in monitoring chemical processes, where it is expected to prevent severe accidents that cause emergency shutdowns, financial losses, environmental disruptions, and risk to human life. Process monitoring starts with fault detection. Fault detection methods are divided into two types: model-based methods, and data-driven methods. Model-based methods depend on first-principles modelling, which is generally difficult to perform because of process complexity. However, data-driven methods are established without a large amount of process knowledge, and mathematical operations, due to which they have versatility in industrial applications. Additionally, they are quicker to implement than the model-based methods. Numerous studies on fault detection and diagnosis have been conducted so far. In the early stages of fault detection, a univariate analysis with a limit check was applied. However, fault data could not be found because of the highly correlated variables in the chemical processes. To find the buried fault data, a multivariate statistical process was adopted after the univariate analysis. For instance, Principal Component Analysis (PCA), Partial Least Squares, and Support Vector Machine, which are widely applied in industrial processes, were used. However, these methods do not consider the dynamics. Wenfu K (1995) introduced Dynamic PCA as the method which can handle process dynamics. Neural network-based methods have attracted attention in many fields because of their flexibility in modelling complex structures, and dynamics. Various neural network models have been proposed in chemical engineering fields, such as Deep Belief Networks (Zhang et al., 2017), Recurrent Neural Networks (Shaodong et al., 2020), Long Short-Term Memory (Han et al., 2020), and Convolutional Neural Network (Hao Wu et al., 2018).

An important step in the establishment of data-driven methods is pre-processing. A high detection rate cannot be achieved without an appropriate pre-processing method, even if the model possesses great ability. So far, signal processing methods, such as Fourier transform, fast Fourier transform, short-time Fourier transform, and wavelet transform have been applied to various fields, and have exhibited good performance. Short-time Fourier transform and wavelet transform generate time-frequency spectrograms from the original signal data. Wavelet transform is expected to be applicable to chemical process data as it is more suitable at handling non-steady signal analysis, when compared to the Fourier transform. Despite the aforementioned facts, the pre-processing step has not been considered in many previous methods of fault detection for chemical processes. Furthermore, they do not consider the changes in the time-frequency domain of the chemical process data.

In this paper, we proposed a novel fault detection method using a three-dimensional convolutional neural network (3DCNN), and continuous wavelet transform (CWT). We transformed multivariate time series samples into time-frequency scalogram using CWT. After this processing, we trained the 3DCNN with the scalogram as inputs. The output was a binary unit representing whether there was a fault or not.

2. Method

2.1. Continuous wavelet transforms

The CWT produces time-frequency domain information from an original time series data. It is also commonly applied to various research fields, such as transient analysis, electrocardiogram analysis, and fault detection and diagnosis of bearings (Guo. et al., 2018) (Zhang et al., 2020). Unlike the Fourier transform (FT), CWT uses a wavelet to map an objective function onto the time-frequency space. FT does not reflect the information of time and space because it assumes that the signal can be divided into infinite sine waves (Serizawa et al., 2017). This method is not suitable for the analysis of non-stationary data. Short-time Fourier transform (STFT) is an improved method that analyzes signals, using a particular window size fixed for all frequencies. Due to this property, STFT cannot analyze signals accurately if non-linearity remains in the viewing window. However, CWT is more flexible than STFT, and it does not require the definition of a window size. CWT expresses the original signal with the mother wavelet by shifting and scaling. The selection of an appropriate type of wavelet can be easily changed according to engineering conditions (Zhang et al., 2020). Accordingly, CWT is appropriate for handling non-steady time-series data such as chemical process data. Equation (1) represents the CWT of x(t). For a given mother wavelet ψ, 'a' is the scale parameter related to the frequency concept, and 'b' is the shift parameter related to position.

$$W_x(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\, \psi^* \left(\frac{t-b}{a} \right) dt \tag{1}$$

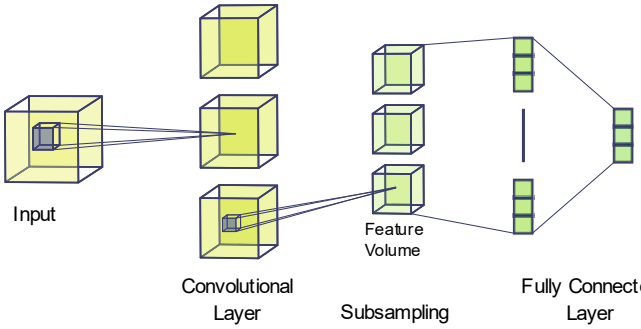


Figure 1 The simple description of 3D convolutional neural network

2.2. 3D convolutional neural network

We adopted a 3DCNN for feature extraction and classification, to handle multi-variable datasets. It is a logical extension of 2DCNN, which works with three-dimensional data such as video, and has an additional temporal dimension in addition to the X and Y coordinates. The first CNN was developed by Fukushima (1980). It could automatically complete feature learning without prior knowledge, and has an efficient recognition ability (Lu Deng, et al. 2021). 2DCNNs have been successful in image-based feature extraction and classification. A 3DCNN was proposed for activity recognition, which required a temporal domain. The basic 3DCNN architecture is illustrated in this slide (Figure 1). The 3D filter is convolved with 3D data to produce the feature volumes. After subsampling and flattening, the features are fed to a fully connected layer for classification.

2.3. Proposed model

First, the chemical process data was divided by a moving time window. In our case, the moving window width was 20 and the stride was 1. Divided data were transformed into a spectrogram using a continuous wavelet transform. We set the size of the spectrograms to (20 x 20) during transformation. We performed pre-processing for each variable. Finally, the shape of the input data was (20 × 20 × the number of variables). The dimensions of the input data were (20 × 20 × 33) because we used 22 process measurement variables, and 11 manipulated variables. Next, the input data was fed into the 3DCNN, and the model was trained. The output was a binary unit representing the existence of the fault.

3. Case study

We evaluated the ability of our proposed method using the Tennessee Eastman process datasets. The Tennessee Eastman (TE) process was first introduced by Downs et al. (2013), and is based on a real process. It is widely used as a benchmark for evaluating fault detection and diagnosis methods (Onel et al. 2018) (Rato et al. 2017). It consists of five main units: a reactor, condenser, compressor, separator, and stripper. Four reactants and two products are included. It has 52 variables, including 41 process variables, and 11 manipulated variables. In this study, the data of the TE process were obtained from the extended database published by Rieth et al. (2017). This dataset includes 500 simulation runs for the training, and testing data of the normal, and 20 fault types. Each simulation

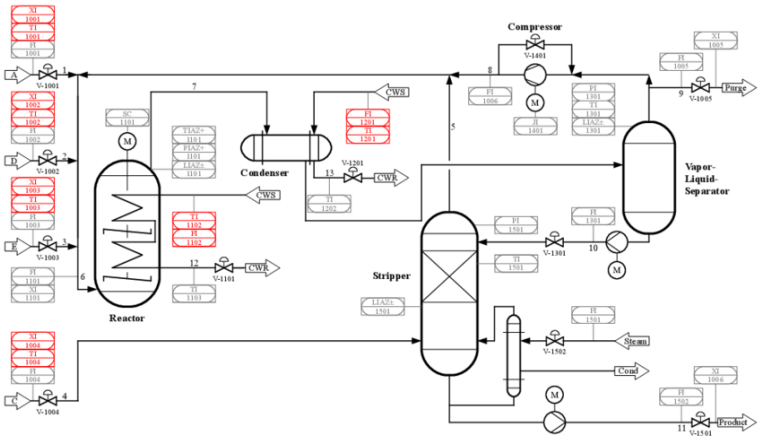


Figure 2 Process flow sheet of Tennessee Eastman process (Andreas, B., et al, 2015)

run involved 500 data samples for training, and 960 data samples for testing. The sampling period was set to 3 min (20 samples per hour). Each fault was introduced to the process after sample numbers 20, and 160 for fault training, and fault testing, respectively. The process flowsheet is shown in Figure 2. We selected this dataset because the conventional dataset was insufficient to train the CNN model. The datasets used in this study were downloaded from the Harvard database. In this study, 22 process variables and 11 manipulated variables were selected, and used to develop the fault detection model.

4. Result

4.1. Metrics for evaluation

We evaluated the proposed method and compared its performance using the fault detection rate (FDR). TP represents true positive, and FN represents false negative.

$$Fault\ Detection\ Rate = \frac{TP}{TP + FN} \tag{1}$$

4.2. Application to TEP

We compared the proposed model with PCA. PCA is the most widely used model for fault detection. The number of principal components was set to nine, and the confidence interval was set to 0.99. Hotelling T^2 and Squared Prediction Error (SPE) were calculated for each fault. A comparison of each model is presented in Table 1. The proposed method tended to improve the fault detection rate of faults in which PCA could not represent high performance.

5. Discussion

We selected fault 5 and considered the effectiveness of the CWT for feature extraction. Fault 5 was a step-like change caused by the condenser cooling water inlet temperature. It affected the variables surrounding the reactor and condenser. We selected variables 32 and 33 from the process flow. Variable 32 was the reactor cooling water flow, and variable 33 was the condenser cooling water flow. They were located around the reactor and condenser. The resultant spectrograms are shown in Figure 3. They were selected

for each of the 10 steps. The spectrogram of each variable showed different characteristics between normal operating condition, and fault 5. This indicated that the change in variables appeared in the time-frequency domain. We believe that this difference implies the effectiveness of the proposed method for detecting faults.

6. Conclusions

In this paper, we proposed a fault detection model for chemical processes based on 3DCNN and CWT. 3DCNN was applied for multivariable analysis. CWT was applied for pre-processing the chemical process data because of its good performance for non-steady signal processing. First, the chemical process data were transformed to a time-frequency spectrogram using CWT. The 3DCNN model was built with convolutional layers, pooling layers, dropout layers, and fully connected layers and trained with a spectrogram. The output was a binary unit representing the normal operation, or faulty condition. The results of the case study with TEP showed that the proposed method performed well on fault detection of TEP faults. Particularly, for faults in which PCA could not show good performance, the proposed method had a better score. It represented the time-frequency domain with the feature of process data.

Table 1 Accuracy comparison of models

Status	PCA(T2)	PCA(SPE)	Proposed	Status	PCA(T2)	PCA(SPE)	Proposed
Normal	0.989	0.990	0.983				
Fault 1	0.991	0.997	0.985	Fault 11	0.255	0.755	**0.857**
Fault 2	0.985	0.949	0.964	Fault 12	0.968	0.949	**0.981**
Fault 3	0.013	0.010	**0.030**	Fault 13	0.934	0.919	0.919
Fault 4	0.086	1.000	0.997	Fault 14	0.862	1.000	0.982
Fault 5	0.245	0.114	**0.936**	Fault 15	0.015	0.012	**0.167**
Fault 6	0.989	1.000	0.989	Fault 16	0.078	0.114	**0.390**
Fault 7	0.364	1.000	0.986	Fault 17	0.719	0.897	**0.946**
Fault 8	0.955	0.861	0.934	Fault 18	0.929	0.937	**0.939**
Fault 9	0.014	0.011	**0.025**	Fault 19	0.022	0.295	0.189
Fault 10	0.210	0.172	**0.419**	Fault 20	0.263	0.460	**0.825**

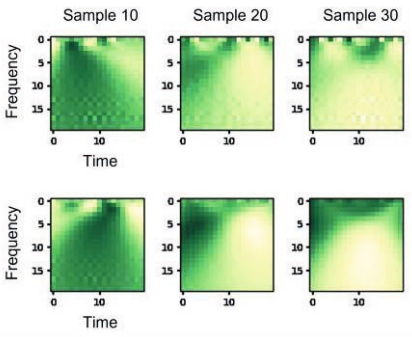


Figure 3 The spectrograms of variable 32 on normal operating condition (the first row) and fault 5 (the second row)

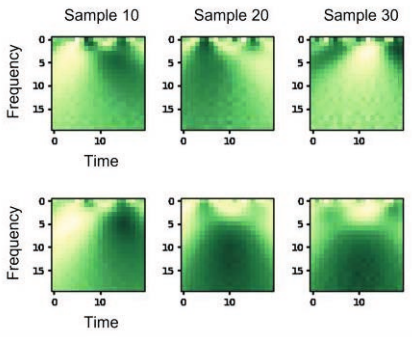


Figure 4 The spectrograms of variable 33 on normal operating condition (the first row) and fault 5 (the second row)

References

A. Bathelt, N. L. Ricker, M. Jelali, 2015, Revision of the Tennessee Eastman Process Model, *IFAC-PapersOnLine*, 48, Issue 8, 309-314

L. Deng, Y. Zhang, Y. Dai, X. Ji, L. Zhou, Y. Dang, 2021, Integrating feature optimization using a dynamic convolutional neural network for chemical process supervised fault classification, Process Safety and Environmental Protection, 155, 473-485

J.J. Downs, E.F. Vogel, A plant-wide industrial process control problem, 1993, Computers & Chemical Engineering, 17, 3, 245-255

K. Fukushima, 1980, Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position, Biological Cybernetics, 36, 193-202

S. Guo, T. Yang, W. Gao, C. Zhang, Y. Zhang, 2018, An Intelligent Fault Diagnosis Method for Bearings with Variable Rotating Speed Based on Pythagorean Spatial Pyramid Pooling CNN, *Sensors*, 18(11), 3857

Y. Han, N. Ding, Z. Geng, Z. Wang, C. Chu, 2020, An optimized long short-term memory network based fault diagnosis model for chemical processes, Journal of Process Control, 92, 161-168

M. Onel, C. A. Kieslich, Y. A. Guzman, E. N. Pistikopoulos, 2018, Simultaneous Fault Detection and Identification in Continuous Processes via nonlinear Support Vector Machine based Feature Selection, Computer Aided Chemical Engineering, Elsevier, 44, 2077-2082,

W. Ku, R. H. Storer, C. Georgakis, 1995, Disturvance detection and isolation by dynamic principal component analysis, Chemometrics and Intelligent Laboratory systems, 30, 1, 179-196

T. J. Rato, M. S. Reis, 2017, Improved Fault Diagnosis in Online Process Monitoring of Complex Networked Processes: a Data-Driven Approach, Computer Aided Chemical Engineering, Elsevier, 40, 1681-1686

C. A. Rieth, B. D. Amsel, R. Tran, M. B. Cook, 2017, "Additional Tennessee Eastman Process Simulation Data for Anomaly Detection Evaluation", https://doi.org/10.7910/DVN/6C3JR1, Harvard Dataverse, V1

T. Serizawa, T. Suzuki, S. Choi, and O. Sasaki, 2017, 3-D surface profile measurement using spectral interferometry based on continuous wavelet transform, *Optics Communication*, 396, 216-220

H. Wu, J. Zhao, Deep convolutional neural network model based chemical process fault diagnosis, 2018, Computers & Chemical Engineering, 115, 185-197

M. Zhang, M. Li, J. Zhang, L. Liu, H. Li, 2020, Onset detection of ultrasonic signals for the testing of concrete foundation piles by coupled continuous wavelet transform and machine learning algorithms, Advanced Engineering Informatics, 43, 101034

Z. Zhang, J. Zhao, A deep belief network based fault diagnosis model for complex chemical processes, 2017, Computers & Chemical Engineering, 107, 395-407

S. Zheng, H. Zhao, 2020, A new unsupervised data mining method based on the stacked autoencoder for chemical process fault diagnosis, Computers & Chemical Engineering, 135, 106755

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50231-1

Methyl sec-butyl ether content estimation in MTBE products via clustering-based adaptive nonlinear soft sensors

Andrei Torgashov[a], Oleg Snegirev[a], Fan Yang[b*]

[a]*Process Control Laboratory, Institute of Automation and Control Process FEB RAS, 5 Radio Str., Vladivostok, Russia*
[b]*Beijing National Research Center for Information Science and Technology and Department of Automation, Tsinghua University, Beijing 100084, China*
yangfan@mail.tsinghua.edu.cn

Abstract

The problem of developing an adaptive soft sensor (SS) using clustering methods is considered to be an example of an industrial reactive distillation process for the production of methyl tert-butyl ether (MTBE). We propose the use of clustering methods to assess the feasibility of updating the model parameters. An algorithm for the functioning of an adaptive SS using a *moving window* and clustering is proposed and tested on industrial data. The reduction of the computational cost in comparison with the adaptive SS using the moving-window method with continuous updating of the model parameters is shown. The accuracy of adaptive SSs based on various nonparametric methods is compared.

Keywords: soft sensing modeling; clustering; adaptation; methyl sec-butyl ether.

1. Introduction

Currently, data from the laboratory and in-line analyzers are used to control the quality of the output products of distillation columns. However, the results of analysis obtained in laboratories do not allow quality management in real time, as they have an insufficient level of efficiency. Therefore, soft sensors are used in production to solve this problem. Soft sensing is a data analysis tool that allows one to get an estimate of a difficult-to-measure output variable using the values of variables available for measurement at any given time. In comparison with rigorous modeling methods based on physical-chemical laws, data-driven soft sensors are more flexible and can be easily implemented in the process industry.

Basically, the soft sensors used in production are usually based on linear regression models, which imply the assumption of a linear structure in the model of a plant. However, this is not always true in real conditions. The use of nonparametric regression methods is free from the requirement of knowledge of the model structure. There exist many nonparametric regression methods. The most popular methods include kernel, spline, and local polynomial kernel. Neural network-based models are also very popular (Wang et al., 2019), allowing extremely complex dependencies to be reproduced based on representative data as a result of training. In this paper, we consider soft sensors based on Alternating Conditional Expectations (ACE) (Snegirev and Torgashov, 2021), Kernel-based orthogonal projections to latent structures (K-OPLS) (Li et al., 2020), and neural networks (Zhu et al., 2018).

In real conditions, the parameters of a plant often change over time, and therefore there is a problem of the degradation of the model, which leads to a deterioration in the accuracy of the evaluation of the output variable. To solve this problem, adaptive soft sensors are used to adjust models using new measurement data. Typical adaptive soft sensors are based on methods such as just-in-time (JIT) (Guo et al., 2020), time difference (TD), and moving window (MW) approaches (Urhan and Alakent, 2020). In this paper, we consider methods for adapting soft sensors based on MW, the essence of which is to update the SS model using training samples. From these training samples, when a new observation appears, the earliest observation is removed, and a new one is added. In well-known methods, when developing an adaptive SS based on MW, clustering methods are used to divide data into clusters and build a separate model for each cluster.

This paper presents an algorithm for the operation of an adaptive soft sensor based on MW using clustering methods to determine the need to update the SS. The superiority of the algorithm in terms of reducing computational costs under the updating of model parameters is shown.

2. Industrial plant description and problem formulation

The chemical plant of research is a mass-transfer (reactive distillation) process (MTP). A specific feature of this MTP is that a reversible chemical reaction occurs simultaneously with the partial or almost complete separation of the resulting mixture by distillation. The product of MTP is methyl tert-butyl ether (MTBE), which is widely used in the production of high-octane gasolines.

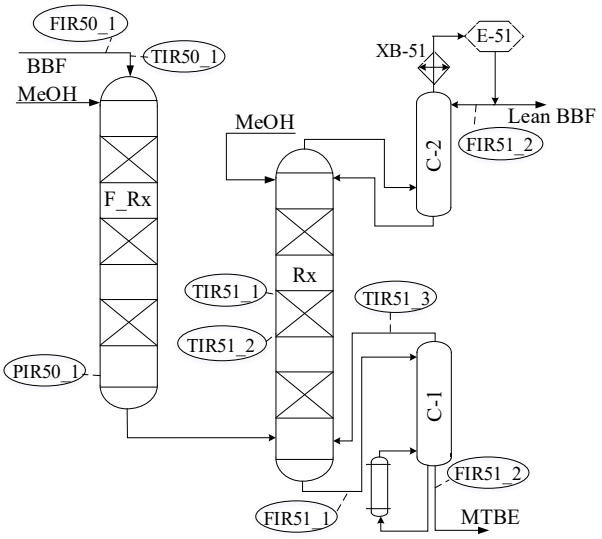


Figure 1: Flowsheet diagram of the reactive distillation process

The key indicator of the quality of the MTBE product (Fig. 1) is the concentration (wt,%) of impurity as methyl *sec*-butyl ether (MSBE). We consider the problem of developing SS for MSBE content estimation. The following measured process variables were used as regressors: u_1 – MTBE flowrate (FIR51_2); u_2 – temperature of fresh BBF

in a for-contact (pre) reactor (F-Rx) (TIR50_1); u_3 – the flowrate of fresh BBF in a for-contact (pre) reactor (F-Rx) (FIR50_1); u_4 – the reflux flowrate in C-2 (FIR51_2); u_5 – the temperature of the reacting mass from the top of C-1 in Rx (TIR51_3); u_6 – the flowrate of reacting mass in C-1 (FIR51_1); u_7 – bottom pressure of F-Rx (PIR50_1); u_8 – the temperature in the top of the middle zone of the catalyst bed of Rx (TIR51_1); u_9 – the temperature in the bottom of the middle zone of the catalyst bed of Rx (TIR51_2).

3. Methods used for model evaluation

In this paper, we present a method based on the ACE in comparison with Kernel-OPLS and neural networks.

3.1. Alternating Conditional Expectation (ACE)

A nonlinear soft sensor model can be obtained using the ACE algorithm (Snegirev and Torgashov, 2021):

$$\theta(Y) = b_0 + \sum_{i=1}^{m} \phi_i(u_i),$$

Where the θ - function of output variable Y; b_0 - bias; ϕ_i - optimal transformation of the *i*-th input variable u_i; *m* – number of input variables.

3.2. Kernel-based Orthogonal Projections to Latent Structures (Kernel-OPLS)

The Kernel-OPLS method is a reformulation of the original OPLS method to its kernel equivalent (Li et al., 2020). A distinctive feature of kernel-based methods is the consideration of the kernel matrix as point products in a multidimensional object space.

3.3. Neural networks

Neural networks can also be used to build soft sensors. In this work, a single-hidden-layer neural network is used. The hidden layer contained *S*=10 nodes, equal to the number of input variables. A set of 300 observations of variables was used to train the neural network.

4. Description of the proposed adaptation algorithm

The proposed algorithm is based on the "Moving Window" (MW) method. The proposed algorithm differs from MW in that when a new observation appears, the need to add a new observation to the training set and to update the SS model is evaluated. In this paper, we propose clustering methods to determine the need to update the current SS model. At the initial stage, the initial set of observations is divided into $\boldsymbol{K}$ clusters. Clustering is carried out in the space of the input variables of the model. Algorithm 1 shows a pseudocode describing how the need to update the SS model is assessed when a new observation is obtained.

The Euclidean distance was used as a distance measure in the algorithm (Cardarilli et al., 2020):

$$d(p,q) = \sqrt{\sum_{i=1}^{m} (p_i - q_i)^2}.$$

As one of the criteria for the need to update the SS, we propose a calculation of the difference in time between the appearance of a new observation and the last one in the cluster (time factor), which will allow for timely consideration of the manifestation of non-stationary properties of the object in the model.

To take into account the time factor at time t for cluster k, we propose using the following indicator: $\Delta\tau_t^k = \frac{\tau_t - \tau_{Q_k}^k}{\overline{T}^k}$, where τ_i is the value of the moment of the appearance of a new observation, $\tau_{Q_k}^k$ is the value of the moment of the appearance of the last element in cluster k, $T^k = (\tau_1^k - \tau_0^k, \tau_2^k - \tau_1^k, ..., \tau_Q^k - \tau_{Q-1}^k)$ is the vector of time intervals between the appearance of observations of the k-th cluster, τ_0^k is the initial moment of time, $\overline{T}^k$ is the average value of the vector T^k

Algorithm 1. Assessment of the need to update the SS

1. Calculate the Euclidean distance $d(u_n, \mu_j)$ from the new observation to the ***j***-cluster center, $j = 1...K$
2. Select the cluster **k** for which $d(u_n, \mu_k)$ is minimal
3. **If** there is a small number of observations in cluster **k,** then
4. Add a new observation to the training set. Recalculate cluster centers taking into account the new observation. **Recalculate the parameters of the SS**
5. **Else If** $d(u_n, \mu_k)$ is greater than the maximum distance from the elements of cluster **k** to its center
6. Add a new observation to the training sample. Increase the number of clusters by 1. Recalculate the cluster centers taking into account the new observation. **Recalculate SS parameters**
7. **Else If** MAE of the estimate of the output variable taking into account the new observation are greater than the MAE at the previous time
8. Add a new observation to the training sample. Recalculate cluster centers taking into account the new observation. **Recalculate the parameters of the SS**
9. **Else If** $\Delta\tau^k > 1$ (How long has an observation similar to the new one been encountered)
10. Add a new observation to the training set. Recalculate cluster centers taking into account the new observation. **Recalculate the parameters of the SS**
11. **Else**
12. The new observation should not be included in the training set. **Leave the SS model unchanged**

5. Results and discussion

5.1. Comparison of models accuracy

The accuracy of the soft sensor models was compared on the data of the MTBE production process plant. The mean absolute error (MAE) was used as the evaluation criterion (Wang and Lu, 2018). Table 1 shows the values of MAE at the end time. The value of MAE at the final time for the K-OPLS-based soft sensor was 52% less than

that of the ACE-based soft sensor and 45% less than that of the neural network-based soft sensor for the proposed adaptation algorithm. For the moving window algorithm, the MAE value for the K-OPLS-based soft sensor was 52.3% less than that of the ACE-based soft sensor and 35.5% less than that of the neural network-based soft sensor.

Table 1. MAE at the end time

	Proposed Method	Moving Window
ACE	0.1596	0.15694
K-OPLS	0.0766	0.07485
Neural Networks	0.13945	0.11604

5.2. Comparison of computational costs for adaptive soft sensors

Figures 2 and 3 are graphs of MAE and computational time of adaptive SS based on the ACE and SS based on the K-OPLS. Table 2 presents the value of computational time spent at work adaptive soft sensors based on the ACE, the K-OPLS and of the neural network working on the proposed method and the method of MW.

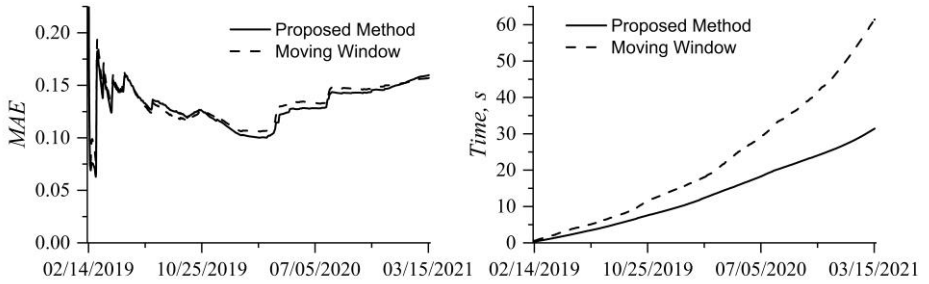


Figure 2: The change of MAE during the process (left) and the time spent on work (right) for the ACE-based soft sensor

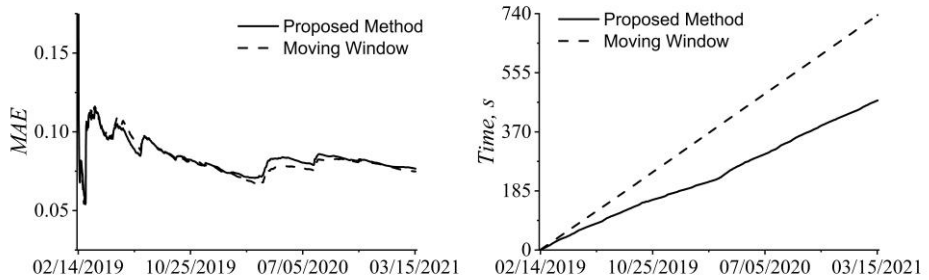


Figure 3: The change of MAE during the process (left) and computational cost (time) spent on updating (right) for the K-OPLS-based soft sensor

Table 2. Computational cost (time) of adaptive soft sensors

	Proposed method	Moving Window
ACE	31.43	61.45
K-OPLS	468.3	735.71
Neural Networks	51.77	86,1

The advantage of the proposed algorithm is to reduce the computational cost. Thus, the ACE-based SS, K-OPLS-based SS and the neural network-based SS during the process

showed approximately the same values of MAE for both the proposed algorithm and the MW, but the reduction in computational time spent updating the model was 48.8% for the ACE-based SS, 36.3% for the K-OPLS-based SS and 39.8% for the neural network-based SS.

6. Conclusions

We have proposed an algorithm for the operation of an adaptive SS based on an MW using clustering methods and taking into account the time factor. During the testing of the algorithm on the data of the reactive distillation process, we have shown that the time for recalculation (updating) of the model parameters was reduced compared to the MW algorithm for the ACE-based SS by 48.8%, for the K-OPLS-based SS by 36.3% and for the neural network-based SS by 39.8%.

We have compared the use of ACE and K-OPLS algorithms and neural networks in the construction of soft sensors. The MAE value at the last moment for a soft sensor based on K-OPLS was 52% less than that of a soft sensor based on ACE, and 45% less than that of a soft sensor based on a neural network for the proposed adaptation algorithm. For the MW algorithm, the MAE value for the K-OPLS-based soft sensor was 52.3% less than that of the ACE-based soft sensor and 35.5% less than that of the neural network-based soft sensor.

7. Acknowledgements

The reported study was partially funded by RFBR and NSFC (project numbers 21-57-53005 and 62111530057).

References

Cardarilli, G.C., Nunzio, L.D., Fazzolari, R., Nannarelli, A., Re, M., Spano, S , 2020, N-Dimensional Approximation of Euclidean Distance, IEEE Transactions on Circuits and Systems II: Express Briefs, 67, 565 – 569

Guo, F., Xie, R., Huang, B., 2020, A Deep Learning Just-In-Time Modeling Approach for Soft Sensor Based On Variational Autoencoder, Chemometrics and Intelligent Laboratory Systems, 197, 103922

Li, J., Li, K., 2020, Application in Soft Sensing Modeling of Chemical Process Based on K-OPLS Method, Journal of Measurement Science & Instrumentation, 11, 17-27

Snegirev, O.Y., Torgashov, A.Y., 2021, Adaptation of the Structure and Parameters of Nonlinear Soft Sensors by the Example of an Industrial Reactive Distillation Process, Automation and Remote Control, 82, 1774–1786

Urhan, A., Alakent, B., 2020, Integrating Adaptive Moving Window and Just-In-Time Learning Paradigms for Soft-Sensor Design, Neurocomputing, 392, 23–37.

Wang, K., Shang, Ch., Lei Liu, Jiang, Y., Huang, D., Yang, F., 2019, Dynamic Soft Sensor Development Based on Convolutional Neural Networks, Industrial & Engineering Chemistry Research, 58 (26), 11521-11531

Wang, W., Lu,Y., 2018, Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model, IOP Conference Series: Materials Science and Engineering, 324, 012049, DOI: 10.1088/1757-899x/324/1/012049

Zhu, W., Ma, Y., Zhou, Y., Benton, M., Romagnoli, J., 2018, Deep Learning Based Soft Sensor and Its Application on a Pyrolysis Reactor for Compositions Predictions of Gas Phase Components, Computer Aided Chemical Engineering, 44, 2245-2250

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50232-3

Early identification of abnormal deviations in nonstationary processes by removing non-stationarity

Cheng Ji, Fangyuan Ma, Jingde Wang*, Wei Sun*

College of Chemical Engineering, Beijing University of Chemical Technology, North Third Ring Road 15, Chaoyang District, Beijing, 100029, China
sunwei@mail.buct.edu.cn, jingdewang@mail.buct.edu.cn

Abstract

In chemical processes, most measurements show stationary characteristic around their set points. However, the means of specific measurements are still time-varying with a wide range of variation, which makes the interval of data under normal operating conditions determined from multivariate statistics process monitoring method relatively large. In such case, the fault signal could be buried by these nonstationary features at its early stage, resulting in a long fault detection time and low fault detection rate. In this work, a process monitoring strategy dealing with nonstationary process is proposed by removing non-stationarity. Stationarity test is first applied to determine measurements with nonstationary characteristic. Support vector regression (SVR) model is then established for each nonstationary measurement using independent variables selected by mutual information. The original nonstationary measurement can be replaced by model residual, which is the difference between the regression value and the measured value, and therefore the nonstationary characteristic can be removed. When the multivariate statistical monitoring model is further established, the interval of normal operating conditions is effectively narrowed, by which the faulty deviation can be identified earlier. In order to verify the proposed method, case studies on Tennessee Eastman process (TEP) are investigated. The results demonstrate a better performance in detecting the abnormal deviations in nonstationary systems compared to other related methods.

Keywords: Chemical nonstationary process; Mutual information; Support vector regression; Process monitoring

1. Introduction

With the rapid development of data measurement and storage technology, multivariate statistical process monitoring (MSPM) methods have received considerable attention because of their unique advantages in processing highly coupled and multivariate data. However, some measurements in chemical industrial processes inevitably show a certain dynamic and nonstationary characteristic due to equipment aging or random disturbances, which violates the assumption of traditional MSPM that the process is time-independent (Li and Yan, 2019). Dynamic principal component analysis (DPCA) applies augmented matrix to describe such time-varying dynamic characteristic of data. However, the selection of lag order is limited for high-dimensional process data and all process variables have to be expended to the same level, while process dynamic has different effects on each variable (Huang et al., 2019). Slow feature analysis, as an unsupervised learning algorithm of extracting constant or slowly changing features of time series, has

been applied for dynamic process monitoring (Shang et al., 2015). The process can be better described by considering both static and dynamic indices, but variances of original nonstationary data are still not processed. When applying MSPM to project high-dimensional data onto a lower dimensional feature space, the interval of normal region will be wide because the main variance of original data is preserved. The fault signal could be buried in nonstationary trends of process variables at its early stage, leading to a long fault detection time. Aiming at the process non-stationarity, cointegration theory has been developed to extract time-invariant characteristics by establishing cointegration relationship among nonstationary variables (Li et al., 2014). Process monitoring can be implemented by distinguishing between normal changes in cointegration relationships and abnormal deviations, but the application of cointegration is still relatively limited as a new method introduced to process monitoring.

In this work, a monitoring strategy for nonstationary process is proposed by removing characteristic of original nonstationary measurements. Process variables are first divided into stationary variables and nonstationary variables by stationarity test. Regression model is then established by mutual information (MI) and support vector regression (SVR) for each nonstationary variable. Model residuals of these variables, which no longer show nonstationary characteristics, are put together with stationary variables to replace the original nonstationary signals. Therefore, the variances of data under normal operating conditions are significantly reduced, the interval of normal operating conditions determined from monitoring statistics of principal component analysis model is effectively narrowed, by which the abnormal deviations can be identified at its early stage. The proposed method is employed to Tennessee Eastman process (TEP) and the performance of process monitoring is compared with related methods. The results show that the fault detection rate can be significantly improved by removing non-stationarity.

2. Methodology

In this section, the preliminaries of methods applied in the proposed process monitoring strategy are introduced.

2.1. Stationarity test

Process stationarity is an important assumption for building statistical models. Unit root test is usually used to test stationarity by judging whether the first-order difference of a time series is stable. In this work, the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test is applied to divide process measurements into stationary part and nonstationary part.

2.2. Mutual information (MI)

MI is a commonly used correlation analysis tool from the aspect of information theory. Considering two random variables X and Y, the MI value can be calculated as follows,

$$MI(X,Y)=\sum_{x,y}p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \tag{1}$$

where $p(x, y)$ is the joint probability distribution function, $p(x)$, $p(y)$ are the marginal probability distribution function of x and y respectively. The probability distribution function can be calculated in Equation (2) by kernel density estimation,

$$p(x)=1/n\sum_{i=1}^{n}K(x-x_i) \tag{2}$$

where K is the kernel function. In this work, MI is applied to select independent variables for its advantage in identifying nonlinear relationship.

2.3. Support vector regression (SVR)

SVR is a statistical machine learning method that has been applied in industrial processes. For a training set $T = \{(\mathbf{x}_i, y_i), i = 1...l\}$, where $\mathbf{x}_i \in \mathbf{R}^N, y_i \in \mathbf{R}$, SVR aims at finding a regression function that can fit all training samples,

$$f(x) = \mathbf{w}^T \Phi(\mathbf{x}) + b \tag{3}$$

where $\mathbf{W}$ is a coefficient vector in feature space, $\Phi(\mathbf{x})$ is a kernel function to map input $\mathbf{x}$ to a vector in feature space and b is an intercept. The solution of $\mathbf{W}$ and b can be obtained by solving the optimization problems in literature (Smola and Schölkopf, 2004). SVR has a great advantage in dealing with nonlinear processes by introducing a kernel function to project the original data into a high-dimensional linearly separable space. Therefore, SVR is selected to build regression models for nonstationary variables.

2.4. Principal component analysis (PCA)

For a high dimensional data set X, PCA can be employed to find a projection direction that maps the data into a lower dimensional space by orthogonal transformation,

$$X = TP^T + E \tag{4}$$

where T is a score matrix, P is a loading matrix, and E is a residue matrix. T^2 statistic and SPE statistic can be further established in the corresponding principal component space and residual space to realize process monitoring.

3. The proposed monitoring strategy by removing process non-stationarity

In this section, the idea of removing process non-stationarity and the implementation procedures of the proposed process monitoring framework are introduced.

3.1. Strategy for removing process non-stationarity

The main contribution of the proposed monitoring strategy is to early identify the abnormal process deviations from normal random disturbances in nonstationary process. If the process non-stationarity can be effectively removed, the fault can be early detected with traditional PCA. On this basis, SVM is applied to extract nonstationary and nonlinear relationships among variables and establish regression models for nonstationary variables. The obtained residues are used as virtual variables to replace nonstationary variables. Therefore, all process variables are processed to be stationary, and then better process monitoring results can be easily obtained with PCA.

3.2. The implementation procedures of the proposed process monitoring framework

The implementation procedures of the proposed process monitoring method are shown in Figure 1. It consists of offline training and online monitoring. For offline training, the first step is to divide training data under normal conditions into stationary part and nonstationary part by KPSS test. For nonstationary variables, MI is first calculated to select independent variables. Then SVR model is established for each nonstationary variable using corresponding independent variables. The optimal parameters are determined by grid search and cross-validation. Once the SVR models are determined,

the difference between the regression value and the measured value of each nonstationary variable is applied as a virtual variable to replace original nonstationary signal. The virtual variables are put back together with original stationary variables, by which the process is converted to stationary. Therefore, a PCA model can be established and control limits of T^2 statistic and SPE statistic can be determined.

For online process monitoring, real-time data are input into the SVR models to get the regression values of nonstationary variables. Real-time virtual variables are calculated from the difference between regression values and measured values. Then virtual variables and stationary variables are put together and brought into the PCA model to calculate T^2 statistic and SPE statistic of the current sample. The monitoring results can be determined by comparing statistics with the control limits.

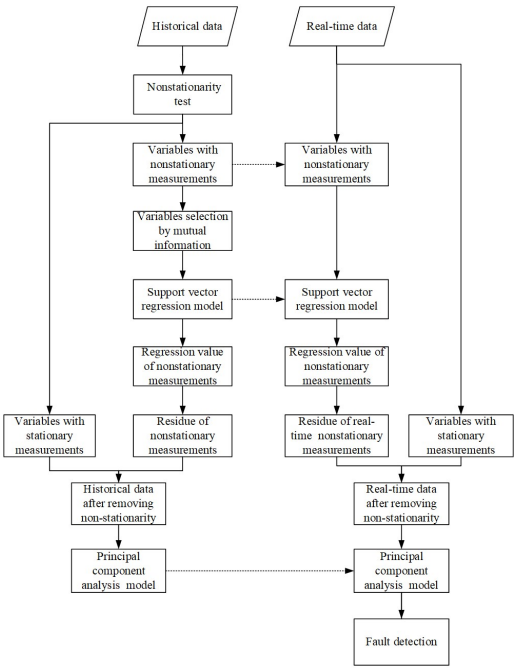


Figure 1 Procedures of the proposed process monitoring framework

4. Case studies on Tennessee Eastman process (TEP)

In this section, the proposed monitoring strategy by removing process non-stationarity is applied to TEP. The results are discussed and compared with other related methods.

4.1. Tennessee Eastman process and data description

TEP is a famous chemical benchmark simulated from an industrial plant (Downs and Vogel, 1993), and has been commonly used to test the performance of newly proposed process control and monitoring methods. The process includes five units, a total of 52 variables, and 21 pre-set faults that can be introduced. The information of the variables and the types of the faults is available in Downs and Vogel's paper (Downs and Vogel, 1993). In this work, a normal data set and 18 faulty data sets are applied. Each data set contains 960 samples and 33 variables because the component variables are excluded for their long sampling time. All the faults are introduced at the 160th sample point.

4.2. Results and discussion

KPSS test is applied to the normal data set, 7 variables are determined as nonstationary variables, which are compressor work, compressor recycle valve, condenser cooling water flow and other four variables in stripper. It is worth noting that stream 4 in TEP is directly delivered to the stripper, and most faults introduced to stream 4 cannot be early detected, which can be considered that the fault signal is buried in process non-stationarity. Independent variables are selected to train SVR models for these 7 variables respectively. Radial basis function is selected as kernel function and model parameter C is 20, gamma is 0.004 according to cross validation. The residues are used to replace original signals and the fault signal can be early identified because the process non-stationarity is removed. Taking fault 8 as an example, original normal data set and fault data set of 4 nonstationary variables are shown in Figure 2. The normal curves for certain variables show a similar nonstationary trend because they are all collected from stripper and highly correlated. The fault signal cannot be distinguished from the normal signal at its early stage because the fault is introduced to nonstationary variables. By removing non-stationarity using virtual variables in Figure 3, the interval of random disturbances under normal conditions is narrowed and the fault signal can be early identified. PCA model is then established for process monitoring, and the monitoring results of all 16 faults are shown in Table 1. It can be obtained that the proposed method shows the best performance in fault detection rate. It is worth noting that the proposed method shows obvious advantages in fault 5, 10, 12, 20, and 21 because these faults are introduced into nonstationary variables. For fault 5, fault detection rate obtained by the comparison methods is low because the response of the control system reduces the fault signal, but the fault signal can still be obviously distinguished by proposed method. For other faults, the proposed method also shows better performance in fault detection time. It is proved that this method provides a feasible idea for monitoring nonstationary processes.

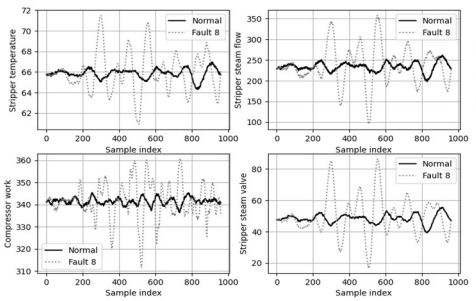


Figure 2 Original data of nonstationary variables

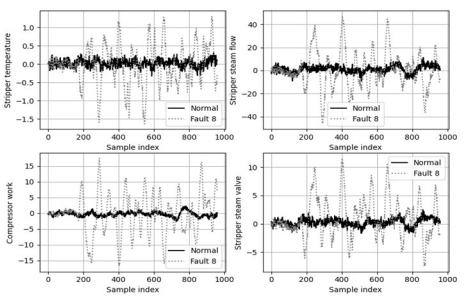


Figure 3 Data of virtual variables in this work

5. Conclusions

In this work, a monitoring strategy for chemical process with nonstationary measurements is proposed. The original signals of nonstationary variables are replaced by virtual variables, which are residues obtained by MI-based SVR models. The process non-stationarity has been removed because the nonstationary and nonlinear feature can be extracted by SVR. Therefore, the interval of normal operating conditions determined from PCA model can be effectively narrowed, by which the faulty deviation can be identified earlier. Case studies on TEP illustrate the effectiveness of this method in nonstationary processes. The proposed method provides a brand-new way to deal with nonstationary measurements in nonstationary process monitoring.

Table 1 Fault detection rates of different methods

Fault	PCA		DPCA		DICA		KSFA		Proposed method	
No.	T^2	SPE	T^2	SPE	T^2	SPE	S^2	SPE	T^2	SPE
1	0.99	1	0.99	1	1	1	0.99	0.99	0.99	**1**
2	0.98	0.99	0.99	0.99	0.99	0.98	0.99	0.99	0.98	**1**
4	0.54	0.96	0.12	1	1	1	0.96	0.99	0.57	**1**
5	0.23	0.25	0.27	0.53	1	1	0.93	0.95	1	**1**
6	0.99	1	0.99	1	1	1	0.99	0.99	0.99	**1**
7	1	1	1	1	1	1	1	1	1	**1**
8	0.98	0.98	0.97	0.97	0.98	0.98	0.99	0.99	0.98	0.98
10	0.33	0.34	0.36	0.57	0.88	0.88	0.78	0.71	0.77	0.84
11	0.21	0.64	0.20	0.83	0.80	0.76	0.87	0.80	0.58	0.82
12	0.97	0.98	0.99	0.97	1	1	0.99	0.99	1	**1**
13	0.94	0.96	0.95	0.96	0.95	0.95	0.95	0.97	0.95	0.96
14	1	0.99	1	1	1	1	1	1	1	**1**
16	0.16	0.25	0.18	0.59	0.89	0.75	0.50	0.70	0.87	0.87
17	0.74	0.89	0.78	0.97	0.97	0.96	0.94	0.96	0.84	**0.97**
18	0.89	0.9	0.90	0.90	0.91	0.91	0.91	0.92	0.91	0.91
19	0.14	0.28	0.20	0.56	0.91	0.9	0.14	0.36	0.78	0.82
20	0.32	0.60	0.36	0.70	0.89	0.81	0.67	0.78	0.61	**0.81**
21	0.26	0.43	0.44	0.60	0.44	0.37	0.536	0.47	0.51	**0.65**
Average	0.648	0.747	0.649	0.841	0.923	0.903	0.841	0.864	0.852	**0.924**

References

J. J. Downs and E. F. Vogel, 1993, A plant-wide industrial process control problem, Computers & chemical engineering, 17, 3, 245-255.

J. Huang, O. K. Ersoy, and X. Yan, 2019, Fault detection in dynamic plant-wide process by multi-block slow feature analysis and support vector data description, ISA transactions, 85, 119-128.

G. Li, S. J. Qin, and T. Yuan, 2014, Non-stationarity and cointegration tests for fault detection of dynamic processes, IFAC Proceedings Volumes, 47,3, 10616-10621.

Z. Li, and X. Yan, 2019, Complex dynamic process monitoring method based on slow feature analysis model of multi-subspace partitioning, ISA transactions, 95, 68-81.

C. Shang, F. Yang, X. Gao, X. Huang, A. K. Suykens, and D. Huang, 2015, Concurrent monitoring of operating condition deviations and process dynamics anomalies with slow feature analysis, AIChE Journal, 61, 11, 3666-3682.

A. J. Smola, and B. Schölkopf ,2004, A tutorial on support vector regression, Statistics and computing, 14, 3, 199-222.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50233-5

AI System for Substance Identification Based on Chemical Substance-Symptom Knowledge Graph

Sangwoo Yoo[a], Hunggi Lee[a], Juri Lim[b], Dongil Shin[a,b*]

[a] *Department of Disaster and Safety, Myongji University, Yongin, Gyeonggido 17058, Korea*
[b] *Department of Chemical Engineering, Myongji University, Yongin, Gyeonggido 17058, Korea*
dongil@mju.ac.kr

Abstract

Current chemical safety management technology is focused on detecting and responding to leaks based on sensors installed in the field, so it is difficult to respond appropriately in case of sensor failure or unexpected accidents. In a laboratory or a site where various chemical substances are irregularly handled, the accuracy of judging exposed substances based on bias is low, especially if prior risk analysis or sensor installation is insufficient. Therefore, loss can be minimized by providing quick and accurate initial response information at the accident site as well as support for accident material identification using the symptoms expressed by the exposed person, which can be viewed as one of the biosensors. Accordingly, this study supported the establishment of a substance-symptom knowledge base and developed a real-time exposure substance identification AI system based on the manifestation symptoms.

This study developed a system that identifies chemicals exposed from symptoms expressed to the exposed person in unexpected chemical leakage accidents and supports response information of determined substances for preemptive response in dangerous situations. In the development, in addition to knowledgeing exposure symptoms and chemical information, a knowledge base (KB) was established in the form of a knowledge graph using AllegroGraph. In order to determine whether knowledge within the established KB is true, knowledge demonstration was conducted through a knowledge graph embedding technique. Based on the established KB, material discrimination was conducted using SPARQL knowledge query inference. In addition, in the case of new substances with insufficient symptom knowledge, it is difficult to support with the KB-based inference system proposed in this study, so a structure-based symptom prediction DNN model using the fingerprint technique (MACCS Keys, Mordred) was proposed, showing high accuracy of 81.82.

Keywords: chemical incident, exposure symptom knowledge, chemical identification, deep learning, artificial intelligence

1. Introduction

In the production, handling, distribution, storage and use of chemical substances, unexpected chemical leakage accidents at the site can cause enormous damage such as personal injury and property damage if the initial response is not appropriate. In the event of a chemical leakage accident, early identification of the leaked material and prompt initial response are essential in minimizing the loss caused by the accident.

Existing chemical detection technology lacks smart technology for detecting and discriminating contact chemicals essential for initial response in the event of an unexpected chemical leakage accident. When an exposure accident occurs, it is difficult to accurately identify the leaked chemical if it is not filtered in the prior risk analysis based only on sensor information. Compared to research on leak response technology and sensor-based abnormal detection and monitoring technology, research on chemical substance diagnosis considering symptoms is very insufficient. If a sensor is not installed at the accident site, an initial estimate must be made based on qualitative exposure symptoms, but the knowledge service on hazardous chemicals considering human bio-sensing information (symptoms) is also relatively insufficiently developed.

This study proposed a knowledge service-based AI system that identifies chemicals exposed from symptoms expressed to the exposed person in unexpected chemical leakage accidents and supports response information of determined substances for preemptive response in dangerous situations.

2. Proposed System: SEARCH

A Symptom-based Expert for Advanced Response (SEARCH) system was proposed and established to provide real-time exposure substance identification support and initial response information based on symptoms of expression when an unexpected chemical accident occurs at the chemical handling site (see Fig. 1). In this paper, we intend to deal only with the knowledge base construction, knowledge verification, exposure substance identification, and knowledge expansion based on predictive models, which are the key issues in the operation of SEARCH.

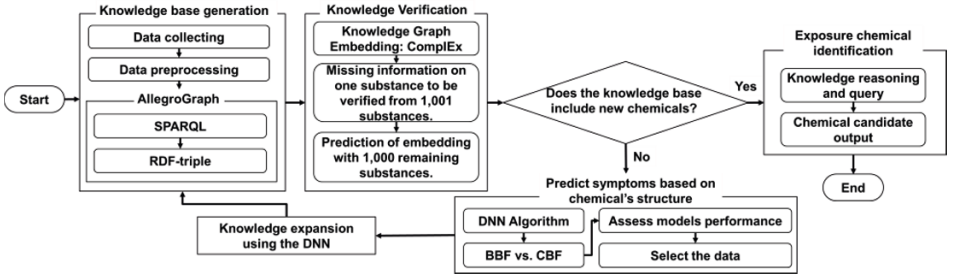


Figure 1. SEARCH system flow chart

3. Chemical-symptom knowledge base generation

A knowledge base was established to support symptom-based exposure substance identification and to verify knowledge using knowledge graph embedding (see Fig. 2). In the case of symptom-based exposure substance identification support, it is not possible to support substances not equipped with a knowledge base, so it is essential to expand knowledge based on predictive models for continuous support. To secure the ease of knowledge expansion, a graph database-based knowledge base that is easy to add/delete/change was built.

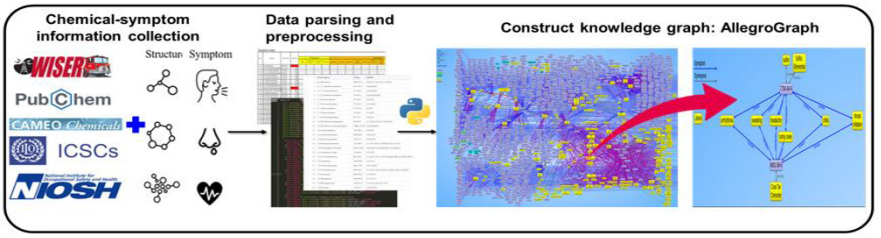


Figure 2. Framework of constructing knowledge graph

3.1. Chemical and symptom data collection

The collected chemicals were collected from 1,001 chemicals, including 499 core substances commonly used in industrial sites mounted on National Institutes of Health (NIH)'s Wireless Information System for Emergency Responders (WISER) and 502 substances provided by PubChem, Camo Chemicals, ICSCs, and NIOSH, which are open chemical databases. Exposure symptom knowledge of chemicals was based on NIH's WISER-mounted information, and symptom knowledge contains 79 symptom information (high body temp, blood noise, etc.) divided into 10 categories such as temperature, nervous system, oral cavity, and skin (Hochstein, C., 2008). In addition, RDKit, a Python package (provided by name, substructure and similarity of chemical substances, SMILES, InChI, etc.), was used to collect additional chemical information.

3.2. Chemical information and molecular structure preprocessing

In order to predict the symptoms of new substances not mounted on the knowledge base, the collected chemical information and molecular structure information were pre-processed using the fingerprint technique. To further confirm which method of bits based fingerprint (BBF) or count based fingerprint (CBF) shows better prediction performance in predicting symptoms, BBF-based MACCS keys and CBF-based Mordred fingerprint methods were used (see Fig. 3).

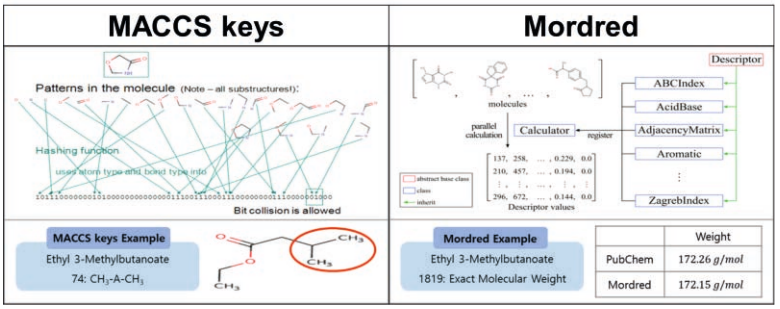


Figure 3. MACCS keys vs. Mordred

MACCS keys convert SMILES into MACCS keys with the most commonly used structural key (Anju Sharma, 2021). Using RDKit, a Python package, it was converted to 166 bits of structural keys of 960 bits and 166 bits. Mordre is a molecular descriptor calculator (Hirotomo Moriwaki, 2018). From various experiments, the chemical information of the molecule is converted into standardized experimental results (real numbers), which are 1825 expressors. RDKit, a Python package, was used for conversion.

3.3. Chemical substance-symptom knowledge base construction

AllegroGraph, a knowledge graph tool, was used to build a knowledge base. AllegroGraph is a triple store system designed to store knowledge in the form of Resource Description Framework (RDF) triple, available through Common Lisp, Python, and other APIs, and supports knowledge graph queries and reasoning such as SPARQL, RDFS+ (Gundla & Chen, 2016). Therefore, AllegroGraph, which can be linked to Python and supports knowledge graph construction, query, and reasoning, was selected as a knowledge graph tool. The previously collected and extracted 1,001 chemical substances, 79 symptom knowledge information, and preprocessed data were saved in Excel and XML format, and converted into RDF triple, one of the knowledge expression methods, to build a knowledge base in AllegrpGraph. For the construction, agraph-python, a Python package provided by AllegroGraph, was used in a Linux environment.

4. Knowledge verification using knowledge graph embedding

Although several DBs have been cross-verified for the expression of chemicals, it is difficult to secure data, and errors may exist because people directly intervene in building knowledge graphs to add knowledge and data, so knowledge graphs were embedded.

4.1. Adaptation of KB embedding algorithm

In this study, the ComplEx model, a knowledge graph embedding model, was used to demonstrate knowledge within the knowledge graph. The formula for the Score Function of the ComplEx model is as follows (T. Trouillon et al., 2016):

$$Score\ Function: RE\left(< v_s, v_p, \overline{v_o} >\right) \tag{1}$$

The data were set to triple 73,692 cases including exposed material-expression symptoms, exposed material-MACCS key, train/test set ratio was 8:2, and parameters were set to k=150, epochs=20, optimizer=adam, batch_count=100, eta=5. For knowledge verification, information on one substance to be verified was omitted from 1,001 substances mounted on the knowledge graph, and then embedding prediction was conducted with the remaining 1,000 substances. In prediction, if the prob value of the missing substance was greater than 0.98, it was accepted as True.

4.2. Knowledge verification results

As a result of knowledge verification through knowledge graph embedding, the results of knowledge verification of substances with less than 10 (Cuminaldehyde), 10 or more and less than 30 (isoamyl acetate), and 30 or more and less than 79 (benzene) among 1,001 substances can be found in Table 1 below. The average FP rate of 1,001 materials, including the three specified materials, was very low at 0.19, confirming the demonstration of triple mounted in the knowledge graph.

Table 1. Result of knowledge verification of Cuminaldehyde, Isoamyl Acetate, Benzene

	FN Rate	**FP Rate**
Cuminaldehyde	0.0	0.0
Isoamyl Acetate	0.1	0.06
Benzene	0.12	0.07

5. Identification of symptom-based exposure substances

Using SPARQL provided by AllegroGraph, symptom-based exposure substance discrimination was performed by knowledge query reasoning. As a case study, case 1 entered five symptom information (Nose_sneezing, chills, chest pain, eye swelling, tinnitus). In addition, if there are unconfirmed symptoms in the knowledge query inference process, the speed and accuracy of inference may be increased by excluding the symptoms, and case 2 is an example: Case 2 entered five symptom information (Nose_sneaking, chills, chest pain, eye swelling, tinnitus), and added skin swelling to the symptom information to enter five expression symptoms and one non-expression symptom, where skin swelling is a non-expression symptom.

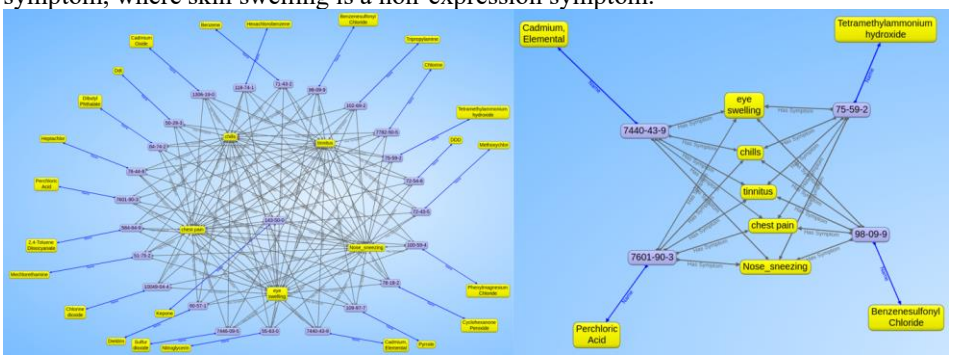


Figure 4. (a) Case 1; (b) Case 2.

As a result of inference, it can be seen that 11 chemicals were inferred in case 1, and 4 chemicals were inferred in case 2. Case 1 showed the same performance as the result of substance estimation only with symptoms among WISER's identity compound function. However, WISER must enter an average of 40 or more symptoms when only symptoms are entered to identify one substance (Bhavnani et al., 2007). Therefore, there is a limit to rapid chemical identification. Case 2 is a symptom exclusion function through unconfirmed symptoms that are not present in WISER. Using SPARQL conditional statements, unconfirmed symptoms were excluded from material estimation, showing that certain substances could be identified with less information input (see Fig. 4).

6. Symptom knowledge expansion using DNN

Since symptom-based substance identification support is not possible in the established knowledge base, a Deep Neural Network (DNN) model was developed to predict symptom knowledge from the molecular structure of the chemical as one of the ways to expand knowledge in the future. In order to compare and analyze the effective fingerprint method in predicting symptoms, it was divided into MACCS keys and mordred. In case 1, MACCS keys are added as input values, and in case 2, mordred is added as input values, and one chemical comes out as output values, resulting in the accuracy of predicting 79 detailed symptoms.

6.1. Adaptation of DL algorithm

In this study, the DNN model was used as a DL method for knowledge expansion.
The data is 482 substances except WISER's solid, 502 substances in PubChem, 79 sub-symptoms, 166 MACCS keys, 1,825 mordred, the train/test set ratio is 8:2, parameters were set as hidden layer=3, activation function=Relu, softmax, optimizer=adam.

6.2. Result of model performance assessment

As a result of the DNN model performance according to Fingerprint, the accuracy of the prediction model using mordred as input was about 10%. The reason Mordred has high performance is that mordred has about 11 times more data than MACCS keys, while MACCS keys only have binary bits of 0 and 1, while mordred contains real information obtained through various experiments.

Table 2. Model accuracy according to two fingerprints

	MACCS keys	**Mordred**
Accuracy	70.24	81.82

7. Conclusions

The starting point of a response to a chemical exposure incident is the identification of a potential exposure substance. To this end, a total of 1,001 chemical-symptom-based knowledge bases were built by adding the symptoms provided by WISER and chemicals provided in the open chemical DB targeting high-risk substances at industrial sites or chemical accident sites. ComplEx, a knowledge graph embedding model, was constructed to demonstrate knowledge in the established knowledge base. Demonstration of knowledge was verified through case studies. The average FP rate of 1,001 materials was 0.19, which was very low. Based on the proven knowledge base, a system for identifying exposure substances that receives symptom information in real time and presents candidates for exposure substances was designed, and its effectiveness was verified through case studies. The case study showed the same results as WISER, the existing standard system, and showed that the speed and accuracy of discrimination can be increased through material discrimination that reflects the conditions that WISER does not have. In addition, since it is difficult to support new substances with little symptom knowledge with the KB-based reasoning system proposed in this study, a structure-based symptom prediction DNN model using the fingerprint method (MACCS Keys, Mordred) was proposed. The DNN model using Mordred as input showed the best performance with an accuracy of 81.82.

References

Anju Sharma, Rajnish Kumar, Shabnam Ranjta and Pritish Kumar Varadwa, 2021, “SMILES to Smell: Decoding the Structure−Odor Relationship of Chemical Compounds Using the Deep Neural Network Approach”, *J. Chem. Inf. Model*, 61, 676-688.

Garrett B. Goh, 2017, “Chemception: A Deep Neural Network with Minimal Che mistry Knowledge Matches the Performance of Expert-developed QSAR/QSPR Models”, *arXiv preprint arXiv:1706.06689*.

Gundla, N. K., & Chen, Z., 2016, “Creating NoSQL biological databases with ontologies for query relaxation”, *Procedia Computer Science*, 91, 460-469.

Hirotomo Moriwaki, Yu-Shi Tian, Norihito Kawashita and Tatsuya Takagi, 2018, “Mordred: a molecular descriptor calculator”, *Journal of Cheminformatics*.

Suresh K Bhavnani, 2007, “Network Analysis of Toxic Chemicals and Symptoms: Implications for Designing First-Responder Systems”, *AMIA Annual Symposium Proceedings*, 51-55.

T. Trouillon, J. Welbl, S. Riedel, E. Gaussier and G. Bouchard, 2016, “Complex embeddings for simple link prediction”, *Proc. 29th Int. Conf. Mach. Learn*., pp. 2071-2080.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50234-7

Model-based monitoring of an intensified unit for continuous pharmaceutical filtration-drying

Francesco Destro[a], Massimiliano Barolo[a], Zoltan K. Nagy[b,*]

[a]*CAPE-Lab, University of Padova, 35131 Padova, PD, Italy*
[b]*Purdue University, West Lafayette, IN 47906, USA*
znagy@purdue.edu

Abstract

Active pharmaceutical ingredient (API) separation from synthesis and crystallization mother liquors is typically carried out in pharmaceutical manufacturing through filtration and drying. These steps are of utmost importance, as impurities herein retained will inevitably end up in the drug product. Recently, a novel carousel has been developed for carrying out filtration and drying in a continuous intensified fashion. The unit represents a step forward with respect to traditional batch filtration and drying, as continuous operation can reduce the variability of the product quality. However, the occurrence of faults compromising product compliance can be assessed only upon discharge of the final cake of API crystals, when its purity can be measured. In this work, we develop a model-based monitoring system for the unit, based on state and parameter estimation. The implemented monitoring system succeeds in tracking the product critical quality attributes (CQAs), and in detecting common faults for the carousel, such as sudden variations of the feed attributes.

Keywords: Process monitoring, State estimation, Quality-by-Design, Continuous pharmaceutical manufacturing, Fluid-solid separation

1. Introduction

In response to the recently registered alarming numbers of drug shortages and recalls, pharmaceutical regulators are encouraging the transition to the so-called “six sigma quality” (Yu and Kopcha, 2017). Under a six sigma quality system, there are six standard deviations between the process mean of each quality variable and the relevant nearest specification limit. Albeit six sigma quality is established in many manufacturing industries (e.g., the semiconductors one), the pharmaceutical industry is still lagging behind, at a two-three sigma quality standard. Among the pharmaceutical emerging technology needed to achieve six sigma quality, a pivotal role is assumed by the transition to continuous processing, and by the implementation of an advanced monitoring system. Such monitoring system should include process analytical technology or soft sensors for directly monitoring the product critical quality attributes (CQAs), and suitable routines for automatic detection of faults that might compromise the product quality.

In this study, we present a monitoring system for a novel continuous carousel for intensified filtration-drying of crystallization slurries (Destro et al., 2021). The carousel is an enabler for end-to-end continuous pharmaceutical processing, as it is one of the few technologies on the market that can connect the upstream and downstream sections of pharmaceutical processes in a continuous fashion. The unit features multiple processing stations embedded in a main cylindrical body, which rotates at every given time interval (cycle duration), transferring the material from one station to the subsequent one. In the

first station, the crystallization slurry is loaded, and the first processing cycle starts. At the end of every cycle, the material is transferred to the following station, until it reaches the final one, and it is discharged in the form of a dry crystals cake. The rotation mechanism enables continuous operation, as different batches of slurry are processed simultaneously in the stations. The CQA of the process is the purity of the discharged crystals cake, which must respect tight solvent and impurity content constraints. However, the composition of the cake being dried cannot be directly measured through sensors during carousel operation. At the same time, other parameters strictly related to the occurrence of disturbances that might compromise the product quality (e.g., abnormal filter mesh fouling or critical change of physical properties of the cake) cannot be measured in real time. Hence, in this study we propose a monitoring system comprising a state estimator (Ray, 1981) for soft-sensing the CQAs of the cakes being processed in the carousel. The monitoring system also features real time parameter estimation routines for tracking critical parameters that can indicate the occurrence of faults. Univariate charts are used on process measurements and estimated states and parameters for fault detection purposes. The monitoring system is tested on a simulated process (Destro et al., 2021) for the isolation of paracetamol crystals from a paracetamol/ethanol slurry with the carousel technology object of this study. Section 2 introduces the process and the case study. Section 3 outlines the conceived monitoring system and presents the monitoring results, before the concluding section.

2. Continuous filtration-drying of paracetamol/ethanol slurries

The carousel setup (Figure 1, equipment legend in Tables 1-2) features five cylindrical ports. In the first one (V101), a volume of paracetamol/ethanol slurry V_{slurry} (controlled by FQC-101) is loaded from a slurry tank (V106), and filtration starts under the action of the pressure P_{compr}, delivered by compressor P101 (controlled by PC-101). Simultaneously, cake deliquoring occurs in V103-V104, and cake thermal drying occurs in V105. Thermal drying is promoted by a flow of hot air, whose temperature when entering V105, $T_{in,g}$, is controlled by TC-101. After a fixed cycle duration Δt_{cycle}, controller KC-101 triggers a cycle rotation. All the material being processed in V101-V104 moves to the following station. The cake entering V105 is ejected through a piston movement, and new slurry is loaded into V101. Different disturbances affect the process in general operating conditions, such as stochastic variabilities of slurry concentration, cake porosity, cake specific resistance and drying kinetic constant, introduced to generate inter-batch variability. Filter mesh fouling also occurs during carousel operation, and must be considered when designing the control strategy of the unit. After every cycle, the meshes installed below ports V101-V104 become more and more fouled, thus increasing the filtration and drying durations. When filter mesh fouling becomes excessive (i.e., every six processing cycles in the simulator), a meshes cleaning-in-place (CIP) procedure is automatically triggered. The material already present in the carousel is processed regularly, but no more slurry is loaded into the carousel. When all the material present at the CIP triggering is discharged from the carousel, ethanol is flowed into all filter meshes, restoring clean filter mesh conditions. A new carousel processing cycle is then initiated by loading new slurry in V101. Note that in the first cycle after the process onset, and after every CIP, only V101 contains material: the remaining ports start operating gradually, cycle after cycle. The simulator is initiated by setting the following values for the controller set-points: $T_{in,g}^{sp}$=50°C (TC-101), V_{slurry}^{sp}= 6 mL (FQC-101),

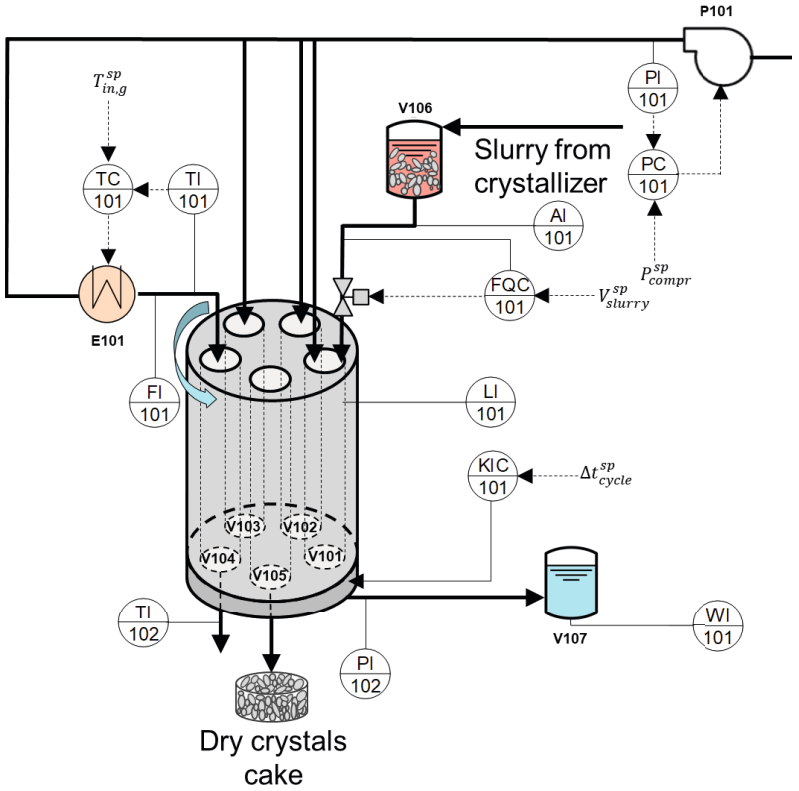


Figure 1. Schematic diagram of the carousel process. V101-V104 present a filter mesh at the bottom, while V105 is open for cake discharge. The equipment legend is given in Table 1 (unit operations and controllers) and in Table 2 (sensors).

Table 1. Legend of the diagram of Figure 1: unit operations and controllers.

Name	Description	Name	Description
P101	Compressor	VI106	Slurry tank
E101	Drying air heater	VI107	Filtrate collector
VI101	Carousel Station 1		
VI102	Carousel Station 2	FQC101	Fed slurry volume controller
VI103	Carousel Station 3	KIC101	Carousel rotation controller
VI104	Carousel Station 4	PC101	Pressure controller
VI105	Carousel Station 5	TC101	Drying air inlet temperature controller

and P_{compr}^{sp}=1 bar_g (PC-101). The cycle duration set-point Δt_{cycle}^{sp} (KIC-101) is instead controlled at closed-loop, as outlined in Section 3. The remaining specification needed for running the simulated process is the nominal slurry concentration, which is fixed at 250 kg/m^3.

Two datasets are generated, both of 1 h duration: Dataset 0, corresponding to normal operating conditions, and Dataset 1, where a sudden 100% increase of the nominal specific cake resistance occurs in the cakes formed by the slurry loaded into the carousel, starting 5 min after the process onset. Table 2 lists the set of measurements generated for each dataset. In the remained of the paper, we use the symbol y to denote a measurement.

Table 2. Legend of the diagram of Figure 1: sensors and corresponding measurements.

Name	Measured variable	Symbol	Unit	Notes
AI-101	Slurry concentration	$y_{c_{slurry}}$	kg/m^3	Ultrasonic probe
FI-101	Air flowrate	$y_{\dot{V}_{in,g}}$	NL/min	
LI-101	Cake height	$y_{H_{cake}}$	m	Camera system
	Fed slurry volume	$y_{V_{slurry}}$	m^3	Camera system
PI-102	Pressure	y_P	Pa	
TI-101	Drying air inlet temperature	$y_{T_{in}}$	K	
TI-102	Drying air inlet temperature	$y_{T_{out}}$	K	
WI-101	Filtrate mass	$y_{M_{filt}}$	kg	

3. Monitoring system: implementation and proof of concept

We propose a monitoring system for the carousel of Figure 1 that features: *i)* real-time parameter estimation, *ii)* state estimation, and *iii)* univariate charts for fault detection and diagnosis, with control limits. The framework for real-time parameter and state estimation (Figure 2) features three routines.

At the end of every cycle that involves slurry processing in V101, a set of filtration-relevant parameters is estimated: the resistance of the filter mesh installed in V101 ($\hat{R}_m$), the specific cake resistance ($\hat{\alpha}$), the cake porosity ($\hat{\epsilon}$), and the filtration duration ($\Delta\hat{t}_{filt}$). The cake porosity is directly estimated through a mass balance:

$$\hat{\epsilon} = 1 - \frac{y_{V_{slurry}} y_{c_{slurry}}}{\rho_s \, y_{H_{cake}} \, A_{dryer}}, \tag{1}$$

where ρ_s is the crystals density, A_{dryer} is the filter cross-section, and the measurements legend is in Table 2. Then, $\hat{R}_m$ and $\hat{\alpha}$ are obtained through maximum likelihood estimation using a filtration model (Destro et al., 2021), P^{sp}_{compr}, $\hat{\epsilon}$, and measurements $y_{c_{slurry}}, y_{H_{cake}}$, $y_{V_{slurry}}$, and $y_{M_{filt}}$ (Table 2). Based on the estimated $\hat{R}_m$ and $\hat{\alpha}$, the

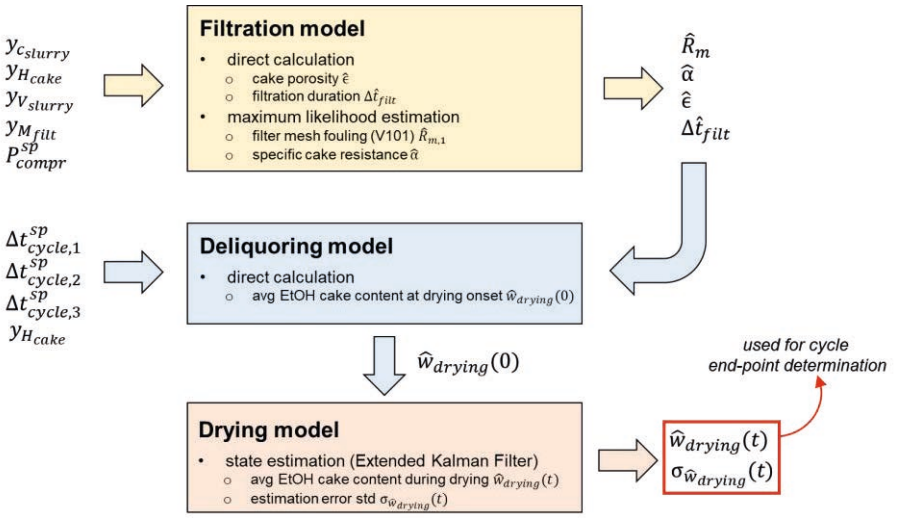


Figure 2. Framework for real-time parameter and state estimation.

filtration model is exploited again, this time for estimating the filtration duration $\Delta\hat{t}_{filt}$ of the cake that has just been processed in V101.

Before the beginning of every carousel cycle that involves cake drying, the duration $\Delta\hat{t}_{deliq}$ of the deliquoring step undergone by the cake that is about to be dried is estimated, from the previously obtained $\Delta\hat{t}_{filt}$, as:

$$\Delta\hat{t}_{deliq} = \Delta t^{sp}_{cycle,1} + \Delta t^{sp}_{cycle,2} + \Delta t^{sp}_{cycle,3} - \Delta\hat{t}_{filt}\ , \tag{2}$$

where $\Delta t^{sp}_{cycle,i}$ (for i = 1, 2, 3) is the duration of the cycles during which the considered cake was, respectively, in V101, V102, and V103. From $y_{H_{cake}}$, $\hat{R}_m$, $\hat{\alpha}$, $\Delta\hat{t}_{deliq}$, and $\hat{\epsilon}$, the ethanol content of the cake entering dryer V104 ($\hat{w}_{drying}(0)$) is calculated through a deliquoring model (Destro et al., 2021).

Based on $\hat{w}_{drying}(0)$, a state estimator (an extended Kalman filter; Ray, 1981), rooted on a drying model developed for the carousel (Destro et al., 2021), is then initiated. The state estimator is used, during the following cycle, for real-time tracking of the time evolution of $\hat{w}_{drying}$ of the cake being dried in V104 (Figure 3). The state estimator also provides $\sigma_{\hat{w}_{drying}}$, the standard deviation of the estimation error of $\hat{w}_{drying}$. In addition to soft-sensing $\hat{w}_{drying}$, the CQA of the process, the state estimator is used for determining Δt^{sp}_{cycle}. When the upper confidence limit of the estimated ethanol content in the cake being dried in V104 reaches the quality threshold, a carousel rotation is triggered (Figure 3). This automated setting for cycle duration determination allows consistently obtaining cakes meeting the target quality specification, because during carousel operation Δt^{sp}_{cycle} is automatically adjusted based on the needed drying duration (Figure 4a). Note that, for processing cycles in which V104 is empty, Δt^{sp}_{cycle} is kept fixed to 30 s. In Dataset 1 (Figure 4a), a sudden increase of the average cycle duration is registered after the 13-th cycle. This cycle actually corresponds to the situation when, for Dataset 1, the first cake with an abnormally large specific resistance (cake #7) enters V104. Since the parameter estimation routine effectively tracks the specific cake resistance increase in the cakes entering V101 (Figure 4b), the state estimator allows to promptly detect the

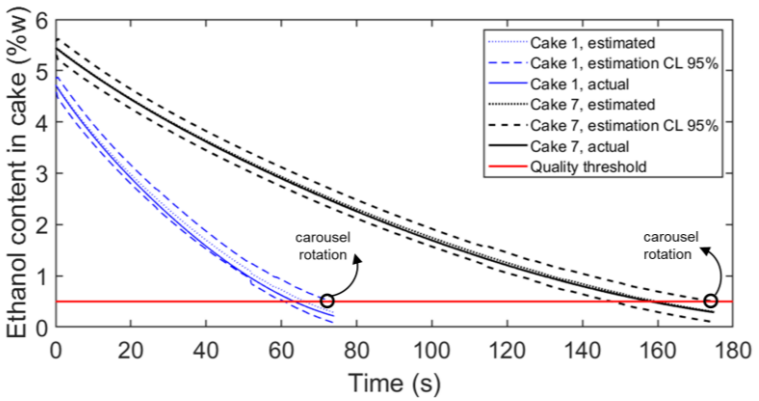


Figure 3. Dataset 1: ethanol content during drying for the first and the seventh cakes processed in the carousel: estimated value, estimation error 95% confidence limits (CL) and actual value.

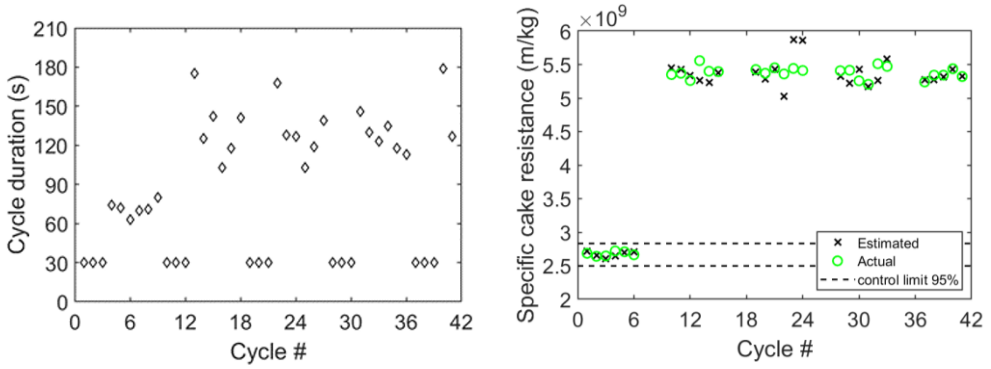


Figure 4 Dataset 1: *(a)* duration of the different processing cycles, *(b)* univariate monitoring chart for specific cake resistance of the cake in V102, reporting estimated value, actual value, and 95% control limits for fault detection, obtained from the normal operating conditions dataset (Dataset 0).

larger drying duration needed by cake #7, compared to cake #1 (Figure 3). Figure 4b also reports the 95% control limits, obtained from Dataset 0, for the estimated specific resistance of the cakes in V101. Abnormal specific cake resistance is immediately detected after cycle #10 (corresponding to when cake #7 enters V102).

Conclusions

We presented a monitoring framework for intensified filtration-drying of a paracetamol/ethanol slurry through a novel continuous carousel. The monitoring system features parameter and state estimation routines for monitoring the product CQAs and key operating parameters. The estimated CQAs are exploited for automatically determining the drying end-point. The monitoring system has successfully been tested under a set of disturbances on a simulated process. Multivariate process monitoring implementation for fault detection and diagnosis is envisioned for the future work.

Acknowledgements

Funding for this publication was made possible, in part, by the Food and Drug Administration through grant (U01FD006738). Views expressed in written materials or publications and by speakers and moderators do not necessarily reflect the official policies of the Department of Health and Human Services; nor does any mention of trade names, commercial practices, or organization imply endorsement by the United States Government. F.D. gratefully acknowledges the “CARIPARO Foundation” for his PhD scholarship and “Fondazione Ing. Aldo Gini” for the financial support.

References

Destro, F., Hur, I., Wang, V., Abdi, M., Feng, X., Wood, E., Coleman, S., Firth, P., Barton, A., Barolo, M., Nagy, Z. 2021. Mathematical modeling and digital design of an intensified filtration-washing-drying unit for pharmaceutical continuous manufacturing. Chem. Eng. Sci., 244, 116803.

Ray, W.H., 1981. Advanced process control. McGraw-Hill, New York.

Yu, L.X., Kopcha, M., 2017. The future of pharmaceutical quality and the path to get there. Int. J. Pharm. 528, 354–359

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50235-9

Plant Fault Diagnosis System using Negative Selection Algorithm

Naoki Kimura*, Yuki Ichikawa, Kazunori Tanihara, Yuichi Makiya, Gen Inoue, Yoshifumi Tsuge

Dept. of Chemical Engineering, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka, 819-0395 Japan
nkimura@chem-eng.kyushu-u.ac.jp

Abstract

Early fault detection and correct diagnosis are required for chemical plants. Therefore, the existing fault detection systems using upper/lower thresholds have difficulties to detect faults when the correlation among process variables breaks without excess of any thresholds. In our previous study, an artificial immune system—especially, negative selection algorithm— had been adopted to fault detection system. Negative selection algorithm is one of methods of artificial immune systems which imitate the vital immune system. We have built up a multiagent based fault detection system using negative selection algorithm. In our system, a set of detectors is generated in each two-dimensional variable space consists of two process variables.

In this study, we extend the system to plant fault diagnosis based on the fault detection result using negative selection algorithm. In this paper, we will illustrate our fault detection and diagnosis system using negative selection algorithm. And we will show the detection and diagnosis results when a malfunction occurs in a dynamic plant simulator of a boiler plant.

Keywords: Fault detection; Fault diagnosis; Artificial immune system; Negative selection algorithm.

1. Introduction

Chemical plants are very complicated because they composed of lots of equipment and instruments. It is an essential to keep safe and stable state in operation of chemical plant, and various countermeasures are taken. A fault detection mechanism of an existing plant protection system generally sets a threshold value for each measuring instrument, and when it exceeds the threshold value, it issues a signal and becomes an alarm. As a method of setting the threshold value, there are methods such as upper limit / lower limit, upper limit difference / lower limit difference, upper limit of change rate / lower limit of change rate. However, we thought that an abnormal condition that could not be detected by only the technique that defines these thresholds exists, and wanted to propose a new fault detection method that can detect the features of anomalies with different properties. Therefore, in our previous study (Kimura 2018), we have proposed an artificial immune system as a new fault detection method capable of detecting features of abnormalities with different properties. An artificial immune system is a system focused on the mechanism of immunity of living organisms and has been proposed in the early 1990's (Dasgupta 2006 & 2011). In the artificial immune systems, negative selection algorithms designed to imitate the mechanism of differentiation and

maturation of immature T cells and the mechanism of discrimination of normality / abnormality of immune are systems specialized for fault detection of researchers attracting attention.

In this research, we aim to develop a fault diagnosis system as a extension of our previously investigated fault detection system using negative selection algorithm.

2. Methodology

2.1. Fault Detection using Negative Selection Algorithm

Negative selection algorithm is one of the fault detection methods in artificial immune systems. Artificial immune system is a mechanism to maintain the system normal, borrowed the idea from the vital immune systems. In this study, the normal operational data of the plant is regarded as 'self', and the abnormal operational data, that is 'fault', is regarded as 'non-self'. Fig.1(a)–(e) illustrates the outline of this algorithm. In our method, there are two phases—detector generation phase and detection phase.In the detector generation phase, there is a variable space consisted of plural process variables which are normalized to [0,1] range. Fig.1(a) shows two-dimensional variable space with two process variables P1 and T1. And the figure also shows there is a path which represents a normal operational data—regarded as 'self' region in this method. The remaining area is regarded as 'non-self' region.Fig.1(b) shows the way to generate detectors. The figure shows four points which are the centers of the detector candidates are generated in the variable space. If the detector candidate has a high affinity with

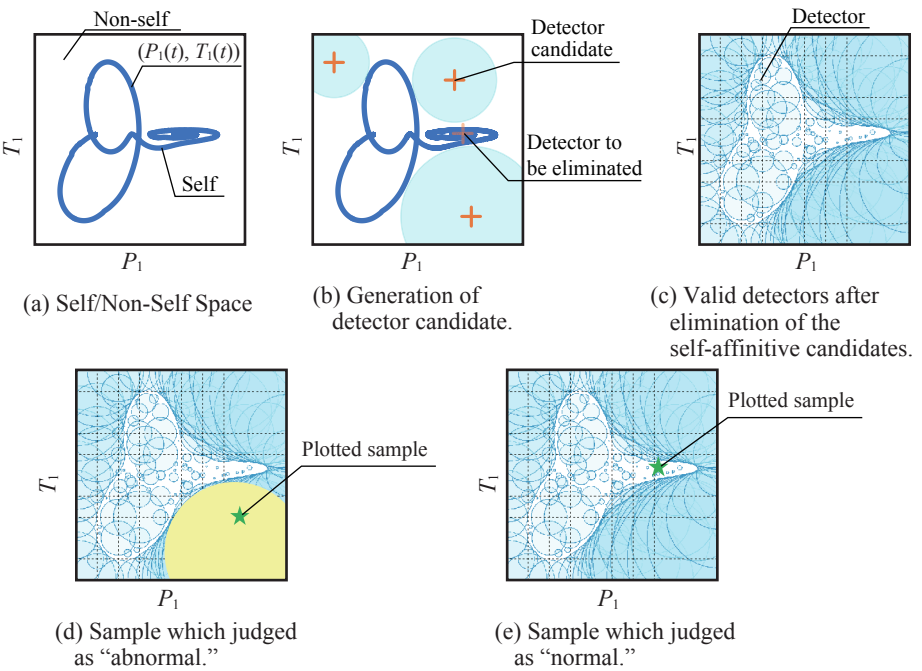


Figure 1: A schematic diagram of the detector generation and detection using negative selection algorithm

the self region, it will be eliminated. One of the four detector candidates in Fig.1(b) is eliminated. On the other hand, if the detector candidate has no affinity with the self region, it will be set the radius which is the shortest distance between self region and the center point and adopted as a detector. This mechanism is called negative selection. To cover the non-self region with detectors efficiently, detector candidates are firstly generated on lattice points in the process variable space, and then detector candidates are generated randomly thereafter. Fig.1(c) shows the detectors fully cover the non-self region of the variable space.

In the detection phase, the data to be subjected to abnormality detection is normalized and plotted in the process variable space. If the plot is placed inside the detector(s) or outside the range of [0, 1], the data is judged as abnormal. In Fig.1(d), a star represents the plotted data, and the data was plotted inside detectors. On the other hand, if the plot is placed without capture by any detectors the data is judge as normal, shown in Fig.1 (e).

2.2. Fault Diagnosis Method

In preparation for fault diagnosis, consideration of the impact on the process when an assumed malfunction occurs. Based on the PFD, P&ID and HAZOP study, it is considered if an assumed malfunction occurs whether the measured value of the sensor increases (denotes as "+"), decreases (denotes as "-") or no effect (denotes as "0") compared to the normal state. After consideration about all the assumed malfunctions, a sign and malfunction table is obtained.

If n variables are used, 3n kinds of sign (+/-/0) combinations can be theoretically made, and if different sign combinations are shown for each cause of malfunction, it should be used to identify the cause of malfunction. However, a variable set in which all *n* signs

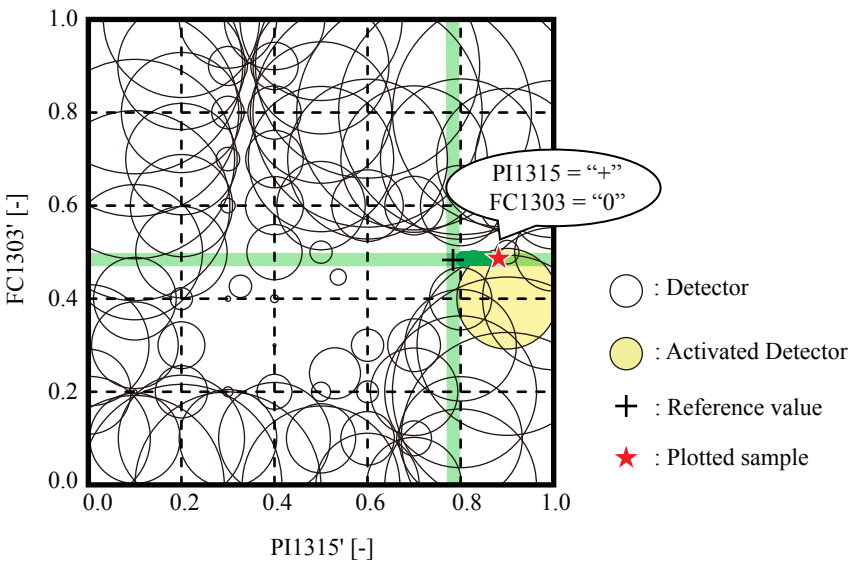


Figure 2: Current operational data plot and reference value in variable space covered by the detectors.

are 0 for a certain malfunction cause or a variable set in which different malfunction causes show the same sign combination cannot be used for fault diagnosis.

The current operation data is plotted in the variable space covered by the detectors every sampling time (Fig.2). If it was plotted in the detection area of the detector, the fault was detected. Then, the sign is judged in three stages: higher (+), no change (0), and lower (-) than the reference value of the operational condition. The sign is determined and summarized for all variable spaces in which a fault is detected, and the diagnostic sign pattern of the monitoring variable set is determined to compare with pattern table.

And then, matching rate for each assumed malfunction is calculated.

$$\text{Matching rate } m = n_{match} / n \tag{1}$$

where, n_{match} is the number of signs that match the sign of the assumed malfunction in the assumed sign pattern and the diagnostic sign pattern, and n is the number of monitoring variables. The malfunction cause(s) are presented to the operator as plausible malfunction cause candidate(s) if the rate(s) is larger than 50%.

3. Simulations Results

Table 1 shows the assumed malfunctions of boiler plant. And the table 1 also shows the diagnostic sign pattern for each malfunction.

Since there are five assumed causes, it can be identified by variables with $n = 2$ or more, but for the reason mentioned in section 2.2, there is no variable set that can be identified with $n = 2$, even though there are more than one hundred sensors in the boiler plant. However, only the combination of { PI1311, PI1315, FC1303 } became an identifiable combination with $n = 3$. Then, these sensor variables are used as a monitoring variable set.

The normal operation data and abnormal operational data was acquired by a dynamic plant simulator "VisualModeler". As for the normal operation data, a total of 12 conditions were acquired when the steam demand was stepwise changed. The abnormal operation data was acquired for a total of 30 conditions: (5 malfunctions; total steam demands: 120, 130 or 140 t/h; 2 types of fluctuation of steam demand).

Table 1: Assumed plant malfunctions

	Assumed malfunction	PI1311	PI1315	FC1303
Mal-1	Heavy oil flowmeter failure	+	+	0
Mal-2	Boiler feed water flowmeter failure	0	0	+
Mal-3	Air flowmeter failure	-	-	0
Mal-4	Boiler water pipe leakage	+	+	+
Mal-5	Burner flame-out	+	0	0

Figure 3 shows the matching rates graph when the malfunction Mal-4 was caused. In Fig.3, the matching rate of correct malfunction cause (Mal-4) shows 100% from detection time to end and the rate for other cause (Mal-1) shows 66.7%. It means that both of the causes are possible cause, because the rates are over 50%. However Mal-1's rate is always maintained at the highest level, it is judged as Mal-1 is plausible cause.

Table 2 shows the diagnosis results for 30 conditions where "Correct" denotes correct diagnosis, "Multiple" denotes that matching rates of both correct and incorrect cause(s) are the same and highest, "False" denotes that matching rate of incorrect cause is the highest, "Missed" denotes that no fault was detected. In this study, two "Multiple", one "Missed" and 27 "Correct" case was obtained.

4. Conclusions

Failure diagnosis has become possible by introducing a sign determination method into the fault detection system that uses a negative selection algorithm. It is necessary to

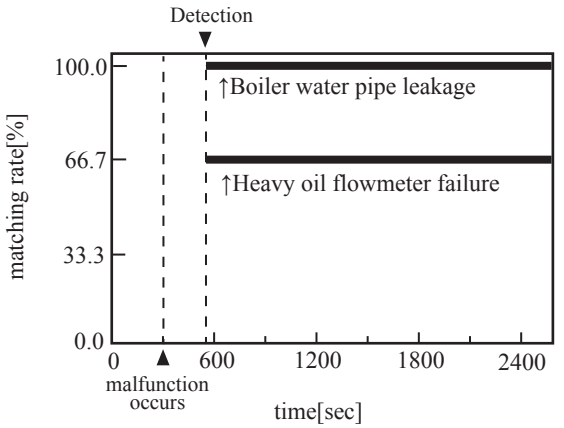


Figure 3: Matching rates after malfunction (Boiler feed water pipe leakage) was occurred.

Table 2: Diagnosis results

	W/O change of steam demand			With change of steam demand		
	Total steam demand			Total steam demad before change		
	140 t/h	130 t/h	120 t/h	140 t/h	130 t/h	120 t/h
Mal-1	Correct	Multiple	Correct	Correct	Multiple	Correct
Mal-2	Correct	Correct	Missed	Correct	Correct	Correct
Mal-3	Correct	Correct	Correct	Correct	Correct	Correct
Mal-4	Correct	Correct	Correct	Correct	Correct	Correct
Mal-5	Correct	Correct	Correct	Correct	Correct	Correct

improve the performance of the fault diagnosis.

References

D. Dasgupta, Advances in artificial immune systems, Computational intelligence magazine, 1, 40–49, 2006.

D. Dasgupta, S. Yu, F. Nino, Recent Advances in Artificial Immune Systems: Models and Applications, Applied Soft Computing, 11, 1574–1587, 2011.

N. Kimura, Y. Takeda, Y. Tsuge, Agent Based Fault Detection System for Chemical Processes using Negative Selection Algorithm, Advances in Science, Technology and Engineering Systems Journal, 3, 2, 90–98, 2018.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50236-0

Digital Twin of a pilot-scale bio-production setup

Monica Muldbak[a], Carina Gargalo[a], Ulrich Krühne[a], Isuru Udugama[a,b], Krist V. Gernaey[a*]

[a]*Process and Systems Engineering Center (PROSYS), The Department of Chemical and Biochemical Engineering, Technical University of Denmark, Kgs. Lyngby 2800, Denmark*
[b]*The Department of Chemical Systems Engineering, The University of Tokyo, Bunkyo-ku 113-8654, Japan*
[*]*kvg@kt.dtu.dk*

Abstract

The term digital twin is now everywhere in the domain of bio-manufacturing. Despite this, the actual implementation of "Digital Twins" in operations is somewhat limited, and the core elements contained within such an implementation are ambiguous. A vital aspect of a fully developed digital twin is the two-way communication capability between the physical plant and the digital model. This contribution will describe the data management and modelling elements needed to achieve a full-fledged digital twin, emphasising the digital model that acts as the prediction engine of a digital twin. To this end, the overall operational workflow of developing a digital replica of a bio-based pilot process will be described. It is expected that the knowledge gained from this endeavour will enable the development of a general framework towards digital model building. Thus, it is expected to generate practical know-how for future digitalisation applications in bio-production.

Keywords: digital twin, digitalisation, modelling, implementation, bio-manufacturing

1. Introduction

Digital Twin (DT) is one of the new buzzwords within the engineering field. A simple literature search shows that the number of scientific publications containing this term has increased significantly over the last couple of years (Udugama, Lopez, et al., 2021). The term DT was first introduced by (Grieves, 2015) where it was defined as a "digital representation of a physical object". Since then, the term DT has been liberally used to describe digitalisation efforts in manufacturing and beyond (Lukowski et al., 2018). In the domain of bio-manufacturing operations, DTs are also gaining popularity. Despite this, a survey conducted with individuals involved with biotech manufacturing in Denmark found that the actual number of implementations that can be considered a DT in plant operations is limited (Udugama, Öner, et al., 2021).

The core of the DT is the digital model(s) being used to simulate the process. Still, the optimal modelling approach and required auxiliary elements to distinguish a digital model from a DT are yet to be defined with the necessary level of detail. At the beginning of 2021, (Udugama, Lopez, et al., 2021) published a five-step pathway to develop a DT for the bio-manufacturing industry. Furthermore, (von Stosch et al., 2021) published an opinion paper about the potential for integrating the DT with artificial intelligence.

In this work, we detail the development of an overall implementation framework for a pilot-scale *E. coli* based production process operated for engineering education at the Bio-manufacturing, Training and Education Center at NC State University (BTEC).

This process has been selected because it is a well-known fermentation process with great relevance in education (e.g., at the BTEC facility) and in industrial settings.. The development of a DT for this process is expected to be used as a demonstration case. The lessons learnt can be used to further the application of DTs in the bio-manufacturing industry. The remainder of this article is organised as follows: Section 2 will describe the benefits and drawbacks of choosing between mechanistic, data-driven, and hybrid options for digital model development. Section 3 gives a detailed description of the implementation framework. Section 4 summarizes the initial lessons learnt and identifies potential opportunities. Finally, conclusions are derived in Section 5.

2. Digital Models

Digital models form the core of any DT application. However, models can quickly become computationally heavy and would thus struggle with the requirement of real-time data handling of a fully realised DT. Therefore, in order to be implemented for real-time simulation, the selection of the modelling approach and the corresponding level of complexity is vital when setting up a digital model.

2.1. Data-driven vs. mechanistic models

Prior to focusing on model complexity, it is essential to select the correct strategy and modelling approach. Mathematical models can generally be divided into three main categories; mechanistic, data-driven and hybrid models. Mechanistic modelling, or parametric modelling, is theory-based modelling. This type of model often contains various parameters and constants, the value of which may be appropriately adjusted to almost any system.

Mechanistic models are widely used in modelling fermentation processes such as *E. coli* based production processes. One such example is (Anane et al., 2017), where a mechanistic approach was used to model overflow metabolism in *E. coli* at a macro-scale. This model was then validated using measurable concentrations within a fed-batch. (Millard et al., 2017) also modelled the growth of *E. coli* but at a smaller scale, which added details regarding metabolite regulation within the cells. However, the number of parameters increases drastically with the complexity of the model, adding a significant computational burden. It can be challenging for complex models to determine the value of these parameters and validate the model. Therefore, adopting a mechanistic model is not trivial and requires expert knowledge.

Data-driven models, on the other hand, tackle the system from a completely different point of view. They are, as the name implies, based on system data. This means that data-driven models can be susceptible to variance within data sets unless large amounts of data are available. However, the collection of large datasets is resource-intensive. Notwithstanding, it can be an advantage if, for model improvements, we can apply a machine learning algorithm (e.g., artificial neural networks) since these types of models usually improve with larger data sets. The advantage of the data-driven models is that they usually require less experience and expert knowledge to set up and can be easier to validate than some mechanistic models.

A disadvantage of data-driven models is related to the fact that they are limited by the type and sensitivity of the sensors available. A sensor setup composed of different sensors can be cumbersome and expensive to set up. Adding additional sensors to an existing system is not without trouble, especially when considering biopharma, where stringent documentation is, and it will always be, a requirement. Flow following sensors form a newer alternative to make use of a more comprehensive sensor setup in a bioreactor. Since

they do not require physical installation to the bioreactor and can acquire data at different points in the tank, flow following sensors seem to be a promising and flexible tool.
Recent examples of flow follower studes include (Lauterbach et al., 2019), who has been working on the development of a truly miniaturised flow follower, which can measure temperature in lab-scale equipment. (Reinecke & Hampel, 2018) have developed a sensor for aerobic digesters. (Bisgaard et al., 2020) presented results on mixing studies using one of the only commercially available multi-parametric flow followers. So far, it has been demonstrated that it provides less extensive and expensive setup modifications, while still providing impactful data collection (Bisgaard et al., 2020).

2.2. Hybrid models

Hybrid models is a term used for any model, which combines the mechanistic and data-driven approaches. Many of the newer hybrid modelling approaches integrate mechanistic modelling and machine learning. There are many different ways this can be done. (von Stosch et al., 2016) have developed a strategy that couples mechanistic modelling with neural networks. Hybrid approaches are also an option when working with models integrating the effect of scale, from macromixing to cellular products. An example of such a multiscale model is the fermentation model developed by (Benalcázar et al., 2020). The issue with neural networks and other traditional machine learning algorithms is that they are discrete by nature. A potential solution to this was proposed by (Chen et al., 2018), who developed a continuous machine learning algorithm based on ordinary differential equations.

3. Proposed Framework

Bio-manufacturing processes are often not fully automated, which is a requirement for implementation of a DT where it can communicate bi-directionally with the plant. To this end, there is a need to develop concepts that can close this gap and provide the operator with actionable predictive information from the DT. The overall framework illustrated in Figure 1 suggests a potential roadmap towards realising a fully-fledged DT for bio-manufacturing processes while respecting the inherent limitations that it implies. Due to the versatile implementation and automation of bio-manufacturing processes, the framework for a DT also needs to be adaptable as well as agile. The strategy to meet this requirement is to develop a ‘plug-and-play’ approach. Individual unit operation models are embedded within the larger framework, allowing the framework to be configured to the DT’s desired purpose and for relatively easy adaptation to other processes.

At the core of the framework is a digital model capable of bi-directional communication. Based on section 2, hybrid models are recommended as the default choice for building this crucial step of the framework. Regarding inputs into the DT, the framework identifies that information from the physical plant can be gathered in real-time as long as the relevant sensors are installed. Since bio-manufacturing processes are mainly operated as batch production processes, the information collected must contain both process data (e.g temperature, pH, flow) and scheduling information (batch sequence, runtime). The data gathered must also be stored in an accessible format so that it can be used for further model validation and improvements.

Connecting the output of the DT to the model is much more complicated. This is because many critical process operations are either carried out manually by the plant operator or require the operator to take action on a subsystem that is not configured to take external commands. Regulatory requirements and the costs of automation mean that these types

of operational configurations will likely remain in bio-manufacturing processes for the foreseeable future. To this end, the action suggested by the DT must be taken by the plant operators. An effective way to communicate these actions is through the use of targeted data visualisatiuon that can provide the operator with actionable information. However, unlike an automated system, operators are susceptible to making mistakes and have individual behavioural traits. While fully describing the nature of a particular human operator is beyond the scope of this work, the framework has identified the need to develop simplified human behaviour models that can be incorporated into the DT.

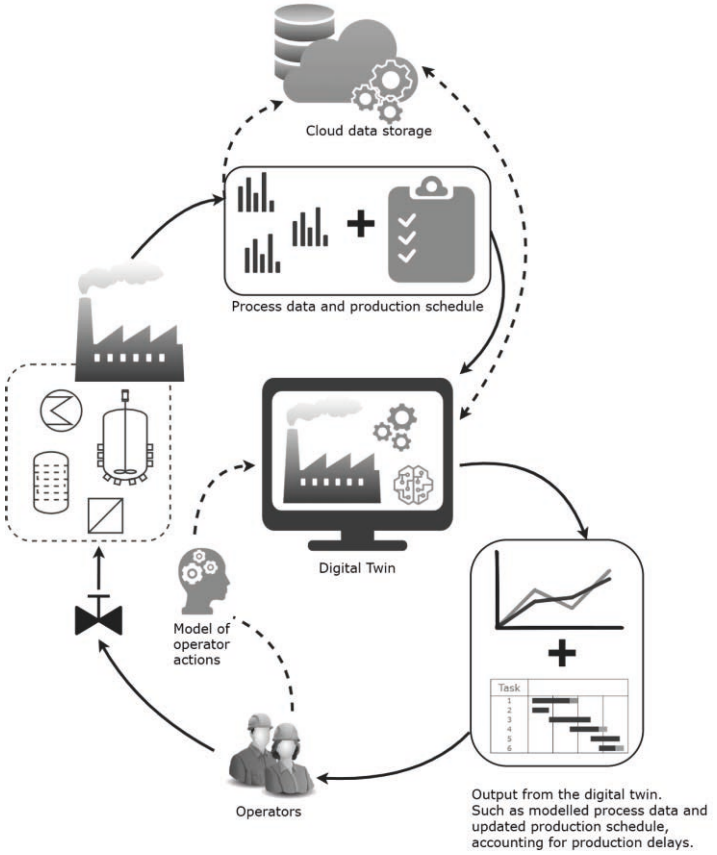


Figure 1: Conceptual framework for the development of Digital Twin for a process, which has not been automated.

4. Lessons learnt

This framework is currently applied to a pilot-scale *E. coli* fermentation, operated at the Bio-manufacturing, Training and Education Center (BTEC) at NC State University. This initial work identifies some minimum communication and modelling requirements, along with the potential for developing a fault detection algorithm for batch prediction.

4.1. Minimum Communication Requirements

For the DT to be realisable, the process in question needs to have some level of automation. If any data-driven or hybrid modelling approach is applied, the models require inputs from the sensors to operate and accurately estimate the state of the process.

To this end, there is a need to establish communication between the digital model and the physical system. Typically, the minimum requirement for bi-directional communication is a system consisting of digital actuators that a central DCS/SCADA can control. A critical constraint that has been identified in the current development process is the need for live extraction of the process data and production schedule. This requires a digital model with a fast simulation time and a high-level programming language to gather DCS/SCADA system data. This is due to the fact that acquiring process data in real-time is crucial for a fully integrated digital twin capable of two-way communication.

4.2. Minimum Modeling Requirements

An important focus area is the applicability and adaptability of the models and the overall implementation framework. The DT can be reasonably adaptable and customisable to different processes within bio-manufacturing. In addition, the overall digital model developed needs to execute in real-time under the constraints imposed by communication limitations and computational requirements. The need to move away from complete mechanistic modelling was identified as a method for fulfilling this requirement. To this end, the proposed strategy is based on hybrid modelling of key unit operations, which can run under a larger modelling framework.

4.3. Potential Fault Detection

Due to documentation requirements within the biopharmaceutical field, all the process data has to be stored. However, in many cases, the full potential of the process data is not explored. (Xu et al., 2019) presented an interesting view on the application of a DT for fault diagnosis. This is an exciting concept that could potentially be applied within bio-manufacturing where process data can be used to predict the variation for each batch and predict when a batch is at a high risk of failure due to off-specification product formation or high byproduct formation. This will allow the operators to decide whether or not to discard a batch or adjust the process conditions due to expected delays or a potential change in product quality.

5. Conclusions

This work detailed the development of a ‘ plug and play’ framework for creating a fully-fledged DTs in the domain of bio-manufacturing. This framework is customizable to the different processes based on the unit operations involved. Sufficient communication infrastructure and appropriate model choices were identified as key requirements for successfully implementing DT in bioprocessing industries. The pros and cons of mechanistic, data-driven, and hybrid modelling approaches were identified and reviewed. Overall, it was shown that a key to successfully applying a DT is the need for robust models, which are simple enough to be run in real- time while still accurate enough to provide adequate decision support capabilities.

Acknowledgements

This work is supported by the Accelerated Innovation in Manufacturing Biologics (AIM-Bio) project funded by the Novo Nordisk Foundation (Grant number NNF19SA0035474).

References

Anane, E., López C, D. C., Neubauer, P., & Cruz Bournazou, M. N. (2017). Modelling overflow metabolism in Escherichia coli by acetate cycling. *Biochemical*

Engineering Journal, *125*, 23–30.

Benalcázar, E. A., Noorman, H., Filho, R. M., & Posada, J. A. (2020). Modeling ethanol production through gas fermentation: A biothermodynamics and mass transfer-based hybrid model for microbial growth in a large-scale bubble column bioreactor. *Biotechnology for Biofuels*, *13*(1).

Bisgaard, J., Muldbak, M., Cornelissen, S., Tajsoleiman, T., Huusom, J. K., Rasmussen, T., & Gernaey, K. V. (2020). Flow-following sensor devices: A tool for bridging data and model predictions in large-scale fermentations. *Computational and Structural Biotechnology Journal*, *18*, 2908–2919.

Chen, R. T. Q., Rubanova, Y., Bettencourt, J., & Duvenaud, D. (2018). Neural Ordinary Differential Equations. *32nd Conference on Neural Information Processing Systems, NeurIPS 2018*, 6571–6583.

Grieves, M. (2015). Digital Twin : Manufacturing Excellence through Virtual Factory Replication. In *A Whitepaper*.

Lauterbach, T., Lüke, T., Büker, M. J., Hedayat, C., Gernandt, T., Moll, R., Grösel, M., Lenk, S., Seidel, F., Brunner, D., Bley, T., Walther, T., & Lenk, F. (2019). Measurements on the fly– Introducing mobile micro-sensors for biotechnological applications. *Sensors and Actuators, A: Physical*, *287*, 29–38.

Lukowski, G., Rauch, A., & Rosendahl, T. (2018). The Virtual Representation of the World is Emerging. In P. Krüssel (Ed.), *Future Telco* (pp. 165–173). Springer International Publishing.

Millard, P., Smallbone, K., & Mendes, P. (2017). Metabolic regulation is sufficient for global and robust coordination of glucose uptake, catabolism, energy production and growth in Escherichia coli. *PLoS Computational Biology*, *13*(2).

Reinecke, S., & Hampel, U. (2018). Investigation of bioreactors by smart sensor particles. *Chemie Ingenieur Technik*, *90*(9), 1268–1268.

Udugama, I. A., Lopez, P. C., Gargalo, C. L., Li, X., Bayer, C., & Gernaey, K. V. (2021). Digital Twin in biomanufacturing: challenges and opportunities towards its implementation. *Systems Microbiology and Biomanufacturing*, *1*(3), 257–274.

Udugama, I. A., Öner, M., Lopez, P. C., Beenfeldt, C., Bayer, C., Huusom, J. K., Gernaey, K. V., & Sin, G. (2021). Towards Digitalization in Bio-Manufacturing Operations: A Survey on Application of Big Data and Digital Twin Concepts in Denmark. *Frontiers in Chemical Engineering*, *3*.

von Stosch, M., Hamelink, J. M., & Oliveira, R. (2016). Hybrid modeling as a QbD/PAT tool in process development: an industrial E. coli case study. *Bioprocess and Biosystems Engineering*, *39*(5), 773–784.

von Stosch, M., Portela, R. M., & Varsakelis, C. (2021). A roadmap to AI-driven in silico process development: bioprocessing 4.0 in practice. *Current Opinion in Chemical Engineering*, *33*, 100692.

Xu, Y., Sun, Y., Liu, X., & Zheng, Y. (2019). A Digital-Twin-Assisted Fault Diagnosis Using Deep Transfer Learning. *IEEE Access*, *7*, 19990–19999.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50237-2

Plant O&M Support System Based on Supervised Data-Clustering Technology

Yoshinari Hori[a*], Takaaki Sekiai[a], Hiroto Takeuchi[b]

[a] *Research & Development Group, Hitachi, Ltd., Ibaraki 319-1292 JAPAN*
[b] *Service & Platforms Business Unit, Hitachi, Ltd., Ibaraki 319-1293 JAPAN*
yoshinari.hori.ns@hitachi.com

Abstract

We propose a plant operations and maintenance (O&M) support system that can correlate plant operational data to both key performance indicators (KPIs) and qualitative data such as plant statuses or anomaly types. The system consists of a supervised data-clustering function and a visualization function for clustered results. The supervised data-clustering function is based on the ART2 network and can correlate plant operational data to KPIs and qualitative data by classifying the operational data. Therefore, the operational data are classified into different categories when the KPIs of the operational data are the same and their qualitative data are different. The visualization function can illustrate the relation between the operational data and the KPIs and the qualitative data as a three-dimensional graph or a two-dimensional graph. An x-y plane showing the gravity centres of the categories is mapped by multidimensional scaling, and the qualitative data are displayed beside dots in both graph types. KPIs are indicated by the z-axis in the three-dimensional graph and by dot colours in the two-dimensional graph. We examined whether the system can be used to analyze plant operational data using process simulation data. The results demonstrate that our system can monitor and maintain industrial plant performance.

Keywords: Operation; Maintenance; Data clustering; Applications.

1. Introduction

To maintain high productivity at industrial plants, it is important to monitor key performance indicators (KPIs) such as plant operational efficiency.

One of the technologies for monitoring plant KPIs is soft sensors. For example, Okada et al. (2012) developed an adaptive soft sensor technique to respond to the degradation of soft sensors. Kim et al. (2013) developed an inferential control system that integrates soft sensors and model predictive control. In addition, Matsui et al. (2014) estimated product quality by using PLS in the deposition process of a thin-film solar cell and visualised the relationship between the main two variables and quality in a three-dimensional graph.

Hori et al. (2019) have developed a performance evaluation system for industrial plants that uses data clustering technology. The system can be employed to analyze plant operational data by using the KPIs and consists of a data clustering function for operational data of the KPIs and a visualisation function of clustering results. The data clustering function is based on the ART2 network (Carpenter and Grossberg, 1987), which is one of the adaptive resonance theory (ART) networks. It can correlate plant operational data to KPIs by classifying operational data. The visualisation function of clustering results illustrates the relation between operational data and KPIs by using a 3D

graph with a z-axis indicating the KPIs and an x-y plane showing where the gravity centres of the categories are mapped by multidimensional scaling (MDS).

In this study, we enhance that system and propose a plant operations and maintenance (O&M) support system that can correlate plant operational data to both KPIs and qualitative data such as plant statuses or anomaly types. We also examine whether the system can be used to analyse plant operational data by using process simulation data.

2. Proposed system

2.1 Overview of the proposed system

Figure 1 shows an overview of the proposed plant O&M support system using data clustering technology.

The system consists of a supervised data clustering function and a visualisation function of clustering results. The basics of each function are described below. Hereafter, the term "qualitative data such as plant status or anomaly type", is simply referred to as "plant status" or "status".

In the supervised data clustering function, the operational data of the plant are classified into multiple categories, and the relationship between a category and a selected KPI and its status is set. The centre of Figure 1 shows the supervised data clustering results. To simplify the explanation, the operational data are two-dimensional data. The points plotted in the graph are the operational data of each time, and the colour of the point represents the KPI. Those with a high KPI value are black and those with a low KPI value are white. The circle surrounding the operational data represents a category that classifies the data. In the example in Figure 1, the operational data are classified into eight categories. The type of line drawing each circle corresponds to the plant status. Solid, dashed, and dotted lines indicate normal (Norm.), abnormal (Abnl.) A and abnormal (Abnl.) B, respectively.

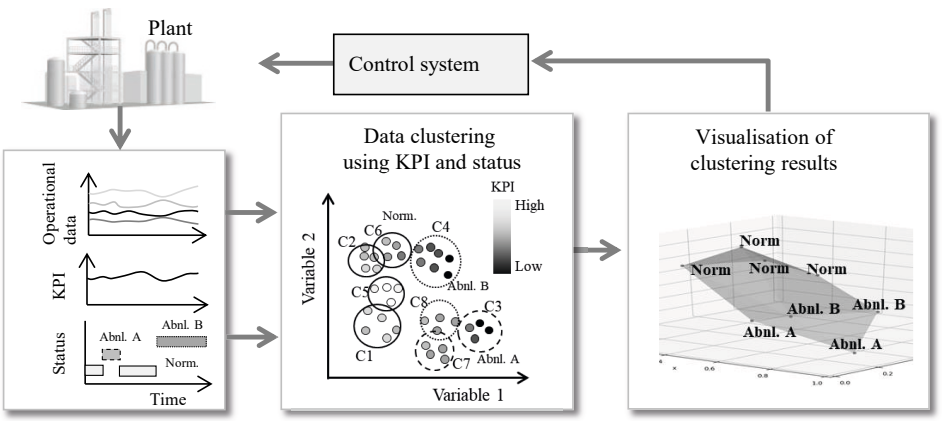


Figure 1 Schematic of the proposed system

In the visualisation function of the clustering results, the relationship between categories and KPI and plant status are visualized in three-dimensional graphs. The Z-axis, that is, the vertical axis indicates the KPI, the XY plane represents the positional relationship of the centre of gravity of the data classified into each category (hereinafter referred to as the gravity centre of the category). Since the gravity centre of the category is N-dimensional data, it is not possible to accurately represent the positional relationship on a two-dimensional plane. However, it is possible to express the approximate positional relationship by using the MDS method. Therefore, as shown in Figure 1, the function can display the positional relationship of the categories with their KPIs and plant status in a three-dimensional graph. That information is useful to support O&M of industrial plant.

In the above, a three-dimensional graph with a KPI axis has been described. However, a two-dimensional graph expressing the value of KPI in colour can be used, instead.

2.2 Data clustering algorithm

The algorithm of the supervised data clustering function is described below.

Step 1: Read multi-dimensional operational data, KPI and plant status and set the vigilance parameter ρ_j for category j to a given initial value.

Step 2: Classify the operational data by using the vigilance parameter ρ_j set in the previous step. The vigilance parameter ρ_j determines the size of category j; namely, the larger ρ_j is, the smaller the size of the category j is.

Step 3: Calculate V_j and M_j by Eqs. (1) and (2).

$$V_j = KPI\_max_j - KPI\_min_j \tag{1}$$

$$M_j = Nm_j / N_j \tag{2}$$

Herein, KPI_max_j is the maximum value of KPI of the data classified into category j and KPI_max_j is the minimum value of KPI of the data classified into category j. N_j is the number of data in category j, N_m is the number of data with the highest proportion of plant status.

Step 4: If the V_j calculated in step 3 is above V_{set}, the set value of V_j, or M_j is below M_{set}, the set value of M_j, or the number of calculations is less than the given value, proceed to step 5. Otherwise, proceed to step 6.

Step 5: Update the value of ρ_j by selecting the larger of ρ_{V_j} and ρ_{M_j} calculated by Eq. (3) and Eq. (4), respectively and return to step 2.

$$\rho_{V_j} = \frac{V_{set} \times \rho_j + a_V(V_j - V_{set}) \times \rho_{max}}{V_{set} + a_V(V_j - V_{set})} \tag{3}$$

$$\rho_{M_j} = \frac{M_{set} \times \rho_j + a_M\left(M_{set} - M_j\right) \times \rho_{max}}{M_{set} + a_M\left(M_{set} - M_j\right)} \tag{4}$$

Herein, in Eqs. (3) and (4) , $\rho_{\max}$ is the maximum value of ρ in this system and a_V and a_M are adjustment coefficients.

Step 6: Calculate the average value of the KPI and select the plant status and map the relationship between the categories and both the KPIs and plant status.

Figure 2 is an explanatory diagram of the above algorithm. Figure 2(a) represents an example of the initial condition of categories, or the clustering result in step 2. In step 2, ρ_j is set to an initial value, then all the categories are the same size. Therefore, a category could have data with a wide range of KPI values and it could have data with different plant status. In Figure 2(a), category C1 has data with high and middle KPI values and category C3 has data in abnormal A and in abnormal B. However, in going from step 3 to step 5, if necessary, the size of the categories gets smaller. In Figure 2(b), category C5 is created for data with high KPI value and category C8 is for data in abnormal B. In this way, each category is supposed to have data with a given range of KPI values and the same plant status.

In the situation where two operational data are very similar and the KPI values are very different or their plant statuses are different because of noise, it is difficult to classify the data into different categories and the category sizes can become very small for such categories. In this case, the value of V_{set} must be made larger or the value of M_{set} must be made smaller. For example, if the value is set to 0.9, 10 % of the data classified in each category can be for another plant status, which can avoid creating categories that are too small.

As described above, the relationship between the multi-dimensional operational data and KPIs and plant status is learned by the proposed algorithm

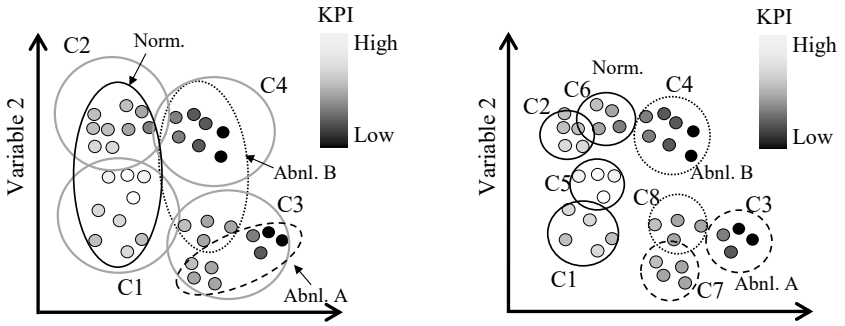


Figure 2 Explanatory diagram of proposed algorithm

3. Data analysis and results

We examined whether the system can be utilised to analyse plant operational data by using Tennessee Eastman process data, which are commonly used as benchmark data for evaluating anomaly detection technology (Downs and Vogel,1993). Several variations of data sets of Tennessee Eastman process data are open to the public. The data used in this study contained 15 data sets with different disturbances: IDV(1) to IDV(15)(N. L. Ricker, 1996). Each data set had time series data which consisted of 41 measured variables, 12 manipulated variables and 10 calculated variables including operational cost, 63 variables in total. The sampling interval was 10 min and each data set contained 301 samples (data covering 50 h). We chose operational cost as the KPI and disturbance as plant status and made test data for the data analysis.

In particular, we chose three data sets IDV(1), IDV(2) and IDV(6) shown in Table 1 for the test data because of their large change in operational cost. Time series data of operational cost for these IDVs are shown in Figure 3. We also chose 9 manipulated variables and for operational data from the 63 variables. The variables whose value were constant in the data sets were not used for the test data. Data analysis was conducted using the test data where V_{set} was 0.4 and M_{set} was 0.9. The results are shown in Figure 4.

Figure 4 is a two-dimensional MDS graph. Each dot indicates the gravity centre of each category and the colour of each dot indicates the KPI value. The legend of each dot indicates plant status. Areas A, B, and C were seen in Figure 4. Operational data of IDV(1) with low KPI values were classified into the categories in area A, operational data of IDV(1) data with high KPI values were classified into the categories in area B, and operational data of IDV(6) with high KPI values were classified into the categories in area C. Then, the KPI value and plant status can be assumed from the area where the category is located.

Table 1 Analysis data

Disturbance	Description	Type
IDV (1)	A/C feed ratio, B composition constant (stream 4)	Step
IDV (2)	B composition, A/C ratio constant (stream 4)	Step
IDV (6)	A feed loss (stream 1)	Step

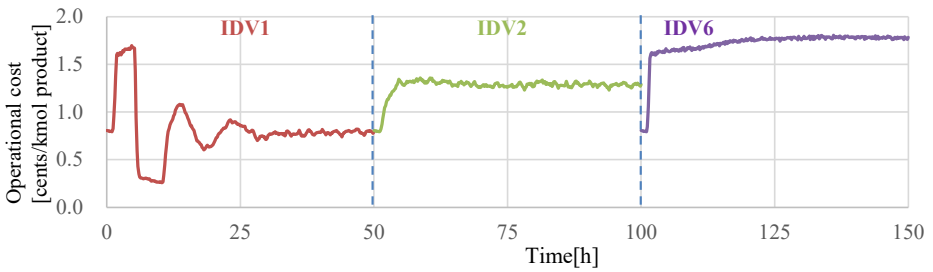


Figure 3 Operational cost of test data

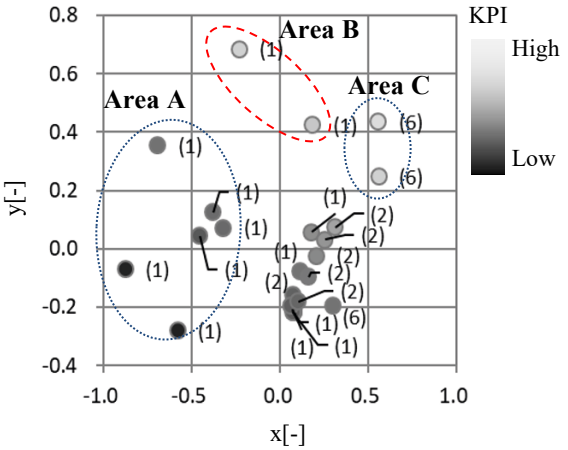


Figure 4 Results of data analysis

4. Conclusions

We proposed the plant O&M support system that can correlate plant operational data to both KPIs and plant statuses. The system consists of a supervised data-clustering function and a visualisation function for clustered results, and it learns the relationship between the multi-dimensional operational data and both KPI and plant statuses.

We examined whether the system was applicable to analysis of plant operational data by using Tennessee Eastman process data sets and setting the operational cost as the KPI and the disturbance as plant status.

The results showed the proposed algorithm classified the operational data into categories according to their KPI values and plant status and our system was demonstrated as useful to support O&M of industrial plants.

References

G. Carpenter and S. Grossberg, 1987, "ART2:Self-Organization of Stable Category Recognition Codes for Analog Input Patterns," *Applied Optics*, 26, 23, 4919–4930

Y. Hori, T. Shiraishi, and S. Hasebe, 2019 "Performance Evaluation System for Industrial Plants Using the ART2 Network That Considers Cluster Size," *Kagaku Kogaku Ronbunshu,* 35, 127-132

S. Kim, M. Kano, S. Hasebe, A. Takinami, and T. Seki, 2013, "Long-Term Industrial Applications of Inferential Control Based on Just-In-Time Soft-Sensors: Economical Impact and Challenges," *Ind. Eng. Chem. Res.*, 52-35, 12346-12356

T. Matsui, K. Murakami, and S. Suzuki; 2014, "Data Analysis Technology in Plant Control," *Fuji Denki Gihoh*, 87, 33-37

T. Okada, H. Kaneko, and K. Funatsu, 2012,"Development of An Adaptive Soft Sensor Method Considering Prediction Confidence of Models," *J. Comput. Chem. Jpn.*, 11, 24-30

N. L. Ricker, 1996, "Decentralized control of the Tennessee Eastman Challenge Process," *J. Proc. Cont.*, 6, 205-221

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50238-4

A Novel Cycle Partitioning Approach to Reliability Based Optimal Sensor Placement for Linear Flow Processes

Om Prakash[a*] and Mani Bhushan[a]

[a]*Department of Chemical Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India*

prakash.om@iitb.ac.in

Abstract

In literature, system reliability for sensor placement design for linear mass flow processes has been defined in terms of cycles (Prakash et al., 2020a). The computation of system reliability involves computation of probability of a union of a set of events, which is NP-hard, thereby limiting its applicability to large processes. To reduce this computational complexity, in the current work, we propose to partition the set of cycles, instead of partitioning the original process. Thus, our approach involves approximating the system reliability function of the original process, rather than approximating the original process. Towards this end, we use a bipartite spectral graph partitioning algorithm to obtain the partitions of cycles. The resulting sensor placement design approach is applied to a case study to show its efficacy.

Keywords: Bipartite graph, Estimation of variables, Graph partitioning

1. Introduction

Sensor placement design (SPD) is the problem of selecting key variables to be measured in the process, so as to maximize some performance criteria while satisfying various design constraints (Ali and Narasimhan, 1995; Bhushan and Rengaswamy, 2002). The focus of the current work is on sensor placement design for the application in estimation of variables in a steady state linear flow process. A pure mass flow process, e.g., water distribution network, is known as a linear flow process, where linear mass balances can be written to describe the process model (Ali and Narasimhan, 1995). Linear flow process is conveniently modeled as a directed graph (process graph) with edges in the graph representing the flow variables (process variables), and vertices representing the process units. SPD is to then appropriately select a set of flow variables to be measured. In the current work, we focus on system reliability based criteria for SPD given that sensors are prone to failure with known probabilities.

The area of SPD for ensuring reliable estimation for linear flow processes has received considerable attention in literature. Most of the works in literature have defined system reliability as the minimum reliability of estimation of individual variables (Ali and Narasimhan, 1995; Bagajewicz and Sánchez, 2000; Kotecha et al., 2008). Recently, for use in SPD, Prakash et al. (2020a) defined the system reliability as the probability of estimating all variables. This definition captures the interacting nature of reliabilities, which

was missing in the previous definitions. Prakash et al. (2020a) used the result that all variables in a steady state linear flow process can be estimated if the unmeasured variables do not form a cycle (Mah et al., 1976). Prakash et al. (2020a) expressed system reliability as one minus probability of having at least one cycle with all unavailable variables. They defined an unavailable variable as a variable which is either not measured by any sensor, i.e., the variable is unmeasured, or all sensors measuring that variable have failed. In their work (Prakash et al., 2020a), the computation of system reliability involves computation of probability of a union of a set of events, which is NP-hard (Veeraraghavan and Trivedi, 1991). (Prakash et al., 2020b) proposed a sum of disjoint product based approach to systematically compute this probability. However, the number of terms in the expression of system reliability exponentially grows with the number of cycles. To illustrate, Prakash et al. (2020b) have considered steam metering process for sensor placement design, which has 12 units, 28 variables, and 3209 cycles. The system reliability expression consisted of 334630 terms. For larger case studies, the number of terms will be even more, and thus will pose challenges related to: (a) computation, (b) storage, and (c) quick evaluation, of the system reliability expression. These challenges limit the applicability of system reliability based SPD for large processes.

Addressing these issues is the motivation for our current work. A traditional heuristic to deal with large processes is to partition the original graph into several disconnected subgraphs by removing a set of edges (Rajeswaran et al., 2018), and then performing SPD in each of these subgraphs. However, this will lead to loss of edges (process variables), and loss of cycles in general as the collection of cycles in the two resulting subgraphs will not lead to the cycles in the original graph. In the current work, we propose a novel approach that partitions the set of cycles and not the original process. Our approach ensures that the collection of cycles in the partitions is equal to the set of cycles in the original graph. Towards this end, we propose to compute the approximated system reliability which is maximized to obtain SPD. Further, we present a case study to demonstrate its efficacy.

The rest of the paper is organized as follows. Relevant existing information is presented in Section 2.. Section 3. presents the novel approach of partitioning cycles, and SPD formulation. Section 4. presents a case study. Finally, Section 5. concludes the work.

2. Preliminaries related to reliability based sensor placement design

For a steady state linear flow process, reliability based SPD for the application in estimation of variables is posed as (Prakash et al., 2020b),

$$\max_{\{q_i\}_{i=1}^{n_v}} R_{sys} = 1 - P\left(\bigcup_{k=1}^{n_c} \mathcal{S}_k\right);\ \text{s.t.,}\ \sum_{i=1}^{n_v} c_i q_i \leq C^*,\ \text{ and, } q_i \in \mathbb{Z}_{\geq 0}, i = 1, ..., n_v \quad (1)$$

where c_i and q_i are cost of a sensor and number of sensors placed to measure the ith variable, respectively, C^* is the total cost available for SPD, and n_v is the number of process variables that can be measured in the process. The system reliability objective R_{sys} in Eq. (1) is the probability of estimating all variables in the process and is defined in terms of cycles. S is the set of cycles in the process graph, and n_c is its cardinality. $\mathcal{S}_k$ denotes the event that kth cycle (S_k) in the process graph has all unavailable variables. Probability that the ith variable is unavailable is $s_i^{q_i}$, where s_i is the failure probability of the sensor

measuring the ith variable, provided full active redundant configuration is followed. The SPD formulation given in Eq. (1) is a non-linear integer programming problem owing to the nature of R_{sys}. Computation of R_{sys} in Eq. (1) involves computation of probability of a union of a set of events, which is NP-hard (Veeraraghavan and Trivedi, 1991). Prakash et al. (2020b) proposed a sum of disjoint product based approach to systematically compute R_{sys}.

3. Proposed cycle partitioning based sensor placement design

The length of expression of R_{sys} which is the optimization objective in SPD (Eq. (1)), grows exponentially with the number of cycles. This poses significant difficulties in solving the SPD for a large scale problem, since computation, storage and quick evaluation of R_{sys} become a challenge. One intuitive heuristic to overcome this issue would be to partition the process graph into several disjoint subgraphs by removing edges in the graph. Subsequently, SPD can be solved for each of these subgraphs. But, this approach suffers from two drawbacks: (a) removing edges in the graph eliminates the corresponding process variables from process graph, and (b) it leads to a loss of cycles in general as the collection of cycles in the subgraphs does not recreate the cycles in the original graph. Thus, this approach involves approximating the process, i.e., solving SPD for different (simpler) processes instead of the original process.

In the current work, we propose to approximate the objective function R_{sys} in the SPD (Eq. (1)) without modifying the process. Towards this end, we propose a novel idea to partition the set of cycles S into two disjoint partitions S^a and S^b, where $S^a \cap S^b = \emptyset$, and $S^a \cup S^b = S$. Thus, we do not eliminate any process variables from process graph, and the cycles are also preserved. The system reliability is then approximated as the product of reliabilities corresponding to these two partitions. To ensure that this approach leads to computational benefits for SPD, the partitions S^a, S^b have to be appropriately chosen. To discuss this issue, let us consider that we have the partitions. Then, we can write the system reliability R_{sys} as (Eq. (1)),

$$R_{sys} = 1 - P\Bigg(\underbrace{\Bigg(\bigcup_{k_a=1}^{n_{c_a}} \mathcal{S}^a_{k_a}\Bigg)}_{\alpha} \bigcup \underbrace{\Bigg(\bigcup_{k_b=1}^{n_{c_b}} \mathcal{S}^b_{k_b}\Bigg)}_{\beta}\Bigg) = 1 - (P(\alpha) + P(\beta) - P(\alpha|\beta)P(\beta)) \tag{2}$$

where $\mathcal{S}^a_{k_a}$ and $\mathcal{S}^b_{k_b}$ denote the event that k_ath and k_bth cycle in S^a and S^b have all unavailable variables, respectively, and n_{c_a} and n_{c_b} are cardinalities of S^a and S^b, respectively. In Eq. (2), $P(\alpha)$ and $P(\beta)$ denote the probability of having at least one cycle in S^a and S^b with all unavailable variables, respectively. Now, consider, $R^a = 1 - P(\alpha)$ and $R^b = 1 - P(\beta)$ as the reliabilities corresponding to the partitions. The product $R^aR^b = 1 - (P(\alpha) + P(\beta) - P(\alpha)P(\beta))$ can be thought of as the approximation of R_{sys}. However, it will be exactly equal to R_{sys} only when the partitions are such that $P(\alpha|\beta) = P(\alpha)$ in Eq. (2). This equality will not hold in general, because many process variables are common to the cycles in partitions S^a and S^b, leading to dependent nature of α and β. Also, note that $P(\alpha|\beta) \geq P(\alpha)$, as the occurrence of β increases the probability of occurrence of α. This gives rise to the following criteria of partitioning the cycles,

$$\min_{S^a, S^b \subseteq S} P(\alpha|\beta) - P(\alpha), \text{ and } \min_{S^a, S^b \subseteq S} \text{abs}(|S^a| - |S^b|) \tag{3}$$

where abs(·) denotes the absolute value. First criteria of Eq. (3) maximizes the independence of α and β, while the second criteria of Eq. (3) ensures that the obtained partitions are balanced. It is required because we do not want length of the reliability expression (R^a or R^b) corresponding to one of the partitions to increase significantly. Note that characterizing independence of the partitions involves computation of $P(\alpha|\beta)$, which is equivalent to computing R_{sys}. One heuristic to maximize independence is to minimize the number of process variables that are common to the partitions. This has to be achieved while ensuring that the partitions are balanced (second criteria in Eq. (3)). This is a well studied problem in graph partitioning literature, and is known to be NP-complete (Dhillon, 2001). Spectral graph partitioning algorithm is a widely used heuristic to effectively solve this problem. In this work, we adapt the bipartite spectral graph partitioning algorithm (Dhillon, 2001) to solve our cycle partitioning problem. It is to be noted that we are not partitioning the original process graph, instead we are partitioning the set of cycles using the graph partitioning algorithm. Next, we formally propose this algorithm in the context of our application.

3.1. Cycle-variable bipartite graph partitioning algorithm and SPD formulation

We propose to model the cycle-variable relationship with an undirected bipartite graph $G = (S, V, E)$. $S = \{S_1, \ldots, S_{n_c}\}$ and $V = \{v_1, \ldots, v_{n_v}\}$ denote the set of vertices, and $E = \{\{S_i, v_j\}, \forall S_i \in S \text{ and } v_j \in V\}$ denotes the set of undirected edges. In our work, S is the set of cycles, V is the set of all process variables, and edge $\{S_i, v_j\}$ exists if the process variable v_j is contained in cycle S_i. The partitioning problem is to find balanced partitions P_1^* and P_2^* of $S \cup V$ such that the number of common edges $\{S_i, v_j\}$ between these partitions is minimized. Consider the following terminologies related the graph. The adjacency matrix $\mathbf{A} \in \mathbb{R}^{(n_c+n_v)\times(n_c+n_v)}$ of the bipartite graph G is given as, $\mathbf{A} = \begin{bmatrix} [\mathbf{0}] & \mathbf{C} \\ \mathbf{C}^\top & [\mathbf{0}] \end{bmatrix}$, where $\mathbf{C} \in \mathbb{R}^{n_c \times n_v}$ represents the cycle by variable matrix with $C_{i,j} = 1$ if $\{S_i, v_j\}$ exists, and is 0 otherwise. First n_c rows and columns in $\mathbf{A}$ corresponds to the cycles, and last n_v rows and columns corresponds to the process variables. The diagonal degree matrix $\mathbf{D} \in \mathbb{R}^{(n_c+n_v)\times(n_c+n_v)}$ is the matrix with its element $D_{l,l} = \sum_k A_{l,k}$. $\mathbf{D}$ is written as, $\mathbf{D} = \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{bmatrix}$, where $\mathbf{D}_1$ and $\mathbf{D}_2$ matrices correspond to the block matrices in $\mathbf{A}$. The spectral partitioning heuristic given in Dhillon (2001) which is used to solve our cycle partitioning problem is now presented as a theorem.

Theorem 1 (Dhillon (2001)) *Given a bipartite graph G, $\mathbf{x}_2 = \mathbf{D}_1^{-1/2}\mathbf{u}_2$ is the partition vector, where $\mathbf{u}_2$ is the left singular vector corresponding to the second largest singular value of $(\mathbf{D}_1^{-1/2}\mathbf{C}\mathbf{D}_2^{-1/2})$. Partitioning of cycles is obtained by assigning $\mathbf{x}_2$ to bi-modal values.*

Theorem 1 presents an efficient heuristic to obtain the partitions S^a and $S^b \subseteq S$. Having these partitions, we propose the SPD as follows,

$$\max_{\{q_i\}_{i=1}^{n_v}} \widehat{R}_{sys} = (1 - P(\alpha))(1 - P(\beta)); \text{ s.t., } \sum_{i=1}^{n_v} c_i q_i \leq C^*, \text{ and, } q_i \in \mathbb{Z}_{\geq 0} \tag{4}$$

where $\widehat{R}_{sys}$ is the approximated system reliability. Next, we present a case study, where we have applied the proposed SPD (Eq. (4)) to obtain the sensor placement.

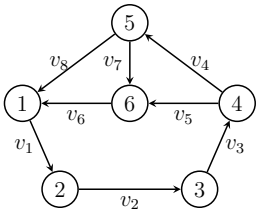


Figure 1: Ammonia process graph (Prakash et al., 2020a)

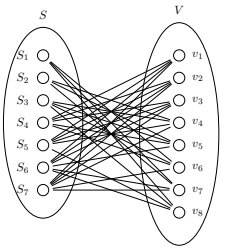


Figure 2: Ammonia bipartite graph G

4. Case Study

We present the Ammonia process case study which has been widely used for data reconciliation and gross error detection (Ali and Narasimhan, 1995). It has 6 nodes including the environmental node and 8 flow variables where sensors can be placed to measure the flow rates. There are 7 cycles in the process graph (Prakash et al., 2020a). The process graph and its cycle-variable bipartite graph are presented in Figures 1 and 2, respectively. The sensor failure probabilities are taken from Prakash et al. (2020a), and unit cost is considered for all sensors. The set of cycles are partitioned using the algorithm presented in Section 3.1., and then sum of disjoint product based approach is applied to compute $P(\alpha)$ and $P(\beta)$. The resulting expressions are,

$$P(\alpha) = s_6^{q_6} s_7^{q_7} s_8^{q_8} + (1 - s_6^{q_6} s_8^{q_8}) s_4^{q_4} s_5^{q_5} s_7^{q_7} + (1 - s_7^{q_7}) s_4^{q_4} s_5^{q_5} s_6^{q_6} s_8^{q_8} \tag{5}$$

and, $P(\beta) = s_1^{q_1} s_2^{q_2} s_3^{q_3} s_5^{q_5} s_6^{q_6} + (1 - s_5^{q_5} s_6^{q_6}) s_1^{q_1} s_2^{q_2} s_3^{q_3} s_4^{q_4} s_8^{q_8} + (1 - s_6^{q_6})(1 - s_4^{q_4})$

$$s_1^{q_1} s_2^{q_2} s_3^{q_3} s_5^{q_5} s_7^{q_7} s_8^{q_8} + (1 - s_5^{q_5})(1 - s_8^{q_8}) s_1^{q_1} s_2^{q_2} s_3^{q_3} s_4^{q_4} s_6^{q_6} s_7^{q_7} \tag{6}$$

Now, we compute the approximated system reliability $\widehat{R}_{sys}$ (Eq. (4)). To compare, we also compute the exact expression of system reliability R_{sys} to be:

$$\begin{aligned} R_{sys} = 1 - \big[& s_6^{q_6} s_7^{q_7} s_8^{q_8} + (1 - s_6^{q_6} s_8^{q_8}) s_4^{q_4} s_5^{q_5} s_7^{q_7} + (1 - s_7^{q_7}) s_4^{q_4} s_5^{q_5} s_6^{q_6} s_8^{q_8} + \\ & ((1 - s_4^{q_4} s_8^{q_8})(1 - s_7^{q_7}) + s_7^{q_7}(1 - s_4^{q_4})(1 - s_8^{q_8})) s_1^{q_1} s_2^{q_2} s_3^{q_3} s_5^{q_5} s_6^{q_6} + \\ & ((1 - s_5^{q_5} s_6^{q_6})(1 - s_7^{q_7}) + s_7^{q_7}(1 - s_5^{q_5})(1 - s_6^{q_6})) s_1^{q_1} s_2^{q_2} s_3^{q_3} s_4^{q_4} s_8^{q_8} + \\ & (1 - s_6^{q_6})(1 - s_4^{q_4}) s_1^{q_1} s_2^{q_2} s_3^{q_3} s_5^{q_5} s_7^{q_7} s_8^{q_8} + (1 - s_8^{q_8})(1 - s_5^{q_5}) s_1^{q_1} s_2^{q_2} s_3^{q_3} s_4^{q_4} s_6^{q_6} s_7^{q_7} \big] \end{aligned} \tag{7}$$

It can be noted that the total number of terms in R_{sys} is 9, while $P(\alpha)$ and $P(\beta)$ have 3 and 4 terms only. Thus, there is an overall reduction in the length of combined expressions of $P(\alpha)$ and $P(\beta)$. Further, we apply the proposed SPD (Eq. (4)) on the Ammonia process to obtain the sensor placement. We also maximize the exact system reliability R_{sys} (Eq. (1)) to enable comparison with the proposed SPD formulation. Enumeration is used to solve these problems. It is observed that for each C^* (Table 1), the solution is same for both the problems thereby validating the utility of the proposed approach.

5. Conclusions

In the current work, we propose a novel cycle partitioning approach to solve SPD for large processes. To achieve this, we model the cycle-variable relationship as a bipartite

Table 1: SPD results for ammonia process

C^*	Exact R_{sys} (Eq. (1))		Proposed SPD formulation (Eq. (4))	
	Selected vars.	R_{sys}	Selected vars.	$\widehat{R}_{sys}$
3	$\{v_1, v_4, v_7\}$	0.4909	$\{v_1, v_4, v_7\}$	0.4909
4	$\{v_1, v_4, v_6, v_7\}$	0.6894	$\{v_1, v_4, v_6, v_7\}$	0.6777
5	$\{v_1, v_4, v_5, v_7, v_8\}$	0.8124	$\{v_1, v_4, v_5, v_7, v_8\}$	0.8083

graph. We then apply a bipartite spectral graph partitioning algorithm to obtain the cycle partitions. The resulting partitions are used to approximate system reliability objective for SPD. The approach is demonstrated on a case study.

References

Ali, Y., Narasimhan, S., 1995. Redundant sensor network design for linear processes. AIChE Journal 41, 2237–2249.

Bagajewicz, M., Sánchez, M.C., 2000. Cost-optimal design of reliable sensor networks. Computers & Chemical Engineering 23, 1757–1762.

Bhushan, M., Rengaswamy, R., 2002. Comprehensive design of a sensor network for chemical plants based on various diagnosability and reliability criteria. 1. Framework. Industrial & Engineering Chemistry Research 41, 1826–1839.

Dhillon, I.S., 2001. Co-clustering documents and words using bipartite spectral graph partitioning, in: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 269–274.

Kotecha, P.R., Bhushan, M., Gudi, R., Keshari, M., 2008. A duality based framework for integrating reliability and precision for sensor network design. Journal of Process Control 18, 189–201.

Mah, R.S., Stanley, G.M., Downing, D.M., 1976. Reconciliation and rectification of process flow and inventory data. Industrial & Engineering Chemistry Process Design and Development 15, 175–183.

Prakash, O., Bhushan, M., Narasimhan, S., Rengaswamy, R., 2020a. Sensor network design based on system-wide reliability criteria. Part I: Objectives. Journal of Process Control 93, 66–82.

Prakash, O., Bhushan, M., Narasimhan, S., Rengaswamy, R., 2020b. Sensor network design based on system-wide reliability criteria. Part II: Formulations and applications. Journal of Process Control 93, 14–27.

Rajeswaran, A., Narasimhan, S., Narasimhan, S., 2018. A graph partitioning algorithm for leak detection in water distribution networks. Computers & Chemical Engineering 108, 11–23.

Veeraraghavan, M., Trivedi, K.S., 1991. An improved algorithm for symbolic reliability analysis. IEEE Transactions on Reliability 40, 347–358.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50239-6

The impact of sampling frequency on chemical process monitoring

Tingting Tao, Jiatao Wen, Yang Li, Cheng Ji, Jingde Wang*, Wei Sun*

College of Chemical Engineering, Beijing University of Chemical Technology, North Third Ring Road 15, Chaoyang District, Beijing, 100029, China
sunwei@mail.buct.edu.cn

Abstract

Data-driven process monitoring technology, which is generally realized by extracting data feature under normal operating conditions, has been developed rapidly to ensure operation safety. Correlation analysis, including auto-correlation and cross-correlation, plays an important role in these feature extraction algorithms. In chemical industrial processes, auto-correlation and cross-correlation in process measurements are always dynamic and could be greatly affected by sampling frequency. The extraction of these features directly determines the generalization ability of the process monitoring model. Therefore, it is important to study the influence of different sampling frequencies on the auto-correlation and cross-correlation of process variables. With the rapid development of data acquisition technology, data of multiple sampling frequencies can be flexibly acquired from distributed control system (DCS), which provides a basis for this research. The aim of this work is to build a process monitoring method by using the optimal sampling strategy and dynamic principal component analysis (DPCA). Time series analysis is first applied to check the degree of auto-correlation of sequence data at different scales according to industrial data with different sampling frequencies. It can be concluded that process dynamic has different effects on different variables. Dynamic principal component analysis is then used for feature extraction and process monitoring. Monitoring results under different sampling frequencies are compared and analysed to obtain an optimal process monitoring method. The research is implemented on Tennessee Eastman Process (TEP). The method shows better results than other methods without the consideration of sampling strategies and dynamic characteristics.

Keywords: Correlation analysis; Feature extraction; Dynamic principal component analysis; Optimal sampling strategy.

1. Introduction

With the development of process industry, chemical production has been gradually getting larger scale and more complicated operation, which is accompanied by a great potential safety hazard. Once a safety accident occurs, it will not only cause economic losses, but also threaten human safety. To avoid safety accidents, process monitoring technology has been developed in chemical processes to assist operators to detect abnormal changes in a process plant timely. With the wide application of distributed control system (DCS), data-driven process monitoring technology has been extensively studied in order to ensure production safety. Whether it is the commonly used multivariate statistical method or recently proposed deep learning-based methods, the main aim is to extract data feature under normal operating conditions by projecting or mapping high-dimensional process data into low-dimensional space. A conventional multivariate

statistical procedure: Principal Component Analysis (PCA), assuming that the observations are static and independent of each other, can simplify and improve process monitoring procedures (MacGregor, et al., 1995). Dynamic principal component analysis (DPCA) was proposed with the adoption of augment matrix to describe the dynamic characteristics of data (Ku, et al., 1995). The data matrix of DPCA contains hysteresis information, which makes it more suitable for monitoring of sequence-related data. Deep learning-based methods, such as neural networks and autoencoder, also achieved good results on simulation data of Tennessee Eastman Process (Cheng, et al., 2019). However, auto-correlation and cross-correlation of industrial process data measurements are always dynamic and could be greatly affected by sampling frequency. The extraction of these features directly determines the generalization ability of the process monitoring model, but the impact of sampling frequency on process monitoring results is hardly investigated. Although dynamic PCA has certain advantages in extracting the dynamic features from process data, the selection of key lag order is generally limited to a low value, which will lead to the loss of auto-correlation features of time series data. With the rapid development of data acquisition technology, data of multiple sampling frequencies can be flexibly acquired from distributed control system (DCS). Data features under different sampling frequencies can be studied at different scales, which provides a basis for studying the impact of sampling frequency on process monitoring results.

The aim of this paper is to extract data features for process monitoring under the premise of considering the data dynamic relevance. When analyzing data with different sampling frequencies, it can be found that process dynamic has different effects on different variables. On this basis, an optimal sampling strategy is proposed to characterize auto-correlation. For the extraction of dynamic cross-correlation features, the statistics of Hotelling-T^2 (T^2) and squared prediction error (SPE) are obtained by PCA and DPCA methods. The alarm point and false alarm rate, obtained from the built statistics, are employed for comparison. The research is implemented on Tennessee Eastman Process (TEP). The results show that DPCA combined with the best sampling strategy could give an earlier alarming with a lower false alarm rate compared with other methods without the consideration of dynamic correlation. It can provide a new idea for the selection of sampling frequency on industrial time series data.

2. Methodology

In this section, the methods applied in this work are introduced.

2.1. The auto-correlation functions

Auto-correlation coefficient is an important index for time series analysis, which can reflect the dynamic characteristics of the process to a certain extent. It is assumed that data have been regularly sampled from a time series $\{X\}$, where $\{x_t, t=1,2,...,n\}$ is a set of observation samples from $\{X\}$. The mean of samples is given as

$$\overline{\mu} = \frac{1}{n}\sum_{t=1}^{n} x_t, \tag{1}$$

For the given samples $\{x_t\}$, the auto-covariance function $\widehat{\gamma_k}$ at the delay number k is given as

$$\widehat{\gamma_k} = \frac{1}{n-k}\sum_{t=1}^{n-k}\left(x_t - \widehat{\mu}\right)\left(x_{t+k} - \widehat{\mu}\right), 1 \le k < n, \tag{2}$$

And when k=0, the corresponding auto-covariance function can be simplified as

$$\widehat{\gamma_0} = \frac{1}{n-1}\sum_{t-1}^{n}\left(x_t - \widehat{\mu}\right)^2 \tag{3}$$

Based on Eq. (2) and (3), the auto-correlation function of the samples $\{x_t\}$ at the delay number k can be defined as follows,

$$\widehat{\rho_k} = \frac{\widehat{\gamma_k}}{\widehat{\gamma_0}}, 1 \le k < n, \tag{4}$$

2.2. Principal component analysis (PCA)

PCA, a classical technique for feature extraction and dimension reduction (MacGregor, et al., 1995), has been widely used in process monitoring with T^2 and SPE statistics. Given n observations of m measurement variables $X_{n\times m}$ after normalization, the covariance matrix of X can be calculated as follows,

$$Cov(X) = \frac{X^T X}{n-1} \tag{5}$$

Then singular value decomposition is employed to $Cov(X)$. The obtained eigenvalues and eigenvectors are sorted according to the size of the eigenvalues. On this basis, a score matrix T and a loading matrix P can be determined by retaining the first k features that contain the most information. The original matrix X can be decomposed as follows,

$$X = TP^T + E = t_1 p_1^T + \cdots + t_k p_k^T + E \tag{6}$$

where p, t are the loading vector and score vector, and E is the residual matrix. T^2 and SPE statistics can be calculated in the principal component space and residual space respectively and the corresponding control limits can be used for process monitoring. PCA has a good performance in the monitoring of stationary process, but the dynamic characteristics of the data cannot be extracted.

2.3. Dynamic principal component analysis (DPCA)

To extract the dynamic characteristics of the data, the DPCA was proposed with a stack of data matrixes as follows,

$$X = \begin{bmatrix} x_t^T & x_{t-1}^T & \cdots & x_{t-l}^T \\ x_{t-1}^T & x_{t-2}^T & \cdots & x_{t-l-1}^T \\ \vdots & \vdots & \ddots & \vdots \\ x_{t+l-n}^T & x_{t+l-n-1}^T & \cdots & x_{t-n}^T \end{bmatrix} \tag{7}$$

where x_t^T is the observation vector in the training set at time t. If enough lags are included in Eq. (7), the serial correlations can be considered in further PCA calculation. However, for high-dimensional data, a large lag order will lead to high computational loads for further matrix decomposition and the process dynamic has a different effect on different variables. Therefore, considering the limited lag order, it is necessary to select an optimal sampling frequency to consider the process dynamic better.

3. Selection of optimal sampling frequency and its application on Tennessee Eastman Process (TEP)

In this chapter, the implementation procedures of the proposed research on the selection of optimal sampling frequency are introduced with the case study of TEP.

3.1. Tennessee Eastman Process (TEP) and data description

TEP is a chemical process benchmark simulated by the Eastman Chemical Company (Downs and Vogel, 1993), which has been widely used by researchers in the field of process control, monitoring and fault diagnosis to validate their newly proposed methods. The revision version of TEP by university of Washington is applied in this work because data of different sampling frequencies and operating conditions can be easily simulated by Simulink (Bathelt, et al, 2015). 50 hours of data are simulated in normal data set and each fault data set. The fault in each fault data set is introduced in the 25th hour. The sampling frequency is set from 1 second to 3 minutes by down sampling. A total of 31 variables are selected in this work. The component variables are not considered here because such variables cannot be measured in real time in chemical operation.

3.2. Determination of the optimal sampling frequency in this work

In this work, two factors are considered to determine the best sampling frequency for chemical process monitoring. On the one hand, the dynamic characteristics of the process should be fully extracted. If the sampling frequency is too high, the dynamic characteristics of time series cannot be extracted by dynamic monitoring methods for the balance of the lag order and computational loads. On the other hand, sufficient process intermediate information has to be retained. If the sampling frequency is too low, the useful information in the process data will be lost and the fault information cannot be obtained in time. Therefore, there is an optimal sampling frequency that can be determined by striking a balance between these two aspects. Early fault detection time (FDT) and low false alarm rate (FAR) are of great significance for measuring the performance of process monitoring models. Therefore, these two factors are used as the standard to find the optimal sampling frequency in this study.

3.3. Selection of optimal sampling frequency in TEP

In this section, the procedures of this work are introduced. As mentioned before, the auto-correlation coefficient can be used to test the stationarity of time series and reflect the dynamic characteristics of data. Therefore, the auto-correlation diagrams are first applied to show the auto-correlation relationship of each variable in TEP, and then a sampling range can be initially obtained. However, it is difficult to directly determine the optimal sampling frequency, because the process dynamic characteristics have different effects on different variables. Different sampling frequencies within the range are selected and the data are brought into the PCA and DPCA models respectively to compare the monitoring results. The FDT and FAR are calculated to select the optimal sampling frequency.

4. Results and discussion

Auto-correlation diagrams of different types of variables in TEP are shown in Figure 1 and 2. It can be seen that most variables show a dynamic characteristic and the process dynamic has different effects on different variables. Therefore, the optimal sampling frequency cannot be directly determined, but it can be obtained from Figure 2 that the

auto-correlation order can be effectively reduced by reducing the sampling frequency. Next, data of sampling frequencies from 1 second to 3 minutes are respectively brought into the PCA and DPCA models to compare the monitoring results.

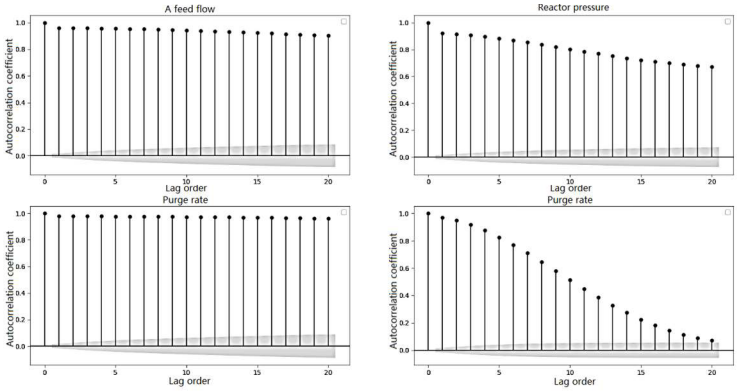


Figure 1 Autocorrelation diagrams of variables with a sampling frequency of 10s

Process monitoring results for fault 1 by PCA and DPCA are shown respectively in Figure 3 and 4. The results of FDT under different sampling frequencies are unified as second-level data. The sampling frequency is 1 second. It's obvious that DPCA has a better performance than PCA because the dynamic characteristics of the data are considered to a certain extent. In addition, the statistics fluctuate around the control limit at the early stage of the fault, making it difficult to determine whether it is a true alarm or not because the sampling frequency is too high. Monitoring results of fault 8 at different sampling frequencies are shown in Table 1. It can be concluded that fast sampling frequency is beneficial to improve FDT to a certain extent, but corresponds to high FAR. When the sampling frequency is at an appropriate value, the FAR is significantly reduced, and the fault can be also detected earlier than other sampling frequencies. Therefore, 60 seconds can be determined to be the optimal sampling frequency for TEP, which can be also validated by other faults in TEP. When the sampling frequency is further reduced, the useful feature information in the process data is lost, resulting in a late FDT. It is worth noting from Table 1 that the process dynamic characteristics will be removed if the sampling frequency is too late, which leads to similar monitoring results of PCA and DPCA.

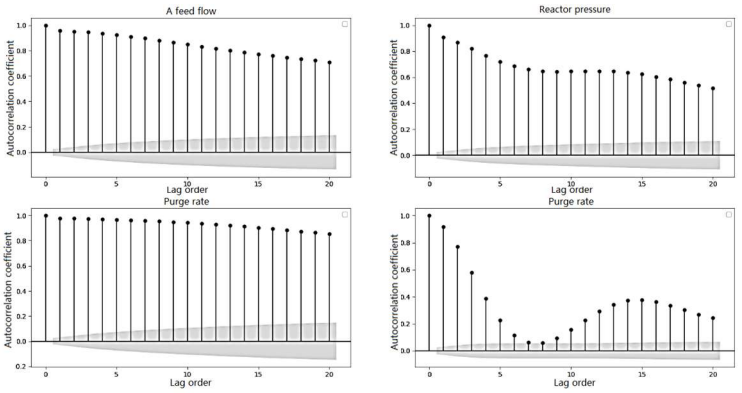


Figure 2 Auto-correlation diagrams of variables with a sampling frequency of 60s

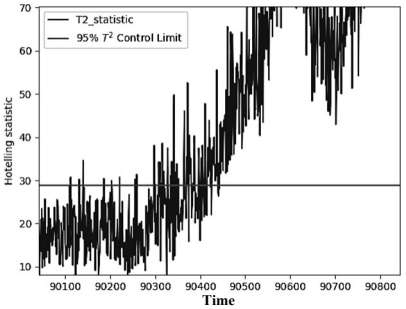


Figure 3 Monitoring result for fault 1 by PCA

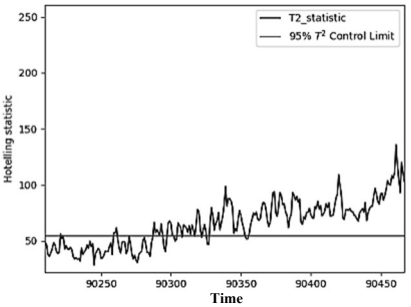


Figure 4 Monitoring result for fault 1 by DPCA

Table 1 Process monitoring results for fault 8 at different sampling frequencies

Sampling frequency	1s	10s	15s	20s	30s	40s	60s	90s	180s
FDT by PCA	94650	94720	94725	94680	94740	94680	94680	94500	94500
FAR by PCA	0.0128	0.011	0.0105	0.0102	0.0093	0.0138	0.0067	0.009	0.006
FDT by DPCA	94441	94460	94455	94460	94530	94520	**94380**	94500	94680
FAR by DPCA	0.0130	0.01	0.0097	0.0084	0.0067	0.0089	**0.0033**	0.006	0

5. Conclusions

In this work, auto-correlation of variables in chemical process at different sampling frequencies is analysed by the auto-correlation diagrams. Considering the dynamic correlations of process, different sampling frequencies are used to establish process monitoring models by DPCA. A strategy for the selection of optimal sampling frequency is then proposed by comparing FDR and FAR, which are two important indicators of process monitoring. Through the case study on TEP, it can be concluded that there is an optimal sampling frequency corresponding to the best process monitoring results. Compared with other studies focusing on feature extraction, the impact of data sampling rate on process monitoring performance is discussed in detail in this work, which can provide a reference for initial data acquisition in industrial practice.

References

A. Bathelt, N. L. Ricker, and M. Jelali, 2015, Revision of the Tennessee Eastman Process Model, IFAC-PapersOnLine, 48, 8, 309-314.

F. F. Cheng, Q. P. He, and J. S. Zhao, 2019, A novel process monitoring approach based on variational recurrent autoencoder, Computers & Chemical Engineering, 129, 4, 106515-106546.

J. J. Downs and E. F. Vogel, 1993, A plant-wide industrial process control problem, Computers & chemical engineering, 17, 3, 245-255.

J. F. MacGregor and T. Kourtl, 1995, Statistical process control of multivariate processes, Control Engineering Practice, 3, 3, 403-414.

W. Ku, R. H. Storer, and C. Georgakis, 1995, Disturbance detection and isolation by dynamic principal component analysis, Chemometrics and Intelligent Laboratory Systems, 30, 1 , 179-196.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50240-2

Autoregressive Distributed Lag Model Based Cointegration Analysis for Batch Process Monitoring

Jiatao WEN, Cheng Ji, Jingde Wang*, Wei Sun*

College of Chemical Engineering, Beijing University of Chemical Technology, North Third Ring Road 15, Chaoyang District, Beijing, 100029, China
sunwei@mail.buct.edu.cn

Abstract

Batch processes have been widely applied in pharmaceutical and chemical industry. Variables in batch process exhibit obvious nonstationary and nonlinear characteristics, which brings challenges to process monitoring. Recently, cointegration theory gets more attention for its applications on the analysis of multivariate nonstationary time series. According to the cointegration theory, if the nonstationary random sequences in a system containing a cointegration relationship, there is at least one stable long-term dynamic equilibrium relationship among these nonstationary variables. Such dynamic equilibrium relationship can also be found in variables in industrial processes, since the nonstationary variables are controlled by physics, chemistry, and other internal mechanisms within a system. For multivariate variables system, Johansen test is a commonly used method to test cointegration relationship and estimate the cointegration vectors, which is based on the multivariate unconstrained vector autoregressive (VAR) model, in which all variables are assumed as indifferent endogenous variables. However, certain variables in real process, such as control variables, is not affected by such long-term equilibrium and is governed by external conditions, which is called a weakly exogenous variable. When there are weak exogenous variables in a system, the cointegration test based on the VAR model needs to be improved, as the impact of exogenous variable is neglected. In this work, considering the impact of weak exogenous variables, autoregressive distributed lag (ADL) model is adopted for cointegration test and parameters estimation. The penicillin fermentation process is presented to illustrate the effectiveness of the proposed method, in which many control parameters exert a significant impact on the state of fermentation. The monitoring results show that the interaction among variables can be better characterized, and abnormal behavior of the process can be correctly detected by proposed cointegration testing method based on autoregressive distributed lag model.

Keywords: Cointegration analysis; Exogenous variable; Batch process; Dynamic equilibrium relationship.

1. Introduction

Batch and semi-batch processes are widely applied in modern industries due to its flexibility. However, product quality can be easily affected by environment conditions, equipment conditions, and raw material quality, which may lead to the waste of raw materials or cause an accident. Thus, it is essential to monitor the batch process to avoid quality loss and potential accidents. However, batch process data present complex

characteristics, which brings challenge to process monitoring. Many approaches have been proposed to monitor batch process in the view of complex data characteristics.

To consider the time series nature of multivariable data, VAR models have been applied to batch process monitoring. Choi et al. (2008) first used VAR models to remove the auto and cross correlation within batch data, and then principal component analysis (PCA) was applied to the residuals obtained from VAR model. Based on their work, Filho and Valk (2020) proposed a new monitoring scheme by estimating coefficients of VAR model, which was proved to hold more information on the variable dynamics than the residual-based method.

However, the VAR model is more appropriate for stationary data, while the variables in batch process exhibits obvious nonstationary characteristics. In recent years, the cointegration theory in time series analysis proposed by Engle and Granger (1987), which is applied to analyze multivariate nonstationary time series, has been introduced to the monitoring of nonstationary processes. Chen et al. (2009) applied the cointegration theory to monitor industrial distillation units and established a monitoring model by extracting the most significant cointegration relationship among variables. To improve the effect of nonstationary process monitoring, Lin et al. (2017) adopted the common trend model for nonstationary process monitoring by eliminating the influence of nonstationary common factors, so that the classic multivariate statistics method is applicable to non-stationary process monitoring. However, cointegration test and cointegration vector estimation are always complemented by Johansen test, which is based on the multivariate unconstrained VAR model where each variable is assumed as undifferentiated endogenous variables. Since weak exogenous variables in a system do not respond to long-term equilibrium, it is not appropriate to estimate the cointegration vector based on VAR model. Pesaran (1995) examined the long-term run relationship among variables based on ADL model, which indicated that ADL cointegration test is proper for a system that contains exogenous variables.

To address the limitation of previous methods, a new batch process monitoring strategy that based on ADL cointegration test is proposed. Considering the impact of weak exogenous variables, ADL cointegration test is first adopted on historical batches to determine if long term equilibrium relationship exists among variables and estimate the cointegration parameters by ordinary least squares. After that, the deviation of the stable relationship among variables of each historical batch can be obtained, which obeys the specific trajectory. Thus, online monitoring can be implemented by judging if the value of deviation is out of the specified range of the trajectory. The monitoring result of the penicillin fermentation process is presented to illustrate the effectiveness of proposed batch process monitoring strategy, which shows that abnormal behavior of the process can be detected accurately with better characterization of the interaction among variables.

2. ADL model based cointegration analysis

According to the cointegration theory, if nonstationary variables in a system contain a common stochastic trend, a stable long-term dynamic equilibrium relationship exists among these variables. The basic idea is that the common random stochastic trend can be eliminated by linear combination of these variables if multiple nonstationary sequences have a common stochastic trend. But an exogenous variable is not affected by the long-term equilibrium deviation in a cointegration system, while an endogenous variable, which is also known as dependent variable, is determined by the system.

For a system with an endogenous variable y_t and two exogenous variables z_{1t} and z_{2t}, the ADL model can be established as follow:

$$\Delta y_t = \alpha_1(y_{t-1} - \beta_1 z_{1t-1} - \beta_2 z_{2t-1}) + \beta_3 \Delta z_{1t} + \beta_4 \Delta z_{2t} + u_t \quad (1)$$

And the long-term equilibrium relation can be represented as

$$\xi_t = y_t - \beta_1 z_{1t} - \beta_2 z_{2t} \quad (2)$$

In equation (2), ξ_t is the equilibrium error sequence which represents the series where the common linear trend is eliminated, and it is assumed that $B = (1, -\beta_1, -\beta_2)$ is the cointegration vector.
To estimate the cointegration parameters, equation (1) is transformed into the following form:

$$\Delta y_t = \alpha_1 y_{t-1} + b_1 z_{1t-1} + b_2 z_{2t-1} + \beta_3 \Delta z_{1t} + \beta_4 \Delta z_{2t} + u_t \quad (3)$$

in which α_1, b_1, b_2, β_3, β_4 are model coefficients which can be estimated by ordinary least squares, and u_t denotes model residual. A common way to test whether there is a cointegration relationship among variables is to test the null hypothesis $\alpha_1 = 0$ with t statistics. If $\alpha_1 = 0$, there is no long-term equilibrium relationship among variables, else if $\alpha_1 \neq 0$, cointegration parameters β_1 and β_2 can be obtained by $\beta_1 = -\frac{b_1}{\alpha_1}$, $\beta_2 = -\frac{b_2}{\alpha_1}$.

3. ADL cointegration test based batch process monitoring approach

CA is an effective way to analyse nonstationary variables with common trend. For nonstationary process monitoring, monitoring index is required to indicate whether current status is abnormal. Considering the existence of endogenous variables, a monitoring strategy for batch process is proposed based on ADL cointegration test. The diagram of proposed method is shown in Figure 1.

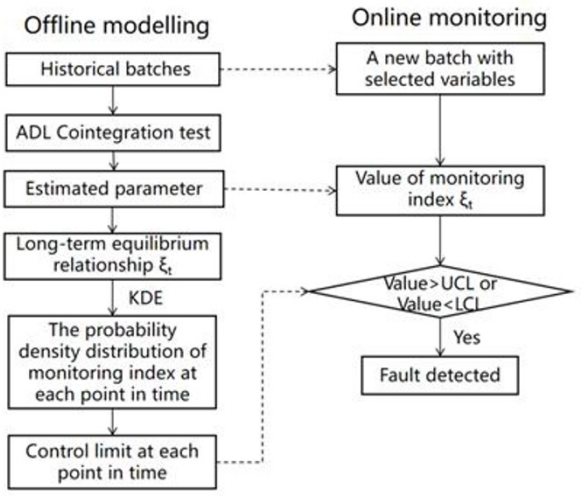


Figure 1 The diagram of proposed monitoring method

Several normal batches containing different initial conditions are labelled as training data. All exogenous variables and a key endogenous variable are selected as model variables. ADL cointegration test is applied to the training batch to determine if cointegration relationship exist among variables. Then the parameters of ADL model are estimated by ordinary least squares (OLS), and the long-term equilibrium error ξ_t which is assumed as monitoring index can be obtained as shown in equation (2). For normal batches, ξ_t present similar trajectory over time, which means that the value of ξ_t is in a certain range at every point of time. When the fault occurs, the original dynamic equilibrium relationship among variables is violated, thus ξ_t will deviate from its original range of value. Kernel density estimation is applied to training batches to estimate the probability density distribution of monitoring index at each point in time, and the upper control limits (UCL) and lower control limits (LCL) are obtained.

For a new batch which is regarded as testing data, monitoring index is calculated with parameters obtained from the training data. For each point of time t, it is assumed as normal status if the value of ξ_t is within the range of lower control limits and under upper control limits, while it is considered as abnormal condition if the value of monitoring index is over upper control limits or under lower control limits with a confidence coefficient of 99%.

4. Case study on penicillin process

4.1. Process description and data preparation

Penicillin fermentation process is a typical batch process. The initial stage of fermentation is intermittent operation, in which many cells are produced. After the glucose substrate is consumed by the cells, it comes into semi-intermittent feeding stage where glucose is continuously added into the fermentation cylinder. During fermentation, the concentration of penicillin will be affected by environment change, initial condition adjustment, and variable fluctuation. It is difficult to measure penicillin concentration online, thus real-time monitoring is required to ensure that the fermentation process runs under normal trajectory. In this work, Pensim software developed by Birol, Ündey, and Cinar (2002) is adopted to simulate the process.

4.2. Batch process monitoring based on proposed method

Ten batches of data under normal condition with different initial conditions are simulated as training data. The information of process variables is listed in Table 1. The reaction time ranges from 395 h to 405 h, and the sampling frequency is 0.1 h. Four exogenous variables and a key endogenous variable are selected as modelling variables that is shown in Table 1, of which ventilation rate is an endogenous variable and the others are exogenous variables.

Table 1 Variables in penicillin fermentation process

Variables	Units
Ventilation rate	L/h
Agitator power	W
Substrate feeding rate	L/h
Substrate feeding temperature	K
Penicillin concentration	g/L

Table 2 Faults information introduced in this work

Fault number	Fault type
1	Step decrease in aeration rate
2	Ramp decrease in agitator power
3	Ramp decrease substrate feed rate

To illustrate the effectiveness of the proposed monitoring strategy, three common faults are set on Pensim simulation platform. The faults are introduced to three different batches separately from 100 h to the end and the information of the faults is available in Table 2. At the same time, a normal batch data is supplied to test the monitoring effect on normal condition of the proposed method as comparison.

4.3. Monitoring result analysis and discussion

Three fault batches and a normal batch are set as testing data to verify the effectiveness of the proposed mothed.

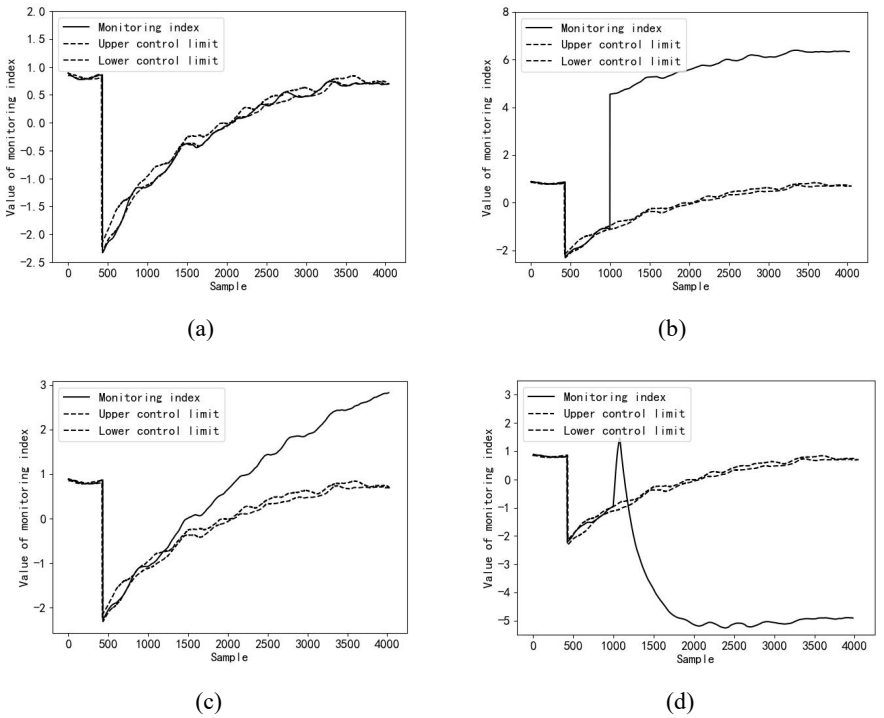


Figure 2 Monitoring results

The monitoring results are shown in Figure 2. For the normal batch as shown in (a), the value of monitoring index remains within the control limits, indicating that the data are under normal conditions and the model has a low false alarm rate that barely exceeds the control limits. For step fault monitoring result as shown in (b), the value of monitoring index exceeds the upper control limit at the first time when fault occurs, which shows that the abnormal condition can be detected in time by applying the proposed monitoring method. For slope faults as shown in (c) and (d), there is a delay to

detect the fault 2 while the fault 3 is detected timely, which indicates that the sensitivities to different fault are distinct. The results show that the proposed method can effectively extract the complex dynamic and nonstationary characteristics in the batch process and monitor the faults in real time.

5. Conclusion

In this work, a new monitoring strategy based on ADL cointegration analysis for batch process is proposed. Taking the complex characteristics of batch process data into consideration, cointegration analysis is applied to the data to eliminate the common stochastic trend and obtain long-term equilibrium relationship of nonstationary variables. In the view of the exogenous variables in process which is not affected by other endogenous variables as well as the equilibrium relationships, ADL cointegration test is used to test if long term equilibrium relationship exists among nonstationary variables, and the model parameters is estimated by ordinary least squares. By considering the impact of exogenous variables, the interaction among variables is better described. On this basis, real-time monitoring of the fermentation process is implemented by monitoring the trajectory of the equilibrium error. In the end, process monitoring results on a benchmark penicillin fermentation data indicate the effectiveness of proposed monitoring strategy.

Acknowledgements

The authors gratefully acknowledge the team of Professor Cinar (Illinois Institute of Technology) for providing the source program of Pensim software which is applied to generate the experimental data in this research.

References

G. Birol, C. Ündey, and A. Cinar. "A modular simulation package for fed-batch fermentation: penicillin production." Computers & Chemical Engineering 26.11(2002):1553-1565.

Q. Chen, U. Kruger, and A. Leung. "Cointegration Testing Method for Monitoring Nonstationary Processes." Industrial & Engineering Chemistry Research 48.7(2009):3533-3543.

R. F. Engle, and C. Granger. "Cointegration And Error-Correction: Representation, Estimation and Testing." Econometrica 55.2(1987):251-276.

D. M. Filho, and M. Valk. "Dynamic VAR model-based control charts for batch process monitoring." European Journal of Operational Research 285(2020).

Y. Lin, U. Kruger, and Q. Chen. "Monitoring Nonstationary Dynamic Systems Using Cointegration and Common Trends Analysis." Industrial & Engineering Chemistry Research (2017): acs. iecr.7b00011.

M. H. Pesaran, "An autoregressive distributed lag modelling approach to cointegration analysis." Cambridge Working Papers in Economics (1995):134--150.

W. C. Sang, J. Morris, and I. B. Lee. "Dynamic model-based batch process monitoring." Chemical Engineering Science 63.3(2008):622-636.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50241-4

A Data-Driven Fault Detection and Diagnosis by NSGAII-t-SNE and Clustering Methods in the Chemical Process Industry

Nahid Raeisi Ardali[a], Reza Zarghami[a*], Rahmat Sotudeh Gharebagh[a], and Navid Mostoufi[a]

[a] *Process Design and Simulation Research Centre, School of Chemical Engineering, College of Engineering, University of Tehran, P.O. Box 11155-4563, Tehran, Iran. rzarghami@ut.ac.ir*

Abstract

This study aims to illustrate a novel unsupervised learning method for fault detection and diagnosis of chemical processes. The data-driven fault detection and diagnosis contains two main steps. a) data preparation and feature selections as preprocessing step and b) fault detection and diagnosis for fault indication. In this study, a non-dominated sorting genetic algorithm (NSGAII) was utilized for selecting the most relevant variables from the measured variables for each fault. The t-distributed stochastic neighbor embedding (t-SNE) algorithm was used for information extraction from the selected variables which will lead to visualization of extracted features. Meanwhile, fault detection was performed by k-means and density-based spatial clustering of applications with noise (DBSCAN) clustering algorithms. The Tennessee Eastman benchmark process and faults of process variation were applied to validate the proposed method. Results show that the combination of NSGAII, t-SNE, and clustering methods is an efficient method for Tennessee Eastman process fault detection and diagnosis. And the proposed method could be used in chemical processes for early fault detection.

Keywords: Process Monitoring; Fault Detection; Tennessee Eastman Process; NSGAII; t-SNE.

1. Introduction

Early fault detection and diagnosis play a critical role in managing abnormal circumstances in dynamic processes (Md Nor et al., 2020). Fault detection and diagnosis methods are mainly divided into three categories as quantitative, qualitative, and data-driven. Quantitative models need an accurate process model based on system governing equations such as momentum, energy, and mass conservation equations, chemical kinetics, and thermodynamic equations. In such models, fault detection and diagnosis are based on the residuals of real process data and the predicted data at normal conditions (Elhsoumi et al, 2017). Qualitative models are based on the causal models and the knowledge of experts about the process such as fault trees (Venkatasubramanian et al., 2003). Due to the complexities and difficulties in obtaining variables needed for quantitative and qualitative modeling, these models could not be practically applied (Ziaei-Halimejani et al., 2021). Meanwhile, by the availability of large volumes of data, data-driven methods for fault detection and diagnosis (supervised and unsupervised) have been considered by many researchers. In industrial plants, historical data are mostly available without labels. In start-up conditions, faults could not be detected clearly, so

unsupervised data-driven methods are more suitable to be applied. Therefore, the purpose of this study is to develop an unsupervised learning method for fault detection and diagnosis of chemical processes. In this paper, a non-dominated sorting genetic algorithm (NSGAII) was used for feature selection and the t-SNE algorithm was used for feature extraction and visualization. Then fault detection and the diagnosis were done by clustering algorithms.

2. Theory

2.1 Data Pre-processing

Basically, a good feature in data-driven models should be informative, discriminative, and independent. Meanwhile, by selecting appropriate features the computation time and overfitting error will be reduced. The NSGAII algorithm was introduced in 2002 by Deb et al. and it is a multi-objective genetic algorithm for non-dominated sorting that includes crossover, mutation, merge, non-dominated sorting, crowding distance calculation, and population sorting functions as shown in Figure 1. Therefore, in this paper NSGAII algorithm was utilized for multi-objective feature selection. And classification error and the number of features were determined as objectives that should be minimized. But it is worth noting that both of these parameters could not be minimized simultaneously, and by decreasing the number of features the classification error will increase. Thus a trade-off between these parameters should be done.

After feature selection, t-distributed stochastic neighbor embedding (t-SNE) was utilized for feature extraction. t-SNE is a nonlinear statistics method for data dimension reduction to two or three dimensions and it is an efficient method for data visualization that was proposed by van der Maaten (2008). The t-SNE algorithm consists of two main steps. First, data transformation from high dimension space to a low dimensional space by probability distribution calculation. Somehow similar points have a high probability and dissimilar points have a low probability of being selected. Then, in low dimension space, a possible distribution is created and the Kullback–Leibler divergence is minimized between the two distributions according to the position of the points.

$$P_{j|i} = \frac{\exp(-\|X_i - X_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|X_i - X_k\|^2 / 2\sigma_i^2)} \tag{1}$$

$$P_{ij} = \frac{P_{j|i} + P_{i|j}}{2\,N} \tag{2}$$

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq i} (1 + \|y_i - y_j\|^2)^{-1}} \tag{3}$$

$$KL(P\|Q) = \sum_{i \neq j} P_{ij} \log \frac{P_{ij}}{q_{ij}} \tag{4}$$

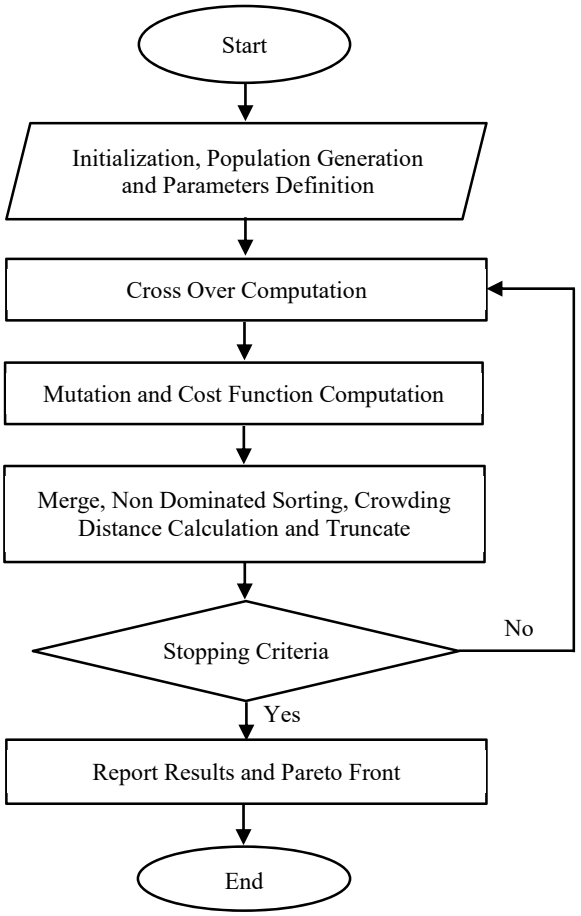


Figure 1: Flowchart of NSGAII Algorithm

2.2 Clustering

Data clustering methods are categorized in unsupervised learning methods whereas these methods don't need datasets labels and training phase. In this paper, data clustering was performed by using k-Means and DBSCAN algorithms. In the k-Means method, cluster centres are chosen based on the predetermined cluster numbers. And cluster members will be assigned to the nearest cluster, then in each iteration, the cluster centres and cluster members will be updated. DBSCAN method is a density-based method that cloud de utilized under noisy conditions. Epsilon and min points number are two main parameters that should be adjusted for each problem. In the DBSCAN Method data points will be categorized as cluster members, cluster borders, and noise depending on their spatial density. The main difference between the k-Means and the DBSCAN algorithm is that DBSCAN could be utilized in nonconvex geometries (Ziaei-Halimejani et al., 2021).

3. Results and discussion

The case study of this paper is the Tennessee Eastman benchmark process. In this process, four main reactants A, C, D, and E react in the gas phase in the presence of B as inert, to produce G and H main products and F as a by-product. The unit operations of the Tennessee Eastman process are a reactor, a condenser, a gas-liquid separator, a stripper, and a compressor. The Tennessee Eastman process contains 12 manipulated variables and 41 measurement variables (Downs and Vogel, 1993). In this paper, 19 flow component analyses were not considered. Tennessee Eastman Process Simulink model (https://depts.washington.edu/control/ LARRY/TE/download.html) is used for data generation, there are 20 faulty conditions and one normal condition. This paper included conditions 0, 1, 2, 4, 6, 7, 13, and 14, as normal, 5 step faults,1 slow drift, and 1 sticking valve fault. The datasets contain 480 samples for each condition.

3.1. Data normalization

Measured data in the chemical processes have different operating ranges, hence by using Eq.5 and dividing by their standard deviation the measured data have been normalized. The normalized data are between -1 and 1.

$$\hat{X}_s^{fn} = \frac{X_s^{fn} - \mu}{\sigma} \tag{5}$$

3.2. Feature selection

As mentioned in section 2, the NSGAII algorithm was utilized for feature selection. The parameters of the algorithm were adjusted as presented in Table 1. And for the 8 conditions dataset, the optimum number of selected features and classification errors is shown in Figure 2. As illustrated, 8 optimum solutions were provided with different numbers of features and classification errors. Therefore, 9 cases for all features and 8 NSGAII, N_F were tested (Table 2).

Table 1: NSGAII Algorithm Parameters

Maximum Iterations	Number of Populations	Crossover	Mutation	Mu
50	20	0.7	0.4	0.1

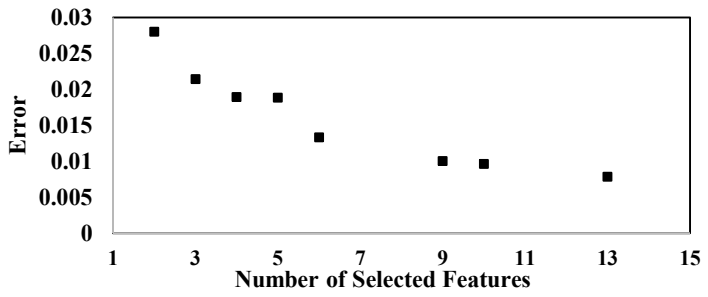


Figure 2: Error of feature selection using NSGAII algorithm versus number of selected features

3.3. Feature extraction

After the feature selection step, t-SNE as a nonlinear dimension reduction method was utilized to extract three features from selected features and feature visualization as shown in Figure 3.a. In Figure 3.b Feature extraction results of the PCA method are shown.

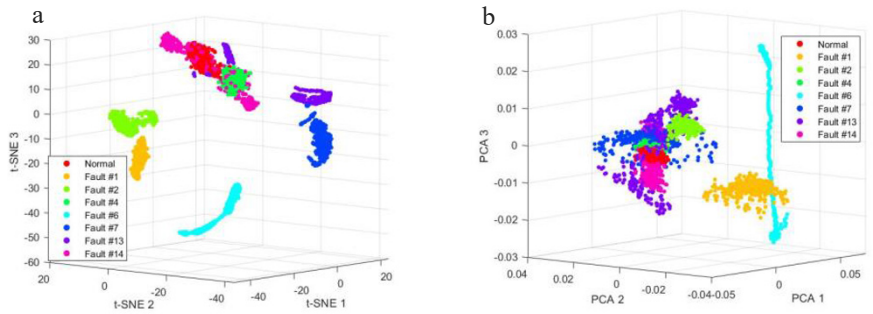


Figure 3: Feature Extraction, a) t-SNE method and b) PCA method

3.4. Fault detection and diagnosis

Fault detection and diagnosis have profited from clustering algorithms. In this research k-Means and DBSCAN methods were utilized. In Figures 4.a and 4.b the results of test 5 are illustrated.

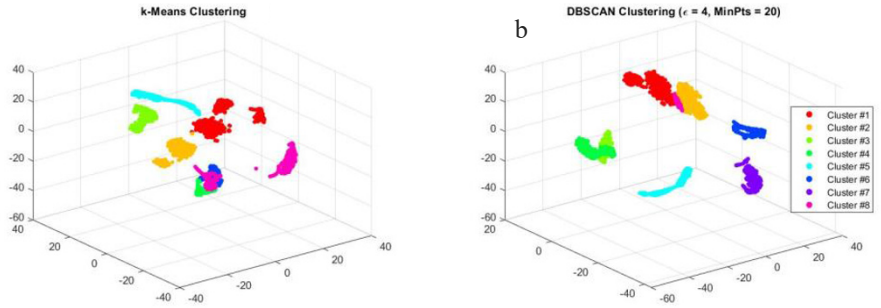


Figure 4: Fault Detection and Diagnosis, a) k-Means method and b) DBSCAN method

3.5. Evaluation

To validate the results of clustering methods, four indexes including adjusted rand index (ARI), adjusted mutual information (AMI), V-Measure, and accuracy have been used. The evaluation results are reported in Table 2. As shown, in all tests, the t-SNE method obtained better results than the PCA method. In test numbers 8 and 9, the t-SNE-k-Means provided better results and in the rest of the tests, DBSCAN results are more accurate. It is worth noting that test number 5 was the best test. In this case, 6 Features included XMEAS (21), XMV (3), XMV (4), XMV (5), XMV (6), and XMV (10) were selected. The classification error and R_F=N_F/N were computed as 0.0133 and 0.18, respectively. Also as shown in Table 2, the t-SNE-clustering performed better in cases with a lower number of features compared with PCA-clustering.

Table 2: Clustering Results Evaluation

	N_F	Extraction	k-Means				DBSCAN			
			ARI	AMI	V-Meas.	Acc.	ARI	AMI	V-Meas.	Acc.
1	33	t-SNE	0.35	0.51	0.64	0.45	**0.39**	**0.54**	**0.66**	**0.52**
		PCA	0.28	0.44	0.50	0.44	0.27	0.45	0.57	0.44
2	13	t-SNE	0.54	0.73	0.78	0.57	**0.64**	**0.8**	**0.84**	**0.73**
		PCA	0.35	0.54	0.66	0.50	0.35	0.56	0.65	0.49
3	10	t-SNE	0.39	0.52	0.67	0.50	**0.59**	**0.74**	**0.78**	**0.69**
		PCA	0.42	0.58	0.63	0.55	0.38	0.56	0.65	0.51
4	9	t-SNE	0.52	0.70	0.79	0.54	**0.61**	**0.75**	**0.83**	***0.96***
		PCA	0.38	0.55	0.69	0.52	0.34	0.48	0.59	0.48
5	**6**	**t-SNE**	0.64	0.76	0.81	0.69	***0.79***	***0.87***	***0.89***	**0.83**
		PCA	0.37	0.57	0.68	0.55	0.28	0.52	0.67	0.47
6	5	t-SNE	0.45	0.66	0.75	0.52	**0.74**	**0.82**	**0.85**	**0.82**
		PCA	0.30	0.56	0.67	0.48	0.29	0.53	0.64	0.47
7	4	t-SNE	0.58	0.70	0.77	0.69	**0.67**	**0.78**	**0.82**	**0.75**
		PCA	0.34	0.56	0.69	0.52	0.29	0.50	0.65	0.48
8	3	t-SNE	**0.62**	**0.72**	**0.76**	**0.74**	0.56	0.71	0.75	0.68
		PCA	0.35	0.55	0.64	0.51	0.35	0.56	0.63	0.48
9	2	t-SNE	**0.67**	**0.76**	**0.77**	**0.82**	0.65	0.75	0.77	0.76
		PCA	0.38	0.49	0.56	0.52	0.32	0.51	0.60	0.51

4. Conclusion

In this research, a method based on NSGAII, t-SNE, and unsupervised learning algorithms was developed for fault detection and diagnosis in chemical processes. Informative features were selected by the NSGAII algorithm and by utilizing 18 % of features, 8 conditions containing normal and 7 faults were detected and diagnosed. Results showed that the proposed method has high applicability to detect and diagnose Tennessee Eastman process faults, and could be applied for other chemical processes.

References

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T., A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II, IEEE Transactions on Evolutionary Computation, 6, 2 (2002). DOI: 10.1109/4235.996017.

Downs, J. J. and Vogel, E. F., A Plant-Wide Industrial Process Control Problem, Computers & Chemical Engineering, 17, 245-255 (1993). DOI: https://doi.org/10.1016/0098-1354(93)80018-I.

Elhsoumi, A., El Harabi, R., Hadj Ali Naoui, S. B., Abdelkrim, M. N., Luenberger Observer based Sensor and Actuator Fault Detection for Chemical Reactor, IFAC Proceedings, 42, 9, 566-571 (2009). DOI: https://doi.org/10.3182/20090921-3-TR-3005.00097.

Md Nor, N., Rosmani, C., and Azlan Hussain, M.,A review of data-driven fault detection and diagnosis methods: applications in chemical process systems, Reviews in Chemical Engineering, 36, 4, 513-553 (2020). DOI: https://doi.org/10.1515/revce-2017-0069.

van der Maaten L.J.P., and Hinton G.E., Visualizing High-Dimensional Data Using t-SNE, Journal of Machine Learning Research, 9, 2579-2605 (2008).

Venkatasubramanian, V., Rengaswamy, R., Yin, K.,. Kavuri, S. N, A review of process fault detection and diagnosis, Part I: Quantitative model-based methods, Computers and Chemical Engineering, 27 , 293- 311 (2003). DOI: https://doi.org/10.1016/S0098-1354(02)00160-6.

Ziaei-Halimejani, H., Zarghami, R., Mansouri, S., and Mostoufi, N., Data-Driven Fault Diagnosis of Chemical Processes Based on Recurrence Plots, Industrial & Engineering Chemistry Research, 60, 3038-3050 (2021). DOI: https://doi.org/10.1021/acs.iecr.0c06307.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50242-6

Cyberattack Detectability-Based Controller Screening: Application to a Nonlinear Process

Shilpa Narasimhan[a], Nael H. El-Farra[a], Matthew J. Ellis[a*]

[a]*Department of Chemical Engineering, University of California Davis, Davis, CA-95616, USA*
mjellis@ucdavis.edu

Abstract

In this work, multiplicative cyberattacks targeting the sensor-controller communication link of a process control system are considered. The interdependence of detectability of an attack with respect to a general class of residual-based detection schemes and the control parameters is characterized. Exploiting this dependence, a controller screening methodology that may be used to incorporate cyberattack detectability into the standard controller design criteria is presented. Using a chemical process example, the application of the controller design screening to a nonlinear process is demonstrated.

Keywords: Multiplicative cyberattack, zero-alarm attack, controller design, cyberattack detectability

1. Introduction

Increased reliance on networked communication has rendered process control systems (PCSs) vulnerable to increasingly complex and frequent cyberattacks (Miller et al., 2018) in the past couple of decades. This has motivated an increasing body of research dedicated to the development of controller-based approaches to attack resilient controller design, cyberattack detection, identification, and mitigation schemes (Giraldo et al., 2018, and Tan et al., 2020 and references therein, Oyama et al., 2020, and Chen et al., 2021). Broadly, cyberattacks may target the PCS by compromising the data integrity of process data in the communication links or by altering the PCS logic. In this work, multiplicative zero-alarm cyberattacks compromising the data integrity of the sensor-controller communication link are considered. These attacks are modeled by a factor multiplied to the measured variable and are particularly threatening as they may be designed to evade detection by keeping the alarms in the detection scheme monitoring the process at zero, with minimal process knowledge.

The detectability of the measured states of a process may be viewed as a systems-theoretic property. In practice, the detectability of a cyberattack on the measured states of a process is dependent on the control parameters (e.g., controller gain and observer gain), and the detection scheme monitoring the process. This interdependence of the detectability of an attack and the controller design has not received much attention in the literature. In a previous work (Narasimhan et al., 2021), an approach to characterizing the interdependence of the attack detectability in terms of terminal set of residuals (a small set containing the origin within which the residual is ultimately bounded) for the attacked and the attack-free process was presented. Based on the characterization, a controller screening methodology that may be used identify and discard control parameters that mask an attack was presented for processes modeled by discrete-time linear time invariant

(LTI) dynamics. This screening may be used to incorporate cyberattack detectability into existing controller design criteria (e.g., closed loop-stability and robustness to uncertainty). The present work explores the application of the controller screening methodology to a nonlinear chemical process. To make the paper self-contained, an abridged version of the screening methodology is presented herein. Interested readers are directed to Narasimhan et al. (2021) for more information.

2. Controller Screening Methodology

Processes modeled by discrete-time linear time invariant (LTI) dynamics, and subject to bounded measurement noise and process disturbances are considered:

$$\begin{aligned} x(t+1) &= Ax(t) + B\,u(t) + Gw(t) \\ y(t) &= \Lambda(Cx(t) + v(t)) \end{aligned} \tag{1}$$

where $x(t) \in \mathbb{R}^n$ is the state of the process, $u(t) \in \mathbb{R}^{n_u}$ is the control input, $y(t) \in \mathbb{R}^m$ is the output from the process, $w(t) \in W \subset \mathbb{R}^{n_w}$ and $v(t) \in V \subset \mathbb{R}^m$ are the bounded process disturbances and measurement noise. The sets W and V are compact and contain the origin in their interior. Multiplicative sensor-controller link attacks are modeled by the diagonal matrix $\Lambda = \mathrm{diag}(\alpha_1, \alpha_2, \ldots, \alpha_m)$ where $\alpha_i \neq 1$ represents attack on the i^{th} sensor-controller communication link. To generate an estimate of the state (denoted by $\hat{x}(t) \in \mathbb{R}^n$), a Luenberger observer with gain L is used. A linear controller with gain K is used to steer the state to the origin. The eigenvalues of $A - LC$ and $A - BK$ are assumed to lie within the unit circle. For analysis, the augmented state vector is defined as $\xi(t) = [x^T(t) \quad e^T(t)]^T$, where $e(t) = x(t) - \hat{x}(t)$ is the estimation error (the dynamic model is defined in Eq. (5) of Narasimhan et al., 2021). The dynamics of the augmented state is $\xi(t+1) = A_\xi(\Lambda, K, L)\xi(t) + B_\xi(\Lambda, L)d(t)$, where $d(t) = [w^T(t) \quad v^T(t)]^T \in F := \{\binom{w}{v} \mid w \in W, v \in V\}$ ($A_\xi(\Lambda, K, L)$ and $B_\xi(\Lambda, K, L)$ are given in Eq. (6) of Narasimhan et al., 2021).

Due to the presence of bounded noise in the measurement, and process disturbances, when the process is at steady state, its augmented state converges to a small set containing the origin, which is the minimum invariant set. When the closed-loop process is stable with $\max_i |\lambda_i(A_\xi(\Lambda, K, L)| < 1$ ($\lambda_i\left(A_\xi(\Lambda, K, L)\right)$ is the i^{th} eigenvalue of $A_\xi(\Lambda, K, L)$), the minimum invariant set of the process depends upon the attack magnitude and the control parameters. It is given by the infinite Minkowski sum $D_\xi(\Lambda, K, L) = B_\xi F \oplus A_\xi B_\xi F \oplus A_\xi^2 B_\xi F \oplus \ldots$ (Kuntsevich et al., 1996). The residual vector is defined as $r(t) = y(t) - \hat{y}(t)$. Writing the residual in terms of the augmented state and the disturbance gives $r(t) = [(\Lambda - I)C \quad C]\xi(t) + [0 \quad \Lambda]d(t) = A_r(\Lambda)\xi(t) + B_r(\Lambda)d(t)$. Thus, the terminal set of residuals for the process may be computed as $D_r(\Lambda, K, L) = A_r(\Lambda)D_\xi(\Lambda, K, L) \oplus B_r(\Lambda)F$.

A general class of residual-based detection schemes using the 2-norm of the residual vector as the detection metric is considered. Elaborate discussion on these detection schemes may be found in Section 2.3-2.4 of Narasimhan et al., 2021. With respect to the class of detection schemes considered, a closed-loop stability preserving attack of magnitude $\Lambda \neq I$ is defined as undetectable if the radius of the 2-norm ball enclosing the terminal residual set of the process under an attack is less than or equal to the radius of the 2-norm ball enclosing the terminal residual set of the attack-free process, i.e, $R(\Lambda, K, L) \leq R(I, K, L)$, where $R(\Lambda, K, L) := \max_{r' \in D_r(\Lambda, K, L)} ||r'||$. Similarly, any closed-loop

stability preserving attack of magnitude $\Lambda \neq I$ on the process is defined as a potentially detectable attack if the radius of the 2-norm ball enclosing the terminal set of the attack-free process is less than the radius of the 2-norm ball enclosing the terminal set of the process under an attack, i.e., $R(I,K,L) < R(\Lambda,K,L)$. Finally, an attack of magnitude $\Lambda \neq I$ is said to be detectable if it renders the closed-loop system in Eq.(1) unstable. Interested readers are directed to Section 3.1 of Narasimhan et al., 2021, for more discussions.

To derive numerically verifiable conditions for characterizing attack detectability, invariant outer polytopic approximation of the minimum invariant set of the augmented states of the stable process is computed as $D_\xi^{est}(\Lambda,K,L)$ (see Raković et al., 2005 for method), such that $D_\xi(\Lambda,K,L) \subseteq D_\xi^{est}(\Lambda,K,L) \subseteq D_\xi(\Lambda,K,L) \oplus B_\infty^{2n}(\epsilon)$ $(B_\infty^{2n}(\epsilon) \coloneqq \{\xi' \in \mathbb{R}^{2n} \mid \| \xi' \|_\infty \leq \epsilon\})$, where ϵ is the error bound. Then, the inner polytopic approximations of the terminal residual sets for the attack-free and the attacked process are computed as $D_{r_e}^{est}(I,K,L)$ and $D_{r_{a,e}}^{est}(\Lambda,K,L)$, respectively. These inner approximations satisfy $D_{r_j}^{est}(\Lambda,K,L) = D_r^{est}(\Lambda,K,L) \ominus A_r(\Lambda)B_\infty^{2n}(\epsilon) \subseteq D_r(\Lambda,K,L)$ $(j = e$ or $j = a,e)$, where, $D_r^{est}(\Lambda,K,L)$ is the outer polytopic approximation of $D_r(\Lambda,K,L)$.

With the disturbance set F, the attack magnitude of interest Λ, the controller gain K, and the observer gain L as the input, the controller screening algorithm is as follows. First, it is checked if $\max_i |\lambda_i(A_\xi(\Lambda,K,L)| \geq 1$. If this is true, the choice of K and L under a multiplicative attack with attack matrix Λ will render the closed-loop process unstable and the attack is detectable. The screening algorithm is terminated in this case. However, if $\max_i |\lambda_i(A_\xi(\Lambda,K,L)| < 1$, then the outer polytopic approximations of the residual sets are computed using the method described in Section 3.2 of Narasimhan et al., 2021. From the outer approximations, the inner approximations of the residual sets are computed. Then, the radii of the 2-norm balls enclosing the sets ($R^{est}(I,K,L)$, $R^{est}(\Lambda,K,L)$, $R_e^{est}(\Lambda,K,L)$, and $R_{a,e}^{est}(\Lambda,K,L)$) are computed. If $R^{est}(\Lambda,K,L) < R_e^{est}(I,K,L)$, the attack is classified as an undetectable attack, else, if $R_{a,e}^{est}(\Lambda,K,L) > R^{est}(I,K,L)$, then the attack is classified as a potentially detectable. However, if it is found that $R^{est}(\Lambda,K,L) > R_e^{est}(I,K,L)$ or $R_{a,e}^{est}(\Lambda,K,L) > R^{est}(I,K,L)$, the test is inconclusive. This may occur when the numerical approximations of the terminal sets satisfy $|R_{a,e}^{est}(\Lambda,K,L) - R^{est}(I,K,L)| \leq \beta$ or $|R_e^{est}(\Lambda,K,L) - R^{est}(I,K,L)| \leq \beta$ where $\beta > 0$ is a small number, implying that the radius estimates are close to each other.

3. Application to a Nonlinear Chemical Process

A chemical process example consisting of a continuously stirred tank reactor (CSTR) with a second-order reaction occurring is considered. The process is subject to bounded process disturbances and measurement noise. In the real-time control of chemical processes, the control actions are usually applied to the nonlinear continuous-time process at discrete time instances. To simulate this, the nonlinear ordinary differential equation model describing the CSTR process is integrated with the linear controller generating a new control action at every 10^{-2} h. To apply the screening methodology, a discrete-time LTI process model is needed. To generate the model, a continuous-time LTI model is first obtained by linearizing the nonlinear model about its open-loop stable steady state. The discrete-time LTI process model is then obtained by applying a zero-order hold discretization. The process models, and process parameter values are given in Narasimhan et al., 2021.

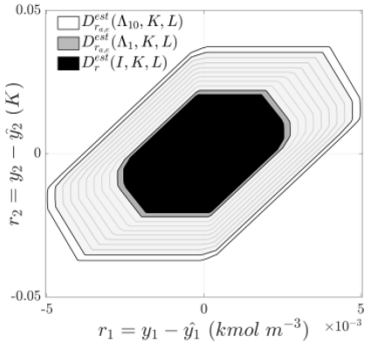


Figure 1. Outer approximation of the terminal residual set for the attack-free process ($D_r^{est}(I,K,L)$), and the inner approximations of the terminal sets for the attacked process ($D_{r_{a,\ell}}^{est}(\Lambda_i,K,L)$).

In this study, the control parameters are to be chosen such that they do not mask multiplicative attacks of magnitude in this range $[1.1, 2]$ on the concentration sensor-controller link. Since there are infinitely many attack magnitudes in the range, the screening algorithm is carried out with ten values from the range, i.e., with values $\Lambda_i = \text{diag}(\alpha_i, 1)$ where $\alpha_i = 1.1 + 0.1(i-1)$, $i = 1,2,\dots,10$. Using pole placement, the controller and observer gains are selected with the controller poles at $[0.5 - 0.1]$ and observer poles at $[-0.3\ 0.4]$. The screening algorithm is applied to the chosen controller design over all Λ_i. The attacked closed-loop process is stable with $\max_j \left|\lambda_j\left(A_{\xi_a}(\Lambda_i,K,L)\right)\right| < 1$ for all $i = 1,2,\dots 10$. The polytopic approximations of the minimum invariant sets of the attacked process over all attack magnitudes, and for the attack-free process are computed with an error bound of $\epsilon = 5 \times 10^{-5}$. Then, the polytopic approximations of terminal residual sets are computed (Figure 1). The size of the terminal residual set for the attacked process increases with α_i. For all cases, $R^{est}(I,K,L) < R_{a,e}^{est}(\Lambda_i,K,L)$, $i = 1,2,\dots 10$, with $R^{est}(I,K,L) = 0.0207$, $\min_i R_{a,e}^{est}(\Lambda_i,K,L) = 0.0222$ for $\Lambda_1 = \text{diag}(\alpha_1,1) = \text{diag}(1.1,1)$ and $\max_i R_{a,e}^{est}(\Lambda_i,K,L) = 0.0375$ for $\Lambda_{10} = \text{diag}(2,1)$. Thus, all the attacks in the range of interest are potentially detectable. To verify the potential detectability of each attack magnitude, two sets of simulations of the process are performed, using the (1) discrete-time LTI process model and (2) nonlinear continuous-time process model. One thousand closed-loop simulation pairs of the attacked process are performed with various realizations of the measurement noise and the process disturbance. For each simulation pair, the same realization was applied. For both sets of simulations, the state is estimated using a Luenberger observer designed based on the discrete-time LTI process model. To monitor the process, a CUSUM detection scheme described by Eq. (12) in Narasimhan et al., 2021 is used. The detection scheme is tuned with the choice of $b = R^{est}(I,K,L) = 0.0207$ and an alarm threshold of $\tau = 0.01$ to guarantee a zero-false alarm rate in the absence of an attack.

For attacks with $\alpha_i \in [1.6, 2]$, the attack is detected over most simulations with the LTI and the nonlinear models. Results from simulating an attack of magnitude $\Lambda_{10} = \text{diag}(2,1)$ on the nonlinear process model are presented in Figure 2. The attack is detected in 969 out of 1000 simulations with the nonlinear continuous-time process model. Of these 969 simulations, the attack is not detected in the corresponding linear model simulation in 4 simulations. Similarly, during the simulations with the discrete-time LTI process model, the attack is detected in 970 out of 1000 simulations. For 5 of the 970 simulations, the attack is not detected for the

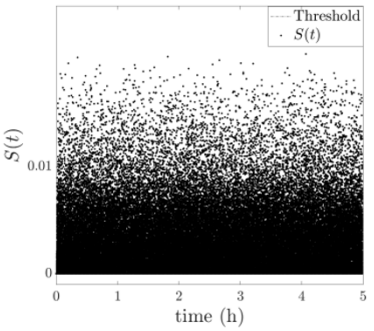


Figure 2. The CUSUM statistic $S(t)$ over 1000 simulations of the closed-loop process under an attack of magnitude $\Lambda_{10} = \text{diag}(2,1)$.

corresponding nonlinear model simulation. In most of the simulation pairs where a discrepancy between attack detection is observed, the difference in the maximum CUSUM statistic value between the two cases was small, meaning that in one case, the value exceeded the threshold by a small amount and in the other, the statistic value was close to the other, but did not exceed the threshold. For example, an attack is detected in the linear case with a CUSUM statistic value of 0.0101, and for the corresponding nonlinear case, the statistic value at the same time is 0.0095, and the attack is not detected. Finally, for the same realization of random variables applied to the nonlinear case and linear case, except for 174 simulations, the time at which the attack is first detected is the same for both process models. Over most of these 174 simulations, the attack detection time for the two process models differed by a single time step.

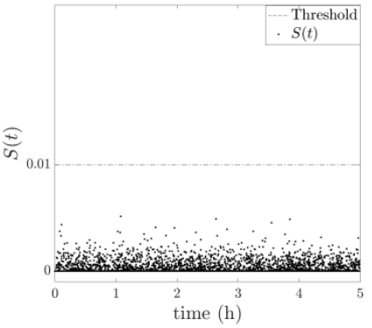


Figure 3. The CUSUM statistic $S(t)$ over 1000 simulations of the closed-loop process under an attack of magnitude $\Lambda_4 = \text{diag}(1.4,1)$.

For attacks with magnitude in range $\alpha_i \in [1.2, 1.5]$, while the CUSUM statistic never breaches the threshold, it is found to be a non-zero value over most simulations for both the nonlinear case and the linear case. The CUSUM statistic resulting from simulating an attack of magnitude $\Lambda_4 = \text{diag}(1.4, 1)$ (with $R_{a,e}^{est}(\Lambda_4, K, L) = 0.027$) on the nonlinear model are given in Figure 3. It is also found that the maximum realization of the CUSUM statistic over each simulation of the nonlinear model is of the same order of magnitude as the maximum realization of the statistic over the corresponding simulation of the linear model. In one case, the maximum CUSUM statistic in simulations with the nonlinear case is 0, however, the maximum statistic for corresponding simulations with the linear case is 3.1×10^{-5}. The mean of the 2-norm of the residual vector over 1000 simulations is 0.0081 with a variance of 2.4563×10^{-5} for the nonlinear case. The mean of the 2-norm of the residual vector over 1000 simulations with the linear case is 0.0081 (same as that for the nonlinear case) and the variance is 2.4553×10^{-5}. While not shown here, statistical distribution of the 2-norm of the residual indicates that tuning the CUSUM detection scheme with a lower alarm threshold may enable the detection of this attack. Note that with the tuning approach adopted in this work, the CUSUM statistic remains at zero for the attack-free process, and any non-zero CUSUM statistic value would indicate an attack. Furthermore, with a decrease in α_i, the total number of simulations with non-zero realizations of the CUSUM statistic decrease, indicating that the attack becomes more difficult to detect.

During closed-loop simulations with an attack of magnitude $\Lambda_1 = \text{diag}(1.1,1)$, the CUSUM statistic remains at zero over all the 1000 simulations for both the linear case and the nonlinear case. Thus, the attack goes undetected, despite the control parameters satisfying $R^{est}(I, K, L) < R_{a,e}^{est}(\Lambda_1, K, L)$. This does not contradict the screening algorithm, as the attack is only potentially detectable. The fact that the attack went undetected may be because $R^{est}(\Lambda_1, K, L) - R^{est}(I, K, L) = 0.0015$ is small. While further analysis is required, for this attack magnitude, a different set of control parameters with a larger separation between the sets $D_r^{est}(I, K, L)$ and $D_r^{est}(\Lambda_1, K, L)$ may enable attack detection. Moreover, for this case, the performance degradation resulting from this attack is small.

The applicability of the controller screening algorithm to the nonlinear case is verified for other control designs and attacks. For example, when the controller screening algorithm predicts that (1) an attack $\Lambda = \text{diag}(0.5,1)$ with control and observer gains selected by placing the poles at [0.3 0.1] and [−0.4 0.3], respectively, and (2) an attack $\Lambda = \text{diag}(1,0.9)$ for control and observer gains selected by placing poles at $[0.5 - 0.1]$ and [0.4 0.4], respectively, are undetectable. During the closed-loop simulations with the nonlinear case and the linear case, the CUSUM statistic remains at zero, and the attacks are not detected. When the controller screening predicts that an attack $\Lambda = \text{diag}(1,1.2)$ is potentially detectable for control and observer gains with poles placed at $[0.5 - 0.1]$ and [0.2 0.2], respectively, the CUSUM statistic has non-zero realizations over most simulations. However, it never breaches the threshold (similar to Figure 3). Thus, the controller screening algorithm based on the discrete-time LTI model of the process is applicable to the continuous-time nonlinear model of the process. This may be because the linear process model adequately represents the dynamics of the nonlinear process near the operating steady-state.

4. Conclusions

In this work, the application of a detectability-based controller design screening methodology for a multiplicative sensor-controller link cyberattacks to a nonlinear chemical process example was demonstrated. Future work will focus on controller-based approaches that enable the detection of a multiplicative sensor-controller link attack.

References

T. Miller, A. Staves, S. Maesschalck, M. Sturdee, B. Green, 2021, Looking back to look forward: Lessons learnt from cyber-attacks on industrial control systems, International Journal of Critical Infrastructure Protection, vol. 35, pp. 100464.

J. Giraldo, D. Urbina, Á.A. Càrdenas, J. Valente, M. Faisal, J. Ruths, N.O. Tippenhauer, H. Sandberg, R. Candell, 2018, A Survey of Physics-Based Attack Detection in Cyber-Physical Systems, ACM Computing Surveys, vol. 51, issue 4, article 76, pp. 1–36.

S. Tan, J.M. Guerrero, P. Xie, R. Han, J.C. Vasquez, 2020, Brief survey on attack detection methods for cyber-physical systems. IEEE Systems Journal, vol. 14(4), pp. 5329–5239.

H. Oyama, H. Durand, 2020, Integrated cyberattack detection and resilient control strategies using Lyapunov-based economic model predictive control, AIChE Journal, vol. 66, issue 12, pp. e17084.

S. Chen, Z. Wu, P.D. Christofides, Cyber-security of centralized, decentralized, and distributed control-detector architectures for nonlinear processes, 2021, Chemical Engineering Research and Design, vol. 165, pp. 25–39.

V.M. Kuntsevich, B.N. Pshenichnyi, 1996, Minimal invariant sets of dynamic systems with bounded disturbances, Cybernetics and Systems Analysis, vol. 32(1), pp. 58–64.

S.V. Raković, E.C. Kerrigan, K.I. Kouramas, D.Q. Mayne, 2005, Invariant approximations of the minimal robust positively invariant set, IEEE Transactions on Automatic Control, vol. 50, no. 3, pp. 406–410.

S. Narasimhan, N.H. El-Farra, M.J. Ellis, 2021, Detectability-based controller design screening for processes under multiplicative cyberattacks, AIChE Journal, vol. 68, issue 1, pp. e17430.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50243-8

OPC UA information transfer via unidirectional data diode for ICS cyber security

Toshiaki Honda[a], Takashi Hamaguchi[a], Yoshihiro Hashimoto[a]

[a]*Nagoya Insititute of Technology, Nagoya 466-8555 Japan*
h.toshiaki.377@nitech.jp

Abstract

Cyber attacks on critical infrastructure pose serious threats. When industrial control systems (ICS) are targeted by cyber attacks, the resulting incidents cause not only security problems but also safety and pestilence issues. Encrypting the communication data passing through the data diode with OPC UA will further improve the security of the communication in ICS. In this paper, we propose a method to establish both a data diode that prevents intrusion and an OPC UA encrypted communication that prevents interception and tampering.

Keywords: OPC UA; Client/Server; PubSub; Key exchange; Data diode.

1. Improved security by combining data diode and OPC UA

Cyber attacks on critical infrastructure pose serious threats. If an industrial control system (ICS) is the target of a cyber-attack, the incidents it causes are not only security issues but also safety and serious pestilence issues. For example, the negative impact of a cyber-attack can be extraordinary, such as an explosion at a manufacturing site, damage to customers or the environment due to improper shipment of products, or the negative impact on customers and society due to temporary suspension of manufacturing services.

If a controller such as a DCS (Distributed Control System) or PLC (Programmable Logic Controller) behaves improperly, it can cause a serious accident. Unfortunately, the controller is unable to determine whether it is under a cyber attack, so even if it receives a malicious command, it will do as it is told. There are many machines that can obtain controller commands and information about the target of the attack and send commands to the controller, such as operational support systems, advanced control systems, MES (Manufacturing Execution System) servers, and SCADA (Supervisory Control And Data Acquisition) system engineering workstations. If the attacker can obtain information about the controller commands and the target of the attack and hijack these devices to send dangerous commands, a serious accident may occur even if the communication protocol is secure OPC UA.

Therefore, we want to monitor what kind of commands are being communicated, even if they are encrypted. We have already proposed and developed a monitoring system that can manage encrypted communications [1]. This monitoring system will be installed close to the controllers. Since there are many controllers, the monitored command information is aggregated and stored in a separate location. If the location is hijacked by an attacker, the controller can be attacked through the monitoring system installed near the controllers. To solve this problem, we thought of installing the unidirectional data

diode in the aggregation path so that even if the aggregation system is hijacked, the attack cannot reach the controller [2].

OPC UA was developed in 2006 as a communication protocol that considers security at the time of design and has been selected as a standard protocol for Industry 4.0, etc. However, since it is a one-to-one communication standard, the engineering cost when setting up communication between many devices has been an issue. The introduction of Publishers and Subscribers (PubSub) communication in addition to Client/Server was announced in 2018 and implemented starting with OPC UA v1.04, released in 2021 PubSub supports UDP, MQTT, and AMQP and can send data to many subscribers at once. Since the flow of data is one-way in PubSub UDP, data diodes can be used in conjunction with it. It can also be used as a network infrastructure able to route datagram-based messages in a broker-less form [3]. Nuclear power plants require the use of unidirectional data diodes when disclosing data. By using unidirectional data diodes to prevent an attacker's invasion and by encrypting the communication data that passes through them with OPC UA, interception and tampering can be prevented, further improving the security of the communication [4] [5].

In this paper, we propose a method for using both data diodes and OPC-UA encrypted communication because security for industrial control systems is an important issue in process system engineering.

2. OPC UA secure communication via the unidirectional data diode

2.1. Basic OPC UA Client/Server models encryption key exchange

The OPC UA's client/server is implemented at the top of the TCP/IP (Transmission Control Protocol/Internet Protocol) stack. The security architecture of OPC UA is a hybrid of the PKI (Public Key Infrastructure) and symmetric key cryptography. The symmetric key cryptosystem is used for encrypting the data, and PKI is used for encrypting the exchange information necessary for authenticating the application and generating the symmetric key. The symmetric key is not reused because it is updated every time the communication connection of OPC UA is started. Moreover, in OPC UA, the confidentiality of the data is improved by using two or more common keys for encryption.

2.2. Basic OPC UA PubSub models encryption key exchange

The OPC Foundation has added the SKS (Security Key Services) specification for use with OPC UA PubSub to the OPC UA PubSub Part 14 specification. SKS is a key exchange architecture. Specifically, SKS is an architecture that provides a common key that publishers can use to sign and encrypt data and subscribers can use to verify and decrypt data signatures. Figure 1 shows the basic key management pattern for OPC UA PubSub. The OPC UA server should support SKS for key exchange. Then, the Publisher should install the encryption key on the SKS. Since SKS is a key management function supported by the OPC UA server, the OPC UA Client/Server model is used for key exchange. Therefore, if the Subscriber obtains the key from SKS, it needs to support the OPC UA client. Furthermore, the key exchange takes place with an authenticated OPC UA client [6].

2.3. Encryption key exchange via the unidirectional data diode

The use of unidirectional data diodes improves security because they can completely block incoming packets. However, since bidirectional communication is not possible, key exchange using SKS through unidirectional data diodes is not possible. Here, we propose a key exchange that can be used when using unidirectional data diodes.

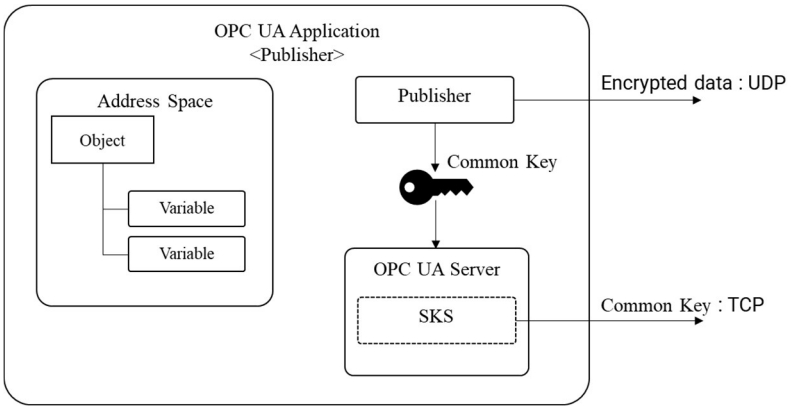


Figure 1 SKS architecture

- Publisher physically hands the key to the Subscriber

  A Publisher has multiple topics, and each topic has multiple subscribers. Each topic has a different common key for reading it but the same common key for subscribers of the same topic. Since the subscription period of a topic may be different for each Subscriber, the subscription for each Subscriber needs to be confirmed at the time of registration, and as shown in the arrow at the bottom of Figure 1, different communication is inevitably required than when reading the topic.

  The subscription is cross-confirmed through a channel that allows mutual communication, which is opened only during registration, and the Subscriber passes the public key corresponding to his private key to the Subscriber. When the Subscriber is allowed to register, the information needed to synthesize the symmetric key information for the subscription is encrypted by the public key and sent to the Subscriber.

  This information can only be decrypted by the Subscriber, so confidentiality is ensured. This procedure corresponds to steps (1) through (6) in Figure 2.

To ensure security by switching the common key, the common key information received as described above is not the value of a single common key but contains multiple common keys, and during communication, the subscriber switches which of these keys to use.

- Encrypted communication to switch the common key to be used

  The Publisher randomly selects a symmetric key to be used for encryption from the set of symmetric keys exchanged in the first step and sends the ID of the key used in addition to the encrypted communication data. The received Subscriber uses the key ID information to identify the common key for decryption and decrypts the communication data. This procedure corresponds to which of the keys from (7) in Figure 2 is used for decryption.

We propose the above two hybrids. The concept of key exchange is based on PKI. However, as mentioned earlier, only UDP can be used for communication when communicating with unidirectional data diodes. Also, it is a broker-less form of communication.

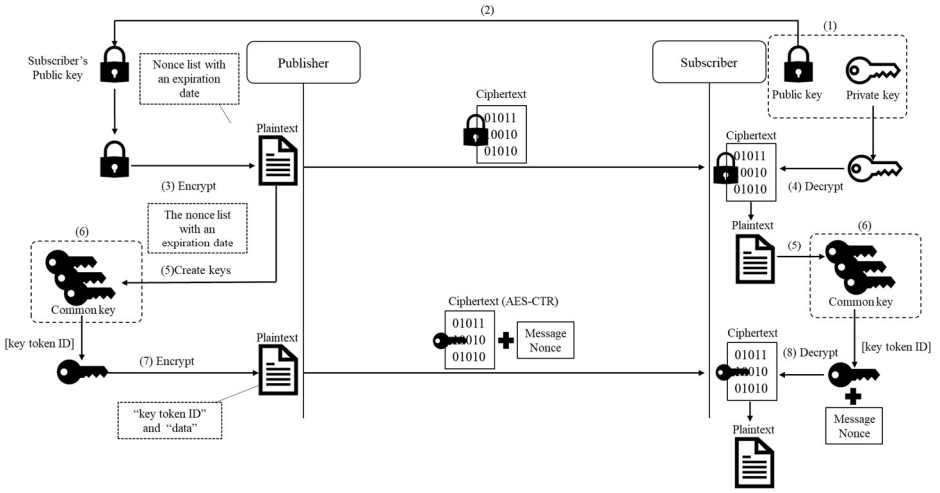


Figure 2 Encryption key exchange sequence

Figure 2 shows the Publisher and Subscriber sequence. It communicates via UDP between Publisher and Subscriber. Figure 2 also shows the key exchange between Publisher and Subscriber. In Figure 2, two pieces of data are sent from Publisher to Subscriber.

The first sequence sends the information needed to generate the common key from the Publisher to the subscribers. The Publisher is required to obtain the public key generated by the Subscriber in advance. The public key cannot be exchanged via the unidirectional data diode, so the public key will need to be physically passed to the Publisher. This can be very laborious and is open to human error, but the public key is safe to send because the Publisher can determine that only the Subscriber with the private key paired with the public key can decrypt it. Subscribers can also specify that the transmission is from an authorized publisher because the encrypted data can be decrypted with their own private key. When exchanging public keys, they need to be carefully exchanged.

The second sequence uses the generated common key to encrypt the data and sends the encrypted data and the key token ID together from Publisher to Subscriber. In OPC UA PubSub, a key token ID can assign to the common key, and the expiration date can be set for the common key. Therefore, OPC UA PubSub can manage multiple common keys and randomly use the common key to encrypt data when sending. The Subscribers can decrypt the encrypted data on the basis of the received key token ID.

2.4. OPC UA PubSub Data encryption and signature

This section describes what encryption OPC UA PubSub does when sending data after key exchange. OPC UA PubSub supports data encryption and signing. The security modes are:

- No security

- Signing but no encryption
- Signing and encryption

Message security is end-to-end security (from Publisher to Subscriber). "No security" means that the message is plain text. "Signing but no encryption" means that the message is only signed. "Signing and encryption" means signing and encrypting a message.

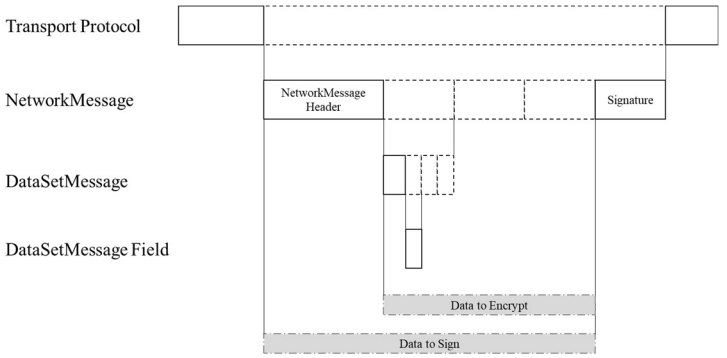


Figure 3 Scope of data signing and encryption

Figure 3 shows the extent of data signing and encryption. The Transport Protocol layer means packets of the protocol. The dashed line in the Transport Protocol layer is the protocol payload. The next layer is the layer defined by OPC UA:

- NetworkMessage
- DataSetMessage
- DataSetMessage Field

NetworkMessage is a container for DataSetMessages that contain information for exchanging data. It also contains security information. The Security information includes signature and encryption processing information. DataSetMessage and DataSetMessage Field are encrypted. DataSetMessage is a collection of DataSetMessage Fields. DataSetMessage Field is the actual value. The scope of encryption is DataSetMessage and DataSetMessage Field.

OPC UA PubSub uses Advanced Encryption Standard Counter Mode (AES-CTR) for data encryption and requires a common key and Message Nonce to decrypt the data. Figure 4 shows the security information contained in NetworkMessage. AES-CTR provides encryption and decryption with pre-exchanged keys and Message Nonce contained in the message. Therefore, data encryption can be more complex, and we can further strengthen our defense against attacks and threats [7].

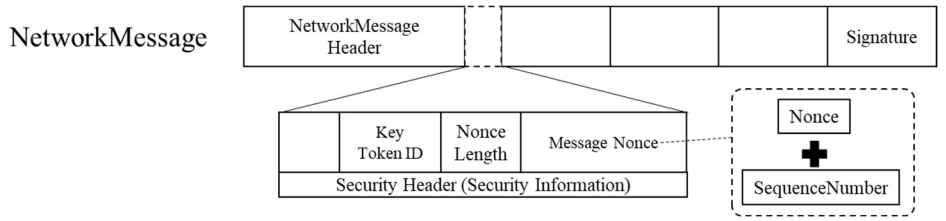


Figure 4 Security information structure

3. Conclusions

This paper proposed a key exchange used when transmitting encrypted data via the unidirectional data diode by OPC UA PubSub. Key exchange does not exchange the common key. The publisher and subscriber exchange randomly generated nonces to generate a common key. Furthermore, OPC UA PubSub uses AES-CTR for data encryption and requires a common key and the Message Nonce to decrypt the data. This can further strengthen our defense against attacks and threats.

Since we have already achieved plaintext communication of OPC UA PubSub, we plan to implement encrypted data via the unidirectional data diode on the basis of this design.

References

[1] W. Machii, I. Kato, M. Koike, M. Matta, T. Aoyama, H. Naruoka, I. Koshijima and Y. Hashimoto, “Dynamic zoning based on situational activitie for ICS security” 10th Asian Control Conference (ASCC), May 2015, pp.1242-1246

[2] T. Honda, Y. Shimazawa, T. Hamaguchi and Y. Hashimoto, "Development of a monitoring system for encrypted data by OPC UA," IIAI 2021 10th International Congress on Advanced Applied Informatics, 2021, pp. 731-736

[3] A. Eckhardt, S. Müller and L. Leurs, "An Evaluation of the Applicability of OPC UA Publish Subscribe on Factory Automation use Cases," 2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA), 2018, pp. 1071-1074

[4] IAEA, "IAEA Nuclear Security Series No. 17," 2011

[5] B. Jeon and J. Na, "A study of cyber security policy in industrial control system using data diodes," 2016 18th International Conference on Advanced Communication Technology (ICACT), 2016, pp. 314-317

[6] OPC Foundation, "OPC Unified Architecture Specification Part 14: PubSub," 2018

[7] IETF, “AES Counter Mode Cipher Suites for TLS and DTLS”, 2006

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50244-X

Study on Device Authentication System for Dynamic Zoning of Industrial Control Systems

Takanori Miyoshi[a], Shota Shimizu[a], Kanata Nishida[a], Masato Izawa[a], and Isao Kato[a*]

[a]*Cyber-Security R&D Office, Sumitomo Electric Industries, Ltd., 1-1-3, Shimaya, Konohana-ku, Osaka 554-0024, Japan*
kato-isao@sei.co.jp

Abstract

In industrial control systems (ICSs) used in critical infrastructure (CI) such as substations and chemical plants, measures are needed to prevent serious physical damage caused by cyber attacks. This is because high volume simultaneous cyber attacks on CI may lead to complete system malfunction and a chain of serious accidents. For example, if a large-scale power outage occurs due to an accident at a substation, it may affect other CI such as transportation facilities and hospitals. In this paper, we propose a new dynamic zoning method for making system functions redundant and connecting communication paths between functions only when communication is required.

Keywords: Dynamic Zoning; Safety and Security; Device Authentication.

1. Introduction

The number of reports of cyber incidents related to ICSs used in electric power plants and chemical plants has been increasing in recent years (NCCIC, 2016) (Symantec, 2019). The purpose and target of cyber attacks are also changing to geopolitical ones (such as political use, a religious issue or terrorism) (Kaspersky, 2019). A different approach is needed because control systems differ in the assets and features to be protected from traditional information systems. In order to ensure safety, ICSs require measures to increase the probability that the system will move to a state without a possibility of physical damage, assuming that a threat that cannot be completely removed may intrude (Safety-II (Hollnagel, 2014)). Cyber security standards for ICSs (e.g. IEC-62443) recommend isolation of threats and blocking of unnecessary communication paths by defense in depth (DiD) using zones and conduits to improve the security of control systems (Knapp, 2011) (IEC62443-3-2, 2020).

To achieve DiD for control systems, types of fixed and dynamic zoning techniques have been proposed (Hashimoto et al., 2013) (Morita et al., 2013) (Moritani et al., 2014). Since the assets to be protected in the control system change over time, it is necessary to switch the focus of zone on a time axis. Therefore, a dynamic zoning method has been proposed in which zones are dynamically switched according to the operating status of the control system (Machii et al., 2014) (Machii et al., 2015). In a system in which activated functions and assets to be protected change over time, communication between devices should be normally-off and connected only when necessary.

In previous studies (Machii et al., 2014) (Machii et al., 2015), the dynamic zoning method was a software based method to control the logical disconnection of communication channels. As such, the communication channel between devices remains physically

connected, and the risk of cyber-attacks on the control system using this communication channel is not completely eliminated. In addition, in the communication control function, the code data used for authentication in device authentication when allowing communication between devices is usually stored and protected in a single location, and a single successful attack could result in the theft of the code data, which could completely hijack the communication control function.

Therefore, for the ideal realization of DiD, the requirements of the communication control function for dynamic zoning include physical connection blocking control of communication paths and distributed management of authentication codes used by the communication control function for device authentication.

The purpose of our study is to implement a more robust dynamic zoning system that detects unauthorized devices on the network and blocks communication on the network to protect society such as a stakeholder's safety and company viability. We proposed a device authentication system that consists of multiple devices with logically independent communication lines (such as separate lines or an aggregation of multiple communication lines). By using our authentication system, when an abnormality in the system due to a cyber-attack is confirmed, it is possible to isolate the threat without stopping the entire system by physically disconnecting the power supply and communication path leading to the connected devices that are in an abnormal state.

2. Device Authentication System for Dynamic Zoning

In this section, we describe the concepts for two types of device authentication systems for dynamic zoning that we are currently developing.

2.1. Device Authentication System with Authentication Switch for ICSs

The concept of this system is shown in Figure 1. This system is a device authentication system using an authentication switch, which has both an authentication function to determine whether or not communication is possible between connected devices, and a communication control function to physically block or switch the connection of communication paths. This system consists of networks (communication lines, power supply lines, and authentication lines), a state control agent in MRP (Manufacturing

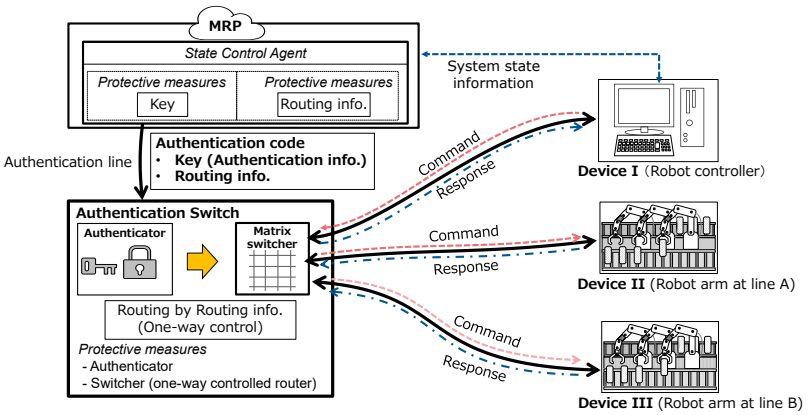


Figure 1: Concept of Device Authentication for ICSs

Resource Planning), an authentication switch, and devices (a robot controller and two robot arms).

The authentication code, which contains authentication information (Key) to verify legitimacy and routing information between connected devices to determine if they can connect to the network, is entered into the authentication switch. After determining the communication availability of the connected devices using the Key in this authentication code, the communication path between the devices is selected using the routing information.

This authentication code can include not only spatial routing information, but also state information of the equipment or control system itself (System state information). By using the system state information, it is possible to dynamically switch the communication route according to the state of the equipment or the operating status of the factory system.

The device authentication procedure for ICSs is as follows:

1. The state control agent in MRP sends the authentication code to the authentication switch via the authentication line.
2. The authentication switch uses the Key to authenticate the device.
3. If the authentication is successful, the authentication switch uses the system state information and the routing information to determine the route between the connected devices by the matrix switcher.

It is also possible to divide and manage this authentication code, and the authentication and routing information of the authentication switch is determined when all of the divided codes are collected. This segmented authentication code is stored in multiple locations that are spatially independent and separated from each other, and each of these locations is protected using different protection methods. Even if a cyber attack is successful and the attacker is able to steal part of the authentication code, the required attack cost is higher than usual because the attacker has to also successfully attack other protection measures that protect the split code to obtain all the segmented codes in order to successfully authenticate.

2.2. Device Authentication System with Security Unit for Mobility Network Systems

Figure 2 shows the concept of device authentication system for mobility network systems that manages automated guided vehicles (AGVs) and other mobility devices in a factory.

This system authenticates devices before they are connected to the network and provides power and a physical connection to the network only to those devices that have successfully authenticated. This system is intended to be used mainly in bus-type networks, and since it only performs connection blocking control of communication, the system can be built with a simpler functional configuration than the device authentication system for ICSs.

Specifically, a physical switch, called a security unit, is used to physically connect the network and devices only after successful authentication. This will prevent unauthorized devices from physically connecting to the network. This system consists of networks (communication lines, power supply lines, and authentication lines), security units, and connected devices. The security unit is a device that has physical network switches for communication and power supply lines and authentication function.

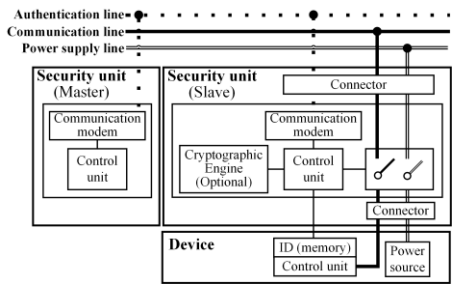


Figure 2: Concept of Device Authentication for Mobility Network Systems

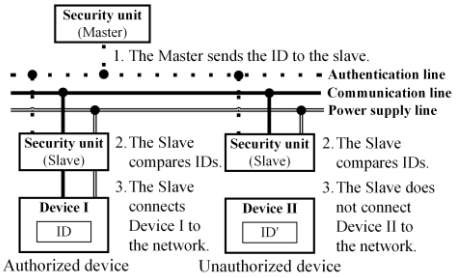


Figure 3: Device Authentication Procedure for Mobility Network Systems

This system uses a Master-Slave type authentication method comprising a security unit, and devices are connected to the network through this security unit (Slave). The security unit consists of a control unit, a physical switch, and a communication modem. The control unit authenticates the connected devices and controls the physical switches. The physical switch connects the device to the communication and power supply lines of the network only if authentication is successful. This makes it possible to prevent unauthorized devices from physically connecting to the network. The communication modem communicates authentication information between security units via the authentication line. Authentication is performed via independent signal lines, which are different from the communication and power supply lines of the network, just like the device authentication system for ICSs. The device has a unique device ID that is used for authentication.

Device authentication procedures using a security unit are shown in Figure 3. The authentication procedure is as follows:

1. The Master sends the ID to the Slave.
2. The Slave compares received ID with the ID held by the connected device.
3. If the IDs match, the Slave will connect the device to the network.

In addition to ID authentication, this system can perform cryptographic authentication between Slave-connected devices using a cryptographic engine, thus enhancing security functions through two-step authentication (Option).

3. System Implementation and Results

To show that our proposed dynamic zoning scheme is feasible using realistic implementation costs and devices available in the market, we have developed a prototype authentication system using a security unit that controls the connection between the CAN bus and the device.

The control unit of the security unit uses an 8-bit microcontroller EFM8BB3 with an operating frequency of 22 MHz. The communication modem uses FSK modulation, and the communication speed can be switched from 4.8 kbps to 22.8 kbps. The ID of the device can be read out between the security unit and the device via 1-wire communication at 16 kbps serial communication.

In our prototype system, we assumed that the devices will be authenticated in a harness network for mobility devices, and that up to 8 devices will be connected to the harness.

Table 1: Evaluation Results and Estimated Time of Device Authentication

	One-to-one (Measured value)	One-to-eight (Estimated value)			
Communication speed [kbps]	19.2	4.8	9.6	19.2	22.8
Authentication time [ms]	34.5	580.8	296.9	163.0	116.5

Assuming that the allowable processing time of an application is 300 ms from the powering-on of a device to the completion of the start-up of all devices, we aimed to keep the authentication time of all devices in this system to within 150 ms, half of that time.

In order to evaluate performance, we confirmed on the actual security unit that the authentication of multiple devices (1-3 devices) connected to the network can be performed successfully. In addition, we calculated the estimated authentication processing time for 8 devices based on the results of this actual measurement, and confirmed that the estimated value could meet the target.

The evaluation included an evaluation of authentication and communication functions and a measurement of authentication time. We checked operation when authentication succeeded and when it failed in order to evaluate the authentication function. In the authentication time measurement evaluation, we measured the time from powering-on until all devices were fully booted after authentication was completed.

Table 1 shows the evaluation results for the authentication time of the prototype system. Using these actual measurements, we estimated the total processing time required for 1-to-8 authentication between Master and Slave. From this estimate, we confirmed that our target time can be met when the communication speed is 22.8 kbps.

4. Discussion

From the performance evaluation results, we have shown that it is possible to construct our proposed device authentication system using small and inexpensive devices. To further reduce the authentication time, it is effective to increase the communication speed in the authentication lines and to reduce the amount of communication data used for authentication.

However, the device authentication system implemented this time does not use two-step authentication with cryptographic authentication using an optional cryptographic engine. Therefore, if the cryptographic strength of the cryptographic engine is high, the processing time of the entire device authentication will increase due to the increase in the processing time of the cryptographic engine. In addition, circuit size may be larger if tamper resistance of the cryptographic engine is included. In environments where communication lines are susceptible to noise, the communication method and speed may be limited to guarantee noise immunity of the authentication lines.

5. Conclusions

In this paper, we proposed a new device authentication system that decentralizes and manages the authentication codes used for device authentication, and controls the physical connection and disconnection of communication and power supply lines. We

also developed a prototype and evaluated the performance of the device authentication system for mobility devices, and showed that our proposed method can be realized with realistic implementation costs and devices available in the market. We also plan to conduct the experiments for the device authentication system for ICSs to show that this system can also be realized using devices that are available in the market and have realistic implementation costs.

In an always-on system, isolation of threats by dynamic zoning is an effective means against threats from cyber attacks. Threat isolation using dynamic zoning allows us to remove threats while the system is running, and at worst, safely shut down the system. This enables early restoration of the system, so our study on dynamic zoning implementation can contribute to the enhancement of resilience for business continuity.

References

Y. Hashimoto, T. Toyoshima, S. Yogo, M. Koike, T. Hamaguchi, S. Jing, & I. Koshijima, 2013, Safety securing approach against cyber-attacks for process control system, Computers & Chemical Engineering, Volume 57, 181-186.

E. Hollnagel, 2014, Safety-I and Safety-II, Routledge, ISBN-13: 978-1472423085.

IEC62443-3-2, 2020, Security for industrial automation and control systems - Part 3-2: Security risk assessment for system design, 1st ed.

Kaspersky, 2019, Kaspersky Security Bulletin 2019. Advanced threat predictions for 2020, Retrieved from https://securelist.com/advanced-threat-predictions-for-2020/95055/, (accessed 2022-01-14).

E. D. Knapp, 2011, Industrial Network Security, Syngress, ISBN-13: 978-0124201149.

W. Machii, I. Kato, M. Koike, M. Matta, T. Aoyama, I. Koshijima, & Y. Hashimoto, 2014, Dynamic Zoning of the Industrial Control System for Security Improvement, The 5th World Conference of Safety of Oil and Gas Industry (WCOGI 2014), Paper No. 1065756.

W. Machii, I. Kato, M. Koike, M. Matta, T. Aoyama, H. Naruoka, I. Koshijima, & Y. Hashimoto, 2015, Dynamic Zoning Based on Situational Activitie for ICS Security, The 10th Asian Control Conference (ASCC 2015), 1242-1246.

T. Morita, S. Yogo, M. Koike, T. Hamaguchi, S. Jung, I. Koshijima, & Y. Hashimoto, 2013, Detection of Cyber-attacks with Zone Dividing and PCA, Procedia Computer Science, Volume 22, 727-736.

H. Moritani, S. Yogo, T. Morita, M. Kojima, K. Watanabe, J. Sun, I. Koshijima, & Y. Hashimoto, 2014, Development of cad for zone dividing of process control networks to improve cyber security, 2014 14th International Conference on Control, Automation and Systems (ICCAS 2014), 1311-1316.

National Cybersecurity and Communications Integration Center (NCCIC), 2016, ICS-CERT Year in Review 2016, Retrieved from https://us-cert.cisa.gov/sites/default/files/Annual_Reports/Year_in_Review_FY2016_Final_S508C.pdf, (accessed 2022-01-14).

Symantec, 2019, Symantec 2019 Internet Security Threat Report, Volume 24. Retrieved from https://docs.broadcom.com/docs/istr-24-2019-en, (accessed 2022-01-14).

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50245-1

Designing Framework for Tabletop Exercise to Promote Resilience Against Cyber Attacks

Yuitaka OTA[a*], Haruna ASAI, Shiho TANIUCHI, Erika Mizuno[a], Tomomi AOYAMA[a],Yoshihiro HASHIMOTO[a,] and Ichiro KOSHIJIMA[a]

[a]*Nagoya Institute of Technology, Nagoya 466-8555, JAPAN*
cjr17009@nitech.ac.jp

Abstract

In recent years, there has been an increase in the number of cyber-attacks targeting ICS (Industrial Control System), also used in the critical infrastructure that supports our daily lives. As a result, companies using ICS need to improve their ability to respond to cyber incidents to ensure business continuity. Cyber incidents at ICS require an enterprise-wide response, as they have a significant impact on IT system security issues, plant safety, and the company's business. For this company-wide response to proceed smoothly, it is essential to create a response plan in advance and create an environment that allows for an instant response when an incident occurs. Many companies do not have enough experience with ICS-targeted incident response to develop this plan. In this study, we developed an exercise to simulate an incident and improve response to a cyber-attack.

Keywords: Cyber Security, Exercise, Cyber-Resilience

1. Introduction

Since controllers installed in a uranium enrichment plant were cyber-attacked in 2010[1], we must recognize that cyber-attacks have become a real threat to our production plants. It is fresh in our memory that in May 2021, the US pipeline [2] interrupted its service due to a cyber-attack by ransomware. In addition, HatMan[3] has been reported as malware that targets "safety instrumentation systems" that protect plant safety.

Until now, companies have created and operated Safety-Response Plans for safety incidents such as equipment failure and natural disasters on the ICS side, and Security-Response Plans for security incidents such as information leakage on the IT side. We have created and implemented. The authors believe it is essential to consider and prepare a business continuity plan (BCP) to ensure the plant's safety and business even in a cyber-attack on the ICS network. In addition, we believe that it is necessary to understand how cyber-attacks affect plants to consider countermeasures against cyber-attack-targeting plants (ICS-BCP). However, many companies have experience with cyberattacks on IT systems, but not on OT systems. (They may be under cyber-attack but may not be able to detect them).

Therefore, we need a mechanism to gain experience with ICS cyber incidents to plan ICS-BCP. Hence, we are developing an exercise that allows to experience a simulated cyber incident to solve this problem. New exercises were developed to compensate for the improvements identified in other exercises designed and implemented.

In this paper, we describe the previous and newly developed exercises.

2. Exercises for improving resilience developed

We have been developing discussion-based ICS security exercises to recognize that hard skills such as control equipment operation and IT forensics and soft skills such as organizational cooperation and information sharing are essential in dealing with cyber incidents.

2.1. Type1: Roleplay based Cyber Defense Exercise [4]

Through this exercise, it is intended that the exercise participants will be able to discuss and learn about communication during incidents, their roles as incident response commanders, and the capabilities (communication and information gathering) required to fulfill their roles properly.
The field exercises simulate cyber incident management and require participants to respond to incidents by assessing the situation according to scenarios that play out in real-time. Figure 1 shows an exercise scenario as a communication path.

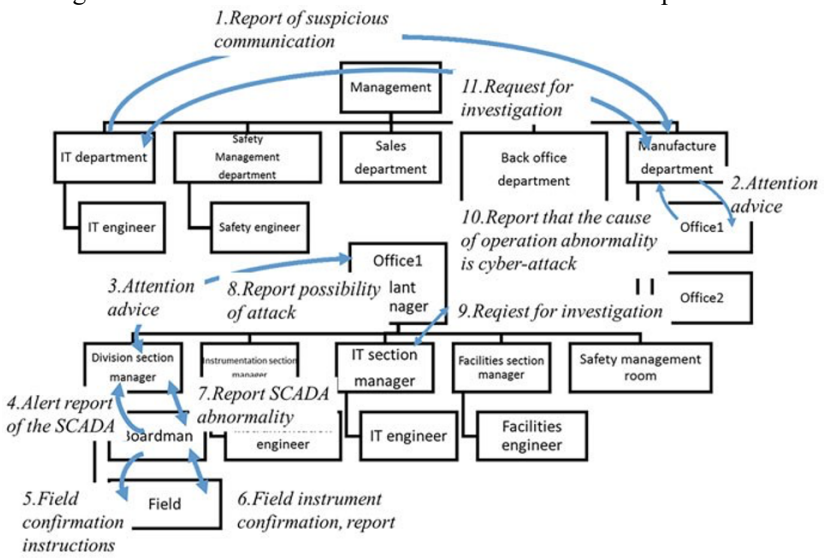


Figure1 Type1 exercise based on emergency communication management

This exercise is a simulation-type exercise based on a scenario that has been prepared in advance. Each exercise participant uses a computer in the exercise, and each is given a unique role in the scenario. There are two roles in the exercise: the role played by the exercise participants (Assigned Role) and the automatic response role (Virtual Role), which is the department or person in charge required to proceed with the exercise scenario. Each role is limited in the actions it can choose and the information it can know. Sharing the correct information at the right moment increases the number of actions available for selection.
As a result, by simulating a real-time accident response, the exercise participants will understand the importance of speed in decision-making when an accident occurs and that decisions can only be made if the proper information is conveyed to the right people at the right time.

2.2. Type2: Workflow Based Cyber Defense Exercise [5]

In this exercise, the following discussions will take place.
1. the actions and inter-organizational coordination required to ensure the safety of the factory where the incident occurred
2. the actions and communication required to suspect that the cause of the incident is a cyber attack
3. actions and communication required to eliminate the cause of the incident.

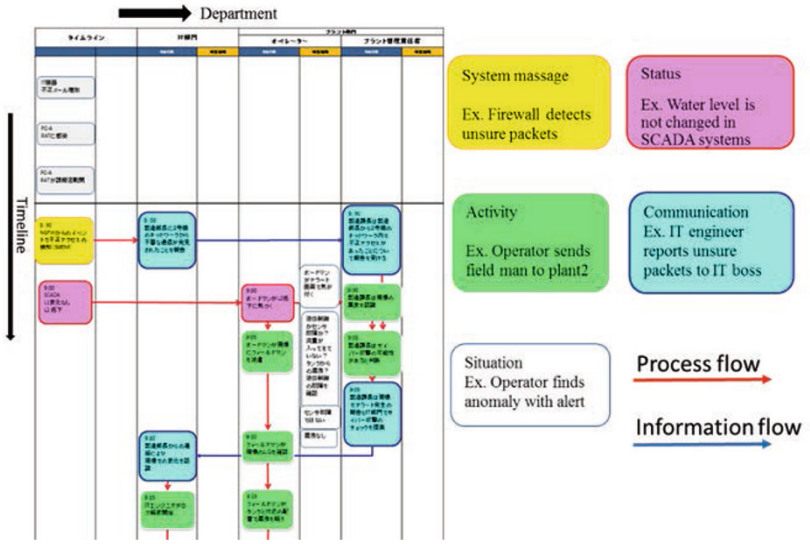


Figure2 Type2 exercise based on emergency workflow

The exercise participants will do the action the following:

1) The exercise participants are given the following information: "An alarm indicating a sudden operational abnormality is triggered, but the instruments are normal on the HMI" and "A security tool has detected an abnormality.

2) Describe the actions to be taken in the appearance where only the swim lane is shown, imagining what may happen as time passes.

3) Add a communication path for information sharing between the described actions.

The exercise participants will create a workflow by performing the tasks 1)-3), as shown in Figure 2. The exercise participants will understand the required organizational structure and information control that is the core of the response by expressing a department that needs to understand both the cyber incident signal and the process event signal on the OT side in each phase (1-3). Such a department should be the leading actor between the IT and OT sides by expressing the departments responsible for actions on the IT and OT sides and the duties involved in decision-making through communication.

These two exercises were conducted at a workshop organized by the Nagoya Institute of Technology and at the ICSCoE (Industrial Cyber Security Center of Excellence) [6] of the Ministry of Economy, Trade and Industry of Japan. As a result, the following results were obtained.

- In the Type 1 exercise, some participants said they understood the importance of an environment where communication is necessary for incident response and where rules can be established in advance so that actions can be taken before thinking.
- However, in Type1 exercise, the scenario proceeds simply by the exercise participants deciding on their actions. As a result, the exercise participants may only enjoy the exercise as a game and may not achieve the meta-learning envisioned in the exercise.
- In the Type 2 exercise, the students must imagine and respond to an incident situation in a plant or network based on the given conditions. Therefore, some say they could enhance their ability to set up temporary structures, which is necessary for formulating ICS-BCP. However, since this imagining is based on what the exercise participants have experienced so far, if they have little experience in planning the organization's incident response, they may not know what to do and may not be able to grasp the whole picture of the reaction, making the exercise less effective.
- Furthermore, since the two types of exercise focuses on organizational coordination and communication in a cyber incident, there were many cases where exercise participants who had no experience with cyber incidents could not envision the actions to be taken. In this situation, the exercise participants focused on the hard skills that they could understand or imagine, such as checking the IT system's log or the PLC's logic, and sometimes failed to achieve the essence of the exercise.

3. Card-Type Incident Response Exercise

The authors developed an exercise based on the results of our exercises described in Chapter 2 that does not depend on the experience of the exercise participants, focuses more on communication, and allows discussion of the concept of what is needed for communication, inter-organizational collaboration, and planning during an incident.
In this new exercise, participants will be given a worksheet as shown in Figure 3 and several cards with the actions taken on the front and the action's results on the back.

1) Cards

The cards are categorized into four types: action cards, situation awareness cards, information sharing cards, and attack cards, as shown below.

- Action cards:
  This card is to ensure plant safety and to continue the corporate business.
- Situation defining cards:
  This card is for recognizing the situation to select an action.
- Information sharing cards:
  This card is for thinking about information sharing that should be done when an incident occurs.
- Attack cards:
  This card is not presented during the exercise, but it is essential for the exercise's progress. This card needs to be aligned with the cyber kill chain.

In addition, to measure the exercise participants' lack of knowledge and misunderstanding, we also prepared cards with unnecessary actions in the scenario in question.

2) Worksheet

As in the Type-2 circumference, the swim lane lists the departments directly or indirectly related to the ICS-BCP, from the person in charge of the on-site operation in

an emergency to the sales and public relations departments. Exercise participants can add departments as needed. (Action cards related to that department can also be added.) However, some departments do not have a card with actions to be taken.
Exercise participants will perform the following actions:

(1) Based on the given information, infer a possible situation in the plant and ICS network.
(2) Select the actions that may be necessary as cards.
(3) Place the cards on the worksheet.

By performing the tasks (1)-(3), the incident response process is examined.
By having the actions presented in advance, the exercise participants do not have to discuss the details of the actions but rather.
The discussion can now focus on the communication content and inter-organizational collaboration required along the incident response process.

- What information is needed to initiate and implement the action and where it originates from
- The information needed to share the results of the action and its source
- Who acts and who directs organizers?

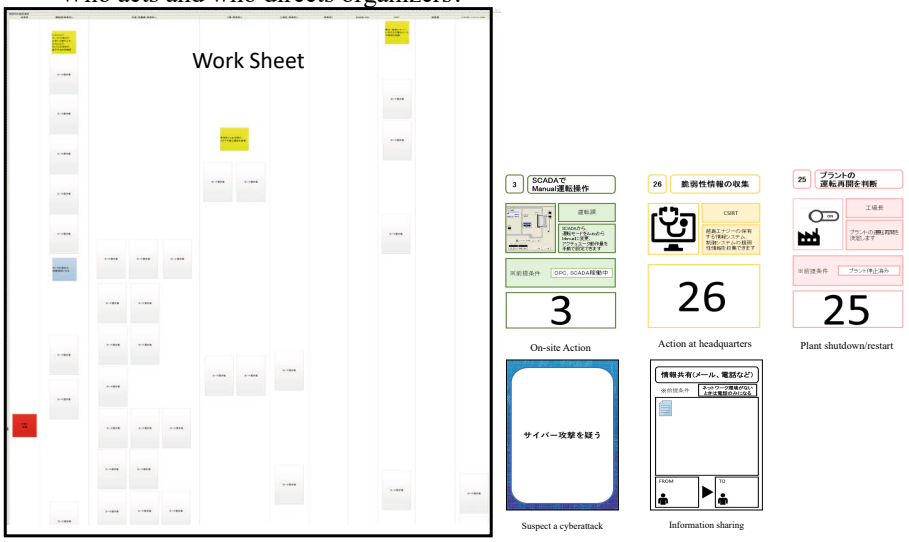


Figure3 Exercise worksheet developed on Microsoft Excel

4. Pilot Exercise

The pilot exercise using newly method was evaluated at ICSCoE. For this trial, over 100 students participated in this trial. Here are some of their comments:

- The visualization of activities by the action cards made it easier to promote discussions around the need for inter-organizational collaboration.
- I was able to think more deeply about incident response by explaining the need for adding selected action card in a logical (5W1H) manner.
- Considering the division of duties in my company, I cannot image of any action to request to the “Safety Management Department” in case of a cyber-attack. I could understand the necessity of revising the internal rules because safety measures are essential.

- The swim lanes in the worksheet included customers, control equipment vendors, and government agencies, which made me realize that incident response is not only for internal use but also for external stakeholders.
- I recognized the need to review the current safety response to recover from cyber incidents.
- The response workflow could be configured with almost the same cards regardless of the method of cyber-attack. This fact helped me to understand the effectiveness of the Consequence-driven Cyber-informed Engineering proposed by INL [7].
- This new exercise method can be customized by your company with little effort, such as changing card contents, modifying swim lanes, and adding cards.

5. Conclusions

In this paper, the authors have described the exercises we have developed to increase the pseudo experience of cyber incidents. We believe that this exercise is essential for companies to plan their ICS-BCP to mitigate damage and recover the current situation as soon as possible in the event of a cyber incident.
In the future, we will introduce a method to make people aware of possible risks due to failure to act [8] to improve resiliency to cyber incidents.

References

[1]SECURELIST by Kaspersky, Stuxnet: Zero victims, https://securelist.com/stuxnet-zero-victims/67483/

[2]BBC NEWS, US fuel pipeline hackers 'didn't mean to create problems', https://www.bbc.com/news/business-57050690

[3]CYBERSECURITY & INFRASTRUCTURE SECURITY AGENCY, MAR-17-352-01 HatMan – Safety System Targeted Malware(Update B), https://us-cert.cisa.gov/ics/MAR-17-352-01-HatMan-Safety-System-Targeted-Malware-Update-B

[4]H. Hirai, Y. Takayama, T. Aoyama, Y. Hashimoto, and I. Koshijima, 2018, Development of the Cyber Exercise for Critical Infrastructures Focusing on Inter-Organization Communication, 13th International Symposium on Process Systems Engineering(PSE 2018) (pp. 1669-1674)

[5]Y. Ota, T. Aoyama, D. Nyambayar, and I. Koshijima, Cyber Incident Exercise for Safety Protection in Critical Infrastructure, Int J. Saf. Secure. Eng. 8,246-257(2018)

[6]Information-technology Promotion Agency (IPA), Industrial Cyber Security Center of Excellence (ICSCoE), https://www.ipa.go.jp/icscoe/english/index.html

[7]Idaho National Laboratory, CONSEQUENCE-DRIVEN CYBER-INFORMED ENGINEERING, https://inl.gov/cce/

[8]E. Mizuno, Y. Saito, T. Hamaguchi, and Y. Hashimoto, 2021, Cyber Incident Response Tabletop Exercise to Raise Ownership, Proceeding of the Japan Joint Automatic Control Conference 64

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50246-3

Self-Organizing Map Based Approach for Assessment of Control Room Operator Training

Mohammed Aatif Shahab[a*], Babji Srinivasan[b,c*], Rajagopalan Srinivasan[b,c**]

[a]*Department of Applied Mechanics, Indian Institute of Technology Madras, Chennai, Tamil Nadu, 600036, India*
[b]*Department of Chemical Engineering, Indian Institute of Technology Madras, Chennai, Tamil Nadu, 600036, India*
[c]*American Express Lab for Data Analytics, Risks and Technology, Indian Institute of Technology Madras, Chennai, Tamil Nadu, 600036, India*
*babji.srinivasan@iitm.ac.in, **raj@iitm.ac.in

Abstract

Operators' knowledge during abnormal situations that are faced in chemical process industries is critical to ensure safety. Operators expand their knowledge base through training programmes that assess their comprehension and skills using simple success and failure criteria, process-based measures, and operator actions. However, these assessment techniques often overlook factors relevant to the evaluation of their cognitive capabilities such as information acquisition pattern, cognitive workload and decision-making strategy. In this work, we present a methodology for evaluating operators' performance during training that blends process-based measurements with eye-tracking-derived cognitive behaviour. Our methodology is based on Self-Organizing Map (SOM), an unsupervised neural network that allows optimum visualization of complex data. Accordingly, we trained two different SOM networks, one using the process data and the other using eye-tracking data to obtain information about operators' performance during training. Results indicate that when operators learn the process dynamics, the number of neuronal clusters hit by the process as well as operators' eye gaze trajectory decrease. The decrease in the number of clusters on SOM trained using process data indicates improved operator performance in terms of successful completion of the task and correct control action with appropriate magnitude. On the other hand, the decrease in the number of clusters hit on SOM trained using eye gaze data signifies that the operator attends to only a few regions on HMI that are critical to the current disturbance/abnormality in the process. Thus, the proposed methodology can be used to gauge the operators' learning progress during training to understand the transition from novice to expert.

Keywords: Operator Training, Self-Organizing Map, Eye-Tracking, Cognitive Behavior

1. Introduction

Control room operators are responsible for monitoring the state of the process and intervene when abnormal situations occur. This requires in-depth knowledge of the process dynamics, and therefore, operators' understanding of the process is crucial to ensure safe plant operation, higher throughput and optimal product quality. However, the complex automation and control techniques have made it challenging for the

operators to understand causal relationships related to process behavior (Liu et al., 2012).

Process industries have resorted to operator training programs that aim to impart operators with the knowledge of process dynamics. Process industries widely use operator training simulators (OTS) to train operators. An OTS provides a credible simulation of the process along with the automation systems. Various studies have sought to evaluate operator performance using OTS. In these studies, the assessment of operators' performance is primarily based on subjective evaluation and/or process and operator actions. Subjective measures are based on expert's opinions and are vulnerable to confirmation biases (Nazir and Manca, 2015). Approaches based on process and action-based metrics usually rely on metrics derived from process behavior (alarm information, deviation of process from steady-state) and operator responses (completion time, sequence of actions). However, there are hardly any approaches that focus on understanding the cognitive workload of operators, crucial to enhancing operators' abilities. Further, existing studies do not focus on the evolution of the operator performance during training. Therefore, it is necessary to develop an approach that can: provide information about the evolution of operators' performance and the cognitive workload of operators during training (Das et al., 2017).

Recent advancements in sensor technology have made it possible to objectively assess human cognitive behavior in high-risk industries (Srinivasan et al., 2019). In our previous works, we have used physiological sensors like eye tracking (Das et al., 2017) and electroencephalography (Iqbal et al., 2020) to study the cognitive components of operators' performance. For instance, we found that gaze entropy can distinguish between expert and novice operators (Bhavsar et al., 2017). Gaze entropy quantifies the spatial distribution of eye gaze on the HMI. Gaze transition entropy was found to be lower for expert operators indicating that they attend only a few regions on Human-Machine Interface, which are critical for dealing with process abnormality. Recently, Shahab et al. (2021) found that the fixation transition entropy decreases and remains at lower values when operators have learned to control the disturbance. This is because operators initially explore the state space of the process. Then, with learning, they understand causal relationships, which manifests in ordered eye gaze transitions between important regions on HMI. With the experience from our previous works, in this study, we propose to evaluate the evolution of operators' performance during training using process data, operator actions and eye-tracking data. Unlike our prior research in which specific regions on the HMI were pre-defined, this study employs an unsupervised neural network called the Self-Organizing Map (SOM) to visualize changes in operator performance with learning. Thus, the application of SOM can help eliminate human interference in operator performance analysis, and provide insights into human operator performance solely based on the process and eye-tracking information.

2. Experimental Studies and Methodology

The experimental study consisted of operators interacting with the Human Machine Interface (HMI) of an in-house chemical process simulator. The operators were asked to monitor the process and intervene during process abnormalities (using sliders provided on the HMI). Failure on the part of operators results in an automatic shutdown of the plant. The study lasted for 24 days, during which ten participants performed several

repetitions of process control tasks, henceforth called as trials. Each trial involves six different scenarios (disturbance rejection tasks). In total, these ten participants performed 81 trials for each scenario leading to a total of 486 tasks. The reader is referred to Iqbal et al. (2021) for more details on the experimental protocol. We recorded process data, alarm information, and operator action data during all these trials. In addition to these measurements, we also recorded eye-gaze data obtained from Tobii TX 300 eye tracker, at a sampling rate of 120 Hz.

We used Self-Organizing Map as a tool to dynamically assess the improvement in operators' performance with the repetition of trials. SOM can project high-dimensional data to a low dimensional space and therefore can serve as a visualization tool to observe the changes in process and operator cognitive behavior. SOMs use an unsupervised learning neural network trained with a competitive learning algorithm to create smaller subspaces. The number of neurons selected for training SOM network is given by $8\sqrt{I}$, where I is the number of samples available for training. For every sample input, "winning" neurons (i.e., neurons that most closely resemble the sample input) are identified. The winning neuron for a sample point is the one which minimizes the Euclidian distance between the neuron and the sample input. Neuronal weights are then adjusted according to the location of "winning" neuron. After several iterations of input data sets, similar neurons come closer to each other based on their weights. A suitable clustering algorithm is then applied to group the neurons into clusters. The SOM can be used for visualization using Unified Distance Matrix (U-matrix). U-matrix depicts the boundary between each pair of neurons by calculating the distance between a neuron and its neighbour. Similar group of neurons are indicated by "Valleys" which are separated by "Mountains" from dissimilar group of neurons. Interested readers are referred to Ng and Srinivasan (2008) for more details on the SOM implementation. We trained two different SOM networks, one using the process data consisting of eleven process variables (data set of size 30781 × 11) described as M^{SOM} and the other using eye-tracking data consisting of eye gaze coordinates on the HMI and gaze duration (dataset of size 97531 × 3) described as N^{SOM}. The data to train SOM networks consists of all the available data from all the tasks (486 tasks). Next, we applied a K-means clustering algorithm to group similar neurons together. The neuronal clusters on M^{SOM} represent different process operating conditions while these represent different regions of HMI on the N^{SOM}.

To assess a particular operator's performance during a task, their process and eye gaze data is projected to the SOM space. The operators' performance during the task can be visualized by the number of neuronal clusters hit by the process and the eye gaze trajectory on M^{SOM} and N^{SOM}, respectively. We hypothesize that the number of neuronal clusters hit by the process trajectory on M^{SOM} (N_{C_P}) should decrease with repetition of trials. This is because when the operator develops an understanding of the process dynamics, they use the correct control strategy to deal with the process abnormalities. Thus, the process trajectory should end in a steady-state cluster without any hit on other process operating states. Hits on additional neuronal clusters indicate incorrect control action from the operator. Similarly, it is expected that number of neuronal clusters hit by the eye gaze trajectory on N^{SOM} (N_{C_F}) should decrease owing to the attention to only a few regions on HMI related to the process abnormality. In addition to these measures, we also evaluated operators' proactive monitoring strategy, crucial for enhancing operators' ability to deal with process abnormalities. Accordingly,

we identified neuronal clusters belonging to the trend panel (area on the HMI where process variables trend is displayed). It is expected that the percentage hits on the neuronal clusters belonging to the trend panel should increase with repetition of trials. This indicates operators' ability to take time-dependent information and foresee the direction of the process. Next, we demonstrate the applicability of the proposed methodology using experimental studies.

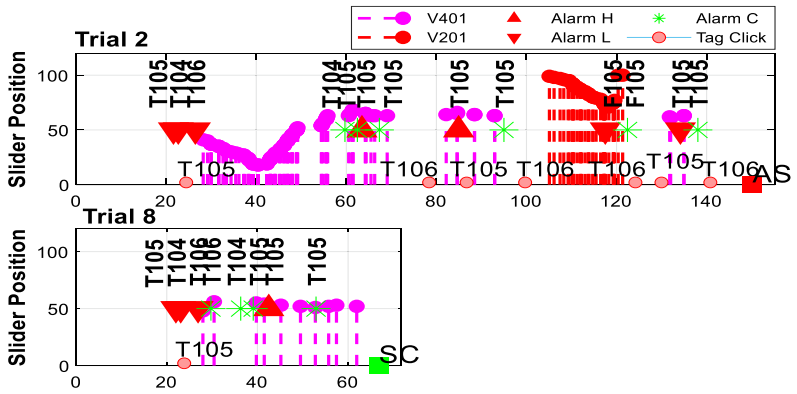


Figure 1: Operators actions with alarm information during trials 2 and 8.

3. Results and Discussion

In this section, we demonstrate the potential of the proposed methodology in capturing operators' learning progress during training using an illustrative example. Consider an operator who performed eight trials of a scenario involving a disturbance in the reflux ratio of the distillation column. The disturbance can be rectified by opening the reflux valve V401 (at least 53% open) provided on the HMI. Operator actions in the second and the last trial are shown in Fig. 1. Consider trial 2 (Fig. 1), at around 22 s, low T104, and low T105 alarms occur, followed by low T106 alarm in distillation column. Operator clicks on the tag of T105 to observe the trend of the process variable T105 and takes the first control action at around 28 s by manipulating V401; however, in the wrong direction. At around 41 s, the operator changes the direction of V401 which clears low T104 and low T105 alarms, but not T106, which results in automatic shutdown of the process. In the same trial, the operator also manipulates feed flow to the distillation column valve (V201) later, depicting that the operator keeps on exploring the possible actions which can clear the disturbance in the process. The use of multiple control actions indicates that the operator is unaware of the root cause of the disturbance. During the last trial (trial 8), as shown in Fig. 1, the operator manipulates the valve V401 by watching T105 (as shown in Fig. 1 by clicking on T105 which opens its trend) and able to clear all the alarms in the process. The manipulation of the valve V401 was very precise, and the participant increased slider opening to a maximum of 3 % during each manipulation, unlike trial 2 where the operator manipulates the slider abruptly (25% increase within short interval of time). The smooth movement of the valve in the eighth trial is an indication that the operator now understands the time constant of the process variables. Thus, with the repetition of trials, the operator learns the root cause of the disturbance along with appropriate amount of manipulation required.

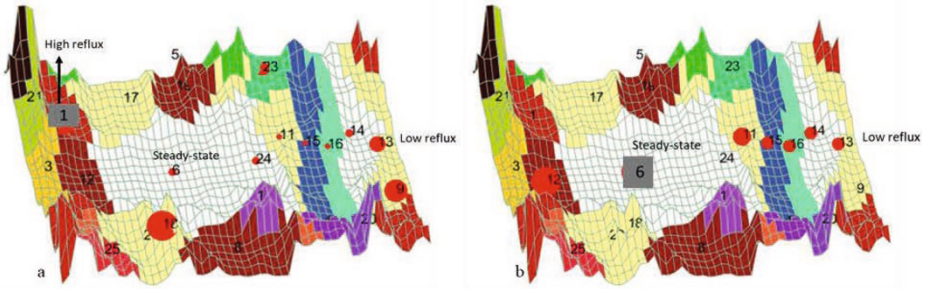


Figure 2: Process trajectory represented on SOM space during (a) Second trial (b) Eighth trial. Cluster hits are depicted by circles.

The process behavior and operator's eye gaze behaviour for the second and eighth trials are depicted in the SOM space in Fig. 2 and Fig. 3. During these trials, the process behavior and effect of control actions on the process can be observed by hits of process trajectory on neuronal clusters. The hits of the process trajectory on the neuronal clusters on M^{SOM} during the second and the eighth trial is shown in Fig. 2 (a) and Fig. 2 (b) respectively, using circles on the clusters. Rectangles on the neuronal cluster represent the end of the process trajectory. During trial 1, it can be observed that the operator's poor control action strategy leads the process to cluster 23 and 18 and further ends the process in cluster 1 (as depicted by a rectangle in Fig. 2 (a)). The total number of cluster hits by the process trajectory (N_{C_P}) during the second trial is 11. However, from Fig. 2 (b), it can be observed that in the eighth trial, the N_{C_P} decreases to 8, and the process trajectory ends in cluster 6 (steady state cluster). This indicates improved operator performance in terms of successful completion of the task and correct control action with appropriate magnitude. Fig. 3 shows operator's eye gaze behavior projected to SOM space during the second and eighth trials. It can be observed that the number of clusters hits by the operator's eye gaze trajectory (N_{C_F}) decreases to 7 in the last trial as compared to the first trial. This signifies that the operator has oriented their eye gaze pattern as per the demands of the task and directed their attention to critical information sources related to the process abnormality. The percentage of hits on neuronal clusters belonging to the trend panel cluster (clusters 4, 5 and 7) also increases from 40% in the second trial to 54 % in the last trial. This indicates operator develops a proactive monitoring strategy with learning. Thus, the proposed SOM based operator performance assessment methodology can act as a powerful technique to evaluate the effectiveness of operator training programs.

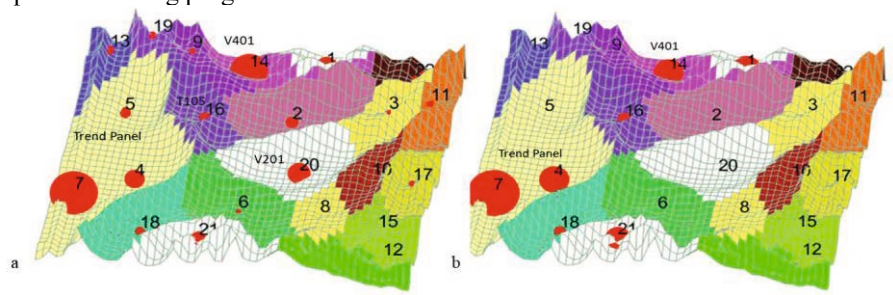


Figure 3: Operator's eye gaze trajectory represented on SOM space during (a) Second trial (b) Last trial. Clusters hits are represented by red circles.

4. Conclusions

The present work proposes an integrated process-based and operator cognitive behavior based methodology to assess operators' performance during training. The approach uses eye-tracking to obtain a trace of attention allocation of operators on the HMI. Self-Organizing Map is used to analyze the process and eye-tracking data. The evolution of cluster hits by the process, and the eye gaze trajectory on SOM can be used to track the learning progress of operators. When the operators develop correct understanding of the process dynamics, the number of clusters hit by process trajectory (N_{C_P}) and eye gaze trajectory decreases. This reflects the development of correct process comprehension and eye gaze orientation in accordance with task demands. Further, our methodology also evaluates operators' proactive monitoring strategy indicated by the percentage of hits by the eye gaze trajectory on the neuronal cluster belonging to trend panel. Our future work is intended to develop an expert operators' model using cognitive markers so as to use that as a benchmark to evaluate the performance of novice operators. The expert model can also be used to transfer knowledge to novices, thus reducing the time of training programs. Finally, a SOM can be trained using multivariate data from several physiological sensors to get deeper insights into the cognitive behavior underlying operators' expertise level.

Acknowledgments: This work is partially funded by American Express Lab for Data Analytics, Risk and Technology, Indian Institute of Technology Madras

References

Bhavsar, P., Srinivasan, B., & Srinivasan, R. (2017). Quantifying situation awareness of control room operators using eye-gaze behavior. Computers & chemical engineering, 106, 191-201.

Das, L., Srinivasan, B., & Srinivasan, R. (2017). Cognitive behavior based framework for operator learning: knowledge and capability assessment through eye tracking. In Computer Aided Chemical Engineering (Vol. 40, pp. 2977-2982). Elsevier.

Iqbal, M. U., Srinivasan, B., & Srinivasan, R. (2020). Dynamic assessment of control room operator's cognitive workload using Electroencephalography (EEG). Computers & Chemical Engineering, 141, 106726.

Iqbal, M. U., Shahab, M. A., Choudhary, M., Srinivasan, B., & Srinivasan, R. (2021). Electroencephalography (EEG) based cognitive measures for evaluating the effectiveness of operator training. Process Safety and Environmental Protection, 150, 51-67.

Liu, X.-W., Shen, X.-W., Fan, D., & Noda, M. (2012) ACP Approach-Based Plant Human–Machine Interaction Evaluation. Service Science, Management, and Engineering:, 247–264

Nazir, S., & Manca, D. (2015). How a plant simulator can improve industrial safety. Process Safety Progress, 34(3), 237-243.

Ng, Y. S., & Srinivasan, R. (2008). Multivariate temporal data analysis using self-organizing maps. 1. Training methodology for effective visualization of multistate operations. Industrial & engineering chemistry research, 47(20), 7744-7757.

Shahab, M. A., Srinivasan, B., & Srinivasan, R. (2021). Evaluating Control Room Operator Training Outcomes Through Eye Gaze Augmented Multi-Scale Data. In Computer Aided Chemical Engineering (Vol. 50, pp. 1307-1312). Elsevier.

Srinivasan, R., Srinivasan, B., Iqbal, M. U., Nemet, A., & Kravanja, Z. (2019). Recent developments towards enhancing process safety: Inherent safety and cognitive engineering. Computers & Chemical Engineering, 128, 364-383.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50247-5

Digital Twin of Alkaline Water Electrolysis Systems for Green Hydrogen Production

Yongbeom Shin[a], Jongyeon Oh[a], Dongkuk Jang[b], Dongil Shin[a,b*]

[a]*Dept. of Chemical Engineering, Myongji University, Yongin, Gyeonggido 17058, Korea*
[b]*Dept. of Disaster and Safety, Myongji University, Yongin, Gyeonggido 17058, Korea*
dongil@mju.ac.kr

Abstract

The digital twin which supports data-based decision making, optimization, control and anomaly detection and diagnosis, can contribute to the improvement of sustainability, agility and productivity in water electrolysis system, which is expected to provide to reduce green hydrogen production cost. In this study, we propose a digital twin for a 500kW alkaline water electrolysis (AWE) to be built at the Saemangeum Renewable Energy National Demonstration Complex in Korea to reduce green hydrogen production cost through optimal operation of AWE system. A simulation model, which is the basis of the digital twin, was developed with Python and gPROMS, and the system efficiency of the AWE process according to pressure was analyzed comparing between excluding the compression process and including the process of hydrogen compression to 200 bar. The optimum operating pressure with the compression showed at 10-30 bar. At high pressure, process equipment cost becomes higher, therefore, it is essential to consider hydrogen compression to the storage pressure in order to decide the optimal operating conditions.

Keywords: Alkaline water electrolysis, digital twin, optimal operating conditions.

1. Introduction

As the spread of renewables, it is required to increase the storage capacity of surplus energy due to intermittent characteristics, and a technology with low energy loss and high efficiency in the storage is required. Green hydrogen (by P2G) is an important eco-friendly energy source for storing surplus energy. However, its production cost is about five times higher than hydrogen produced by other methods (natural gas reforming), and alkaline water electrolysis (AWE) systems which have the highest technological maturity still have high hydrogen production costs.

Many researchers have interest of methods for reducing green hydrogen production cost included scale-up of water electrolysis system, development of high-efficiency electrode or separator material, and operation optimization. Ulleberg (2003) developed a mathematical model of an AWE system and demonstrated it for the photovoltaic-hydrogen energy plant in Julich, and Jang (2021a) proposed high-pressure operation to reduce the energy consumption of the process because hydrogen produced from the AWE system is compressed to high pressure for storage. Energy consumption and efficiency were analyzed through a model developed by Aspen Plus. Then, Jang (2021b) tried to optimize the operating conditions by analyzing the effect of temperature on the system using the developed Aspen Plus model.

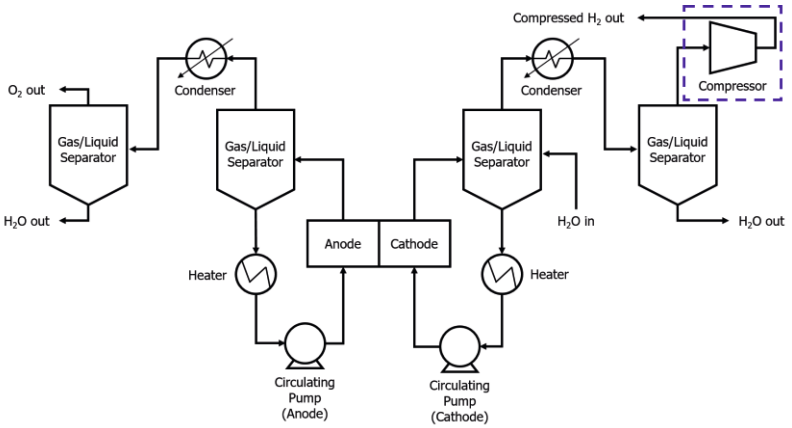


Figure 1. Diagram of the AWE system including the compressor

In this study, we propose a digital twin that integrates functions, water electrolysis process simulation, control, monitoring, optimization, to reduce green hydrogen production cost. A digital twin will be built for a 500kW AWE to be built at the Renewable Energy National Demonstration Complex in Saemangeum, Korea.

In this study, a simulation model, which is the basis of the digital twin, was developed in Python and gPROMS, and the system efficiency of the AWE process according to pressure was analyzed. For the simulation, gPROMS' gML Process and Electrochemical Cell Reactor (ECR) module are used, but the ECR module does not reflect the characteristics of the liquid electrolyte of the AWE, so the cell model in Python is developed. The process simulation model is proposed using custom modeling template of gPROMS to apply the AWE cell model. The process of compression to 200 bar, the load pressure of the trailer truck, is considered as a part of the water electrolysis system. To analyze the impact of the compression to the system efficiency, the optimal operating conditions of the AWE system are compared with the results of excluding compression.

2. Alkaline Water Electrolysis System

Figure 1 is a process diagram of an alkaline water electrolysis system for green hydrogen production. The alkaline water electrolysis system consists of a water electrolysis stack, gas/liquid separator, condenser, heater, pump, and compressor. Oxygen is generated at the anode, hydrogen is generated at the cathode, and the amount of water consumed to maintain the electrolyte concentration is supplied to the separator on cathode side. The generated gas is separated through a gas/liquid separator, and water is removed in the condenser. The hydrogen produced is then compressed to a pressure of 200 bar for loading into a trailer truck.

3. Model Description

3.1. AWE cell model in Python

The electrochemical reaction module of gPROMS, ECR, does not reflect the effects of liquid electrolytes, such as the bubble effect of AWE systems. An AWE cell model to apply in gPROMS is required, cell modeling is performed using Python.

3.1.1. Reversible voltage

Water electrolysis is a non-spontaneous reaction and requires the supply of electrical energy. The reversible voltage, which is the theoretical electrical energy required for the electrolysis of water, can be calculated by the Nernst equation, and is as follows:

$$V_{rev} = V_{rev}^{\circ} + \frac{RT}{2F}\ln(\frac{(P-P_{H_2O})^{1.5}P_{H_2O}^{*}}{P_{H_2O}}) \quad (2)$$

where V_{rev}° is the reversible voltage at the standard conditions, P_{H_2O} is the vapor pressure of the KOH solution, $P_{H_2O}^{*}$ is the vapor pressure of pure water, $P, T, R, and\ F$ are the operating pressure, temperature, gas constant, and Faraday constant.

The first term represents reversible voltage at the standard conditions can be expressed as temperature (Hammoudi, 2012):

$$V_{rev,T}^{\circ} = 1.50342 - 9.956 \times 10^{-4}T + 2.5 \times 10^{-7}T^{2} \quad (3)$$

where T is the operating temperature, m is the molar concentration, w is the w weight percent concentration of KOH solution.

3.1.2. Activation overvoltage

For the water electrolysis reaction of an actual alkaline water electrolysis cell, electrical energy above the reversible voltage is required due to resistance, and it can be expressed as follows:

$$V_{cell} = V_{rev} + V_{act} + V_{ohm} \quad (4)$$

where V_{act} is the activation overvoltage, V_{ohm} is the ohmic overvoltage.

Activation overvoltage the energy loss due to the decrease of the electrochemical reaction rate at the anode and cathode, which can be expressed by the Butler-Volmer equation:

$$V_{act} = \frac{RT}{zF\alpha_{an}}\ln\left(\frac{i}{i_{o,an}}\right) + \frac{RT}{zF\alpha_{ca}}\ln\left(\frac{i}{i_{o,ca}}\right) \quad (5)$$

where α is the charge transfer coefficients, i is the current density (A/cm^2), i_0 is the exchange current density.

When hydrogen and oxygen gases are generated, the effect of gas bubble coverage in the liquid electrolyte can be expressed in terms of temperature and pressure as follows (Jang, 2021a):

$$\theta = 0.23(i)^{0.3}(\frac{T}{T_{ref}}\frac{P_{ref}}{P})^{\frac{2}{3}} \quad (6)$$

Activation overvoltage can be calculated by applying the bubble coverage coefficient of Eq.(6) to Eq.(5).

3.1.3. Ohmic overvoltage

Ohmic overvoltage is energy loss due to resistance and can be expressed through Ohm's law as follows (Jang, 2021a):

$$V_{ohm} = I \times (R_{ca} + R_{an} + R_{KOH} + R_{Sep}) \quad (7)$$

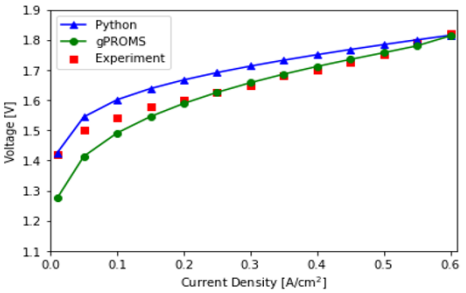


Figure 2. Model validation results by cell voltage

where I is the current of the electrolysis, $R_{ca}, R_{an}, R_{KOH}, R_{Sep}$ are the resistance of the cathode, anode, KOH solution, and separator.

The electrode resistance and electrolyte resistance are calculated based on the conductivity.

3.1.4. Hydrogen production rate

Hydrogen production can be expressed as follows:

$$\dot{n_{H_2}} = \eta \frac{I}{zF} N_{cell} \tag{8}$$

where η is the Faraday efficiency, z is the number of electrons transferred per reaction (z=2), N_{cell} is the number of cells in the stack.

3.2. AWE system modelling in gPROMS

The AWE cell model is integrated into the ECR module of gPROMS using a custom model template, and a simulation of the AWE system is constructed using the gML Process module. The energy consumption, stack power and balance of plant (BOP), of the AWE system is:

$$W_{System} = W_{Stack} + W_{Heater\_an} + W_{Heater\_ca} + W_{Condenser\_O2} + W_{Condenser\_H2} + W_{pump\_an} + W_{pump\_ca} + W_{Compressor} \tag{9}$$

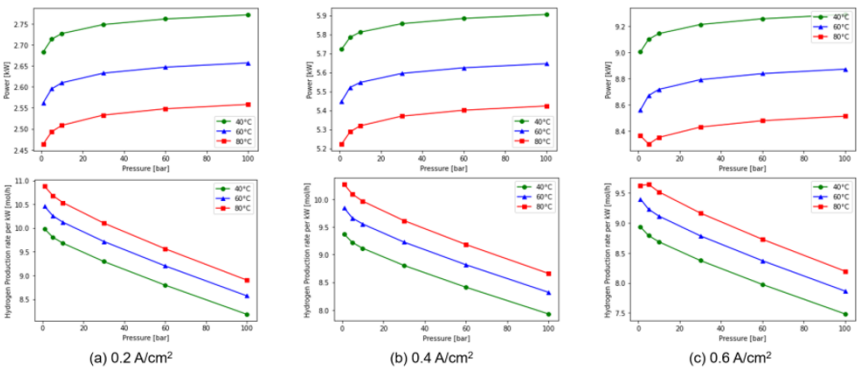


Figure 3. Power and hydrogen production efficiency without compression

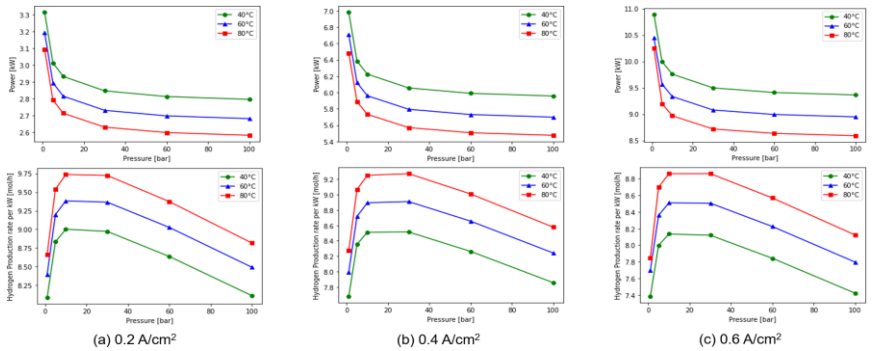


Figure 4. Power and hydrogen production efficiency with compression

4. Simulation Results

4.1. Model validation

The developed model was verified through cell voltage according to current density for AWE cells developed by Korea Institute of Energy Research (KIER), Figure 2 is the results of model validation. The Python-based cell model showed high accuracy (R^2=0.813). gPROMS utilized the cell model also indicated high accuracy (R^2=0.8), and revealed an R-squared of 0.985 at a current density of 0.2-0.6 A/cm^2.

4.2. AWE system simulation

Assuming that the hydrogen produced from the AWE system is compressed to 200 bar, 3 current densities (0.2, 0.4, 0.6 A/cm^2), 3 temperatures (40, 60, 80 ℃), 6 pressures (1, 5, 10, 30, 60, 100 bar), a total of 54 cases were simulated.

4.2.1. Without compressor power

Figure 3 shows the total power consumption and hydrogen production efficiency when the compression process of produced hydrogen is not considered. At 0.2, 0.4 A/cm^2, as the pressure increases, the power consumption increases and the hydrogen production efficiency decreases. At 0.6 A/cm^2, when the pressure increases from 1 bar to 5 bar, the power consumption decreases, which is a phenomenon that temporarily decreases because the decrease effect of the bubble effect is greater than the increase of the reversible voltage caused by the increase in pressure.

Table 1. Stack, total BOP, compressor, and total power consumption at 0.6 A/cm^2, 80℃

Pressure [bar]	Stack Power [kW]	Total BOP [kW]	Comp. Power [kW]	Total Power [kW]	Hydrogen Production [mol/h]	Hydrogen Production per kW
1	7.893	0.470	1.892	10.255	80.493	7.849
5	7.879	0.422	0.900	9.201	80.043	8.700
10	7.934	0.418	0.619	8.971	79.482	8.860
30	8.016	0.414	0.289	8.719	77.262	8.861
60	8.067	0.412	0.157	8.636	74.001	8.569
100	8.105	0.408	0.077	8.591	69.775	8.122

4.2.2. With compressor power

Figure 4 indicates the total power consumption and hydrogen production efficiency with the compression power consumption. In all cases, at the higher pressure, the lower power consumption is indicated because the power consumption for compression up to 200 bar occupies a high proportion in the total power. The hydrogen production per kW of energy consumption presented optimal values at 10 to 30 bar, as shown in Table 1. Compared with the case with no compression at design condition (0.6 A/cm^2, 80 °C), the optimum operating pressure is increased by 6 times from 5 bar to 30 bar.

5. Conclusions

The digital twin supports data-based decision making, optimization, control, and anomaly detection and diagnosis, so it will contribute to the improvement of sustainability, agility and productivity in chemical plant. In this study, a AWE simulation model was developed using Python and gPROMS, which is basis of the development of digital twin, it is expected to improve the productivity of the AWE system resulting in reducing green hydrogen production cost.

The effect of pressure on the performance was evaluated by the developed model. The optimal operating conditions of the AWE system were compared between excluding the compression and including the compression to 200 bar to analyze the impact of the compression process. When excluding the process of compression, it showed the optimal operating pressure range of 1 to 5 bar, and including compression demonstrated the optimal range of 10 to 30 bar. However, process equipment cost becomes higher at high pressure, therefore, it is essential to consider hydrogen compression to the storage pressure, including the equipment cost, in order to decide the optimal operating conditions.

References

M. Hammoudi, C. Henao, K. Agbossou, Y. Dube, M. L. Doumbia, (2012), "New multi-physics approach for modelling and design of alkaline electrolyzers", *Int. J. Hydrog. Energy*, **37**, 13895-13913

D. Jang, H. Cho, S. Kang, (2021a), "Numerical modeling and anlaysis of the effect of pressure on the performance of an alkaline water electrolysis system", *Appl. Energy*, **287**, 116554

D. Jang, W. Choi, H. Cho, W. C. Cho, C. H. Kim, S. Kang, (2021b), "Numerical modeling and analysis of the temperature effect on the performance of an alkaline water electrolysis system", *J. Power Sources*, **506**, 230106

O. Ulleberg, (2003), "Modeling of advanced alkaline electrolyzers: a system simulation approach", *Int. J. Hydrog.*, **28**, 21-33

Acknowledgements

This research was partly supported by Korea Institute for Advancement of Technology(KIAT) grant funded by the Korea Government(MOTIE)(P0008475, Development Program for Smart Digital Engineering Specialist) and Korea Institute of Energy Technology Evaluation and Planning(KETEP) grant funded by the Korea government(MOTIE)(20207200000070, Development of performance standardization and operation risk estimation for renewable energy-linked alkaline water electrolysis hydrogen production systems using digital twins).

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50248-7

Cyber Security Risks of aspects of operations of OPC Unified Architecture

Mariko FUJIMOTO[a], Takuho MITSUNAGA[b], Tatsuki MATSUZAWA[a], Yoshihiro HASHIMOTO[a]

[a] *Nagoya Institute of Technology, Aichi 466-8555, JAPAN*
[b] *Toyo University, Tokyo 115-0053, JAPAN*
m.shigemori.037@stn.nitech.ac.jp

Abstract

OPC Unified Architecture (OPC UA) is a new OPC standard that is the successor of OPC Classic, and it has various security features. Generally, OPC UA is considered to be secure, but if it is not appropriately implemented or configured, there might be cyber risks. In this research, evaluate the cyber risks of OPC UA using several OPC UA products. We focus on security features of OPC UA: application authentication, user Authentication, encryption, and signing. We will show the evaluation result under each typical security configuration in several OPC UA products, then introduce points of attention for OPC UA security configurations and operations based on the evaluation result. Cyber risks and corresponding measures described in this research should be conscious by OPC UA product venders, system integrators and operators of Industrial Control System.

Keywords: OPC UA, Cyber security, Penetration test, Operational Technology

1. Introduction

OLE[1] for Process Control (OPC)[2] is the interoperability standard for reliable data exchange in industrial automation. It ensures the seamless flow of information among devices from multiple vendors. OPC Classic (OPA Data Access) is a legacy OPC standard, and its security has been under discussion. Although a new OPC standard OPC Unified Architecture (OPC UA) was released in 2016, OPC Classic is still widely used.

OPC UA is a recommended communication standard in Reference Architecture Model Industrie 4.0 (RAMI4.0)[3], and all devices will be connected through OPC UA. Since OPC UA has many security features, it is recommended to migrate to OPC UA as early as possible. However, since OPC UA has also potential security risks if it is not appropriately configured and operated, recognizing security risks and corresponding measure of OPC UA are very important. In this research, we introduce the flexibility of OPC UA security settings and cyber risks in aspects of operation under each security setting. We prove cyber risks through penetration tests against several OPC UA

[1] Object Linking and Embedding: A technology that allows embedding and linking to documents and other objects, developed by Microsoft

[2] OLE for Process Control: A set of standards developed by a joint collaboration of leading automation industry suppliers

[3] A three-dimensional consolidation of the most important aspects of Industrie 4.0, ensuring that all participants in Industrie 4.0 share a common perspective and build a common understanding

commercial products, then propose recommended OPC UA security configurations and operations. We focus on the following security features of OPC UA.

-Application authentication: Authentication of OPC UA client application using application certificates

-User Authentication: Authentication of users of OPC UA clients using passwords or user certificates

-Encryption and signing of OPC UA communication

2. Related Research

R. Huang et al. (2010) discussed OPC UA system security issues from the views of the OPC UA application environment and OPC UA communication security. L. Roepert et al. (2020) discuss methods to discover OPC UA servers, test their authentication, obtain their configuration, and check for vulnerabilities. Both R. Huang et al. (2010) and L. Roepert et al. (2020) did not focus on cyber risks and potential attack scenarios. J. Polge, et al. (2019) identified the threats and countermeasures that may occur/be applied when using OPC-UA and proved the impact of the eavesdropping and message flooding attacks on an OPC-UA application using a real testbed. However, they evaluated an encryption suite that has average security strength. W. Matsuda et al. (2021) proved cyber risks of OPC UA using the strong encryption suite, but they evaluated only one product, detailed consideration on the result is not conducted. For the above reasons, we evaluate the cyber risks of OPC UA including strong security features through penetration tests using several OPC UA products and discuss the result.

3. Proposed approach

We evaluate the security features (authentication and encryption) of OPC UA through penetration tests using several OPC UA products and introduce security practices based on the test result.

3.1. Attack scenarios

The purpose of the attack scenario is to prove the cyber risks of attacks inappropriate implementation of OPC UA. We assume that the pentesters have been intruded into the legitimate OPC UA client PC and started attacks from the legitimate OPC UA client PC assuming that the PC had been infected with the malware. We make the hypothesis that if security configurations of OPC UA are not appropriately implemented, there could be the following security risks.

1. **Unauthorized connection**: If attackers can get information to connect OPC Server, they could connect to the OPC Server from an unauthorized client.
2. **The decryption of OPC UA message**: Forward security is a feature of specific key agreement protocols that gives assurances that session keys will not be compromised even if long-term secrets used in the session key exchange are compromised. According to W. Matsuda et al. (2021) OPC UA specification

does not refer to forward security. Therefore, if attackers could steal secret keys for OPC UA, they could decrypt the encrypted communication of OPC UA.

3.2. Evaluation environment

Table 1 shows the evaluation target OPC UA products, and Table 2 shows used tools for evaluation. For security and privacy reasons, we do not describe each product name and vendor used in the evaluation. Table 3 and Table 4 show evaluation target OPC UA security features.

Table 1 OPC UA server products for evaluation

Product	**Product usage**
A	PLC with OPC UA server unit
B	OPC server software including OPC UA server
C	An industrial middleware including OPC UA server

Table 2 Tools used for evaluation

Product name	**Description**	**Purpose**
UaExpert	Free OPC UA client application	For legitimate OPC UA client
Python OPC-UA	Open source OPC UA server and client application	For malicious OPC UA client (attacker)
Wireshark	Network protocol analyser	For snooping communications

Table 3 Security Mode of OPC UA

Mode name	**Description**
None	Setting which does not need any message signature and encryption
Sign *1	Setting which generates message signature using the private key of senders
Sign & Encrypt *1	Setting which generates message signature using the private key of senders, and encrypts messages using the private key of receivers

*1: We evaluate a strongest policy, "Basic256Sha256".

Table 4 User authentication type of OPC UA

Setting	**Description**
Anonymous	Setting which does not need any user authentication
ID and password (ID / pass)	Setting which authenticates the client using the password
User certificate	Setting which authenticates client using user certificate

4. Evaluation Result

4.1. Unauthorised connection to OPC UA server

Table 5 shows the result of the unauthorized connection to the OPC UA Server for each security setting condition. Without depending on products, it was possible to connect OPC servers from computers and applications that are not used in legitimate operations if attackers were able to get the necessary information. After the successful unauthorized connection, pentesters succeeded to read and changing each parameter value of the OPC servers.

Table 5 Result of unauthorized connection to OPC UA servers

OPC UA server setting		**Conditions for success of attack**		**Encryption**
Security mode	**User authentication**	**Client private key & certificate**	**Client credential**	
None	Anonymous	Unnecessary	Unnecessary	No
	ID / pass	Unnecessary	Mandatory (ID / pass)	No
	User certificate*2	Unnecessary	Mandatory (User certificate)	No
Sign	Anonymous	Mandatory	Unnecessary	No
Sign& Encrypt	Anonymous	Mandatory	Unnecessary	Yes
	ID / pass	Mandatory	Mandatory (ID / pass)	Yes
	User certificate	Mandatory	Mandatory (User certificate)	Yes

*2: Evaluated only for 1 product that supports user certificate.

The followings are discussions on the result.

- When security mode is None and Anonymous is allowed, pentesters can connect to the OPC server without any condition.

-If user authentication is enabled, pentesters need a legitimate user ID and password or certificate. Note that user ID is sent in plain text when the security mode is None, and some products do not support user certificates.

When security mode is "Sign" or "Sign&Encrypt", pentesters need the application certificate and corresponding private key of the legitimate OPC UA client. Note that communication is not encrypted when the security mode is "Sign".

- Application certificates and corresponding private keys are located in the installation directory of the OPC UA client application. Thus, pentesters who intruded into the OPC UA client computers could easily obtain them.

4.2. Decryption of OPC UA message

Table 6 shows the result of the decryption of the OPC UA message. Without depending on products, it was possible to decrypt OPC UA messages if attackers were able to get the client's and server's private keys. After the success of decryption and decoding, pentesters succeeded to extract the parameter values and node IDs of the OPC servers.

Table 6 Result of decryption of OPC UA message

OPC UA server setting		**Conditions for success of attack**			**Decryption**
Security mode	**User authentication**	**Client private key**	**Server private key**	**Client credential**	
Sign& Encrypt	Anonymous	Mandatory	Mandatory	Unnecessary	Success
	ID / pass	Mandatory	Mandatory	Unnecessary	Success

The followings are the decryption procedures.

1. Snoop communications between OPC UA clients and servers.
2. Stole OPC server private keys stored in the products installation directory.
3. Stole OPC client private keys stored in the legitimate OPC UA client PCs.
4. Extract and decrypt clients' and servers' nonce from snooped messages. nonces are included in the Open Secure Channel request and response, and they were encrypted by the server's / client's secret keys.
5. Make symmetric keys from clients' and servers' nonce.
6. Decrypt OPC UA Secure Conversation messages using the symmetric keys.
7. Decode parameter values from decrypted messages.

The followings are discussions on the result.

If attackers can obtain secret keys of the OPC UA server and client, they could decrypt OPC UA messages. However, since the message is encoded, attackers should get information about the OPC UA message structures and decode data.

5. Recommendation for secure practice using OPC UA

This section describes OPC UA security practices based on the evaluation result.

- Set Security Mode to Sign & Encrypt to enable digital signature and encryption for OPC UA messages. "Sign" reduces the risk of unauthorized connection but note that communication is not encrypted.

-Even if the OPC server does not configure with "Sign and encrypt", it might be possible to encrypt messages if the client connects with "Sign and encrypt" in some products.

- Disable anonymous access and enable user authentication for OPC UA. Then use supported user authentication methods (ID and password, certificate, etc.) The ID and password, user certificate should be managed in a secure way.

- Many OPC UA applications provide functions to export private keys and certificates and change the security level of the OPC UA. Thus, access control or authentication of OPC UA applications and configuration tools of OPC UA should be enabled to prevent attackers to steal sensitive information or changing security configuration maliciously.

- Manage private keys and certificates of both OPC UA servers and clients in a secure way. For instance, private keys should be protected with passwords. OPC server and client application vendors should support private keys protected by passwords.

If disclosure of private keys is suspected, security risks can be reduced by invalidating the corresponding certificate in OPC UA client and server applications.

- When the OPC UA server and client firstly communicate, operators should carefully confirm each certificate is issued by a trusted host. That is because once certificates are trusted, they are used until certificates are regenerated or revoked. Some OPC UA products provide options to skip the verification or automatic acceptance of certificates, but they should not be enabled.

- The cipher suites for signature and encryption depend on security policy. Use the recommended security policy in the OPC UA specification.

- Many OPC UA applications store certificates and private keys as files without protection. Thus, computers that run OPC UA servers and clients should be protected in OS level and network-level security in addition to the application security.

6. Conclusions

In this paper, we evaluated the cyber risks of OPC UA through penetration tests against several OPC UA products. OPC UA is useful for improving security, but there could be cyber risks if security features are not configured appropriately. We introduced the security practice in aspects of operations of OPC UA based on the penetration test result.

References

R. Huang, F. Liu and D. Pan, 2010, "Research on OPC UA security," 2010 5th IEEE Conference on Industrial Electronics and Applications, pp. 1439-1444

W. Matsuda, M. Fujimoto, Y. Hashimoto and T. Mitsunaga, 2021, "Cyber Security Risks of Technical Components in Industry 4.0," 2021 IEEE International Conference on Omni-Layer Intelligent Systems (COINS), pp. 1-7

B. Mihir and Y. Bennet, 2003, Forward-Security in Private-Key Cryptography, Springer Berlin Heidelberg

J. Polge, J. Robert and Y. L. Traon, 2019, "Assessing the impact of attacks on OPC-UA applications in the Industry 4.0 era," 2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC), pp. 1-6

L. Roepert, M. Dahlmanns, I. Berenice Fink, J. Pennekamp, M. Henze, 2020, Assessing the Security of OPC UA Deployments, the 1st ITG Workshop on IT Security (ITSec)

Z. Veichtlbauer, Armin and Ortmayer, Martin and Heistracher, Thomas, 2017, OPC UA integration for field devices, 2017 IEEE 15th International Conference on Industrial Informatics (INDIN)

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50249-9

Managing Experimental-Computational Workflows in Robotic Platforms using Directed Acyclic Graphs

Federico M. Mione[a], Alexis N. Silva[a], Martin F. Luna[a], M. Nicolas Cruz B.[b,c], Ernesto C. Martinez*[a]

[a]*INGAR (CONICET-UTN), S3002GJC Santa Fe, Argentina*
[b]*KIWI-biolab, Bioprocess Engineering, TU Berlin, Ackerstrasse 76, 13355 Berlin, Germany*
[c]*DataHow AG, Zürichstrasse 137, 8600 Dübendorf, Switzerland*
ecmarti@santafe-conicet.gov.ar

Abstract

Robotic platforms can gather informative data sets to accomplish different modeling or optimization goals for bioprocess development by resorting to on-line redesign of multiple parallel experiments. For reproducible data analysis is key to formally represent and manage experimental-computational workflows in high-throughput experimentation by enforcing FAIR principles. To represent workflows of a robotic platform, directed acyclic graphs (or DAGs) are proposed. Computational implementation of DAGs using open-source software (Apache Airflow) not only helps FAIRizing data and experimental protocols but also obliges making explicit all methods, models, assumptions and hyperparameters used to carry out modeling and optimization tasks. Model-based productivity optimization of a bioprocess based on data from nine fed-batch parallel cultivations is used as an example. Data generated in the parallel experiments are first used to re-estimate online the model parameters and the updated model is used to optimize the feeding profile. Managing experimental-computational workflows as DAGs in the Airflow ecosystem using containers is key to foster the use of FAIR principles in modeling and optimization, and to facilitate access/reuse of costly experimental data.

Keywords: bioprocess development, FAIR principles, high-throughput experimentation, online experimental redesign, workflow automation.

1. Introduction

Reproducible results are key to compare computational methods and validate experimentally solutions found and their related models. However, in the recent survey work of Baker (2016) it was revealed that more than 70% of researchers have been unsuccessful in reproducing the outcomes of research experiments made by others and more than 50% cannot reproduce solutions and data obtained by their own research studies. For reproducible bioprocess development, experimental data used for predicting the dynamic behaviour of genetically modified microorganism and the effect of different combinations of process parameters on the productivity levels must be obtained using FAIR principles (Celebi et al., 2021). To this aim, these principles must be accounted for in the specification of all protocols and workflows used to generate data sets for different modeling or optimization goals by resorting to on-line (re)design of parallel experiments (Haby et al., 2019). This work deals with FAIRizing experimental workflows involving

tasks such as feeding, sampling, at-line analytical processing, etc., which generates data. These workflows are intertwined with data processing pipelines for (re)estimating model parameters, experimental execution monitoring and on-line redesign which must also be FAIR. For widespread usage of FAIR principles in bioprocess development, experimental-computational workflows are represented here as directed acyclic graphs (DAGs) that make experimental protocols reproducible within the Apache Airflow environment (Harenslak and de Ruiter, 2021) using software containers (Dockers) that foster portability and reusability of data, methods and experimental protocols.

2. Experimental-computational Workflows

The FAIR principles (Wilkinson et al., 2016), describe a set of requirements for data management and stewardship to make research data **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable. The FAIR principles are a set of guidelines that aim to maximize the value and usefulness of experimental data and highlight the importance of making experimental protocols and workflows digital objects findable and reusable by others. To this aim, the combination of DAGs with a Common Workflow Language in the Apache Airflow ecosystem is an appealing alternative for introducing FAIR principles in high-throughput experimentation facilities for bioprocess development.

Experimental-computational workflows describe the complex multi-step methods that are used for experimental design, data sampling, data preparation, model building, and active learning that led to new data products with maximum information content. Thus, this type of workflows describes a process (a set of activities) for computational or physical tasks, where different parts of the process (the tasks) are interdependent, e.g., a task can start processing after its predecessors have been (partially) completed and where data flows between tasks define the synchronization constraints. In robotic platforms for high-throughput experimentation, experimental protocols (feeding, sampling, at-line sample processing, etc.) which generates data are intertwined with computational tasks involving update of model parameters using new data and online redesign of the parallel asynchronous experiments to make the resulting data more informative for a given purpose, such as productivity optimization, strain screening or model selection.

In Fig. 1, an abstract DAG for parallel experimentation in a robotic platform involving several mini bioreactors (MBRs) arranged as subsets (columns) is shown. Cultivations can run asynchronously, but sampling must comply with some precedence constraints due to robot operation limitations. Also, for model update and online redesign data from samples for all MBRs or a group thereof must be known (a synchronization constraint). Depending on the robotic facility, the overall experiment structure can be represented at different abstraction levels using DAGs that account, to the desired level of detail, for all precedence and synchronization constraints between macro-tasks such as sample collection or liquid handling blocks at a more abstract level. Each macro-task is also internally described by a DAG where simple tasks are related to liquid handling for substrate/medium feeding or pH/Dissolved Oxygen Tension (DOT) control (see Haby et al., 2019, for details).

To guarantee FAIR data and experimental-computational workflows, a shared database is proposed for interfacing a computational pipeline that manages the parallel experiments in the robotic platform (bottom) with a pipeline of physical macro-tasks for sample collection and liquid handling (top) as it is show in the DAG of Fig. 2. Once the dummy node “Start” triggers both “parallel cultivations” and “experiment execution” nodes, the

sampling times are calculated in the "Get Sample" node, which in turn triggers the execution of both the "sample collection" node and the sensing node "Get Sample Results". Once samples taken have been analysed, resulting experimental data as well as experimental design parameters related to sampling times and liquid handling tasks are saved in a local database. As soon as sampled data can be retrieved from the shared database, the "Parameter Update" node is executed followed by the node "Online Redesign," which in turn triggers the "Liquid Handling" macro-task that executes liquid handling tasks as a macro-action or block until the next change to the experimental planned tasks in the remaining part of the experiment based on new data.

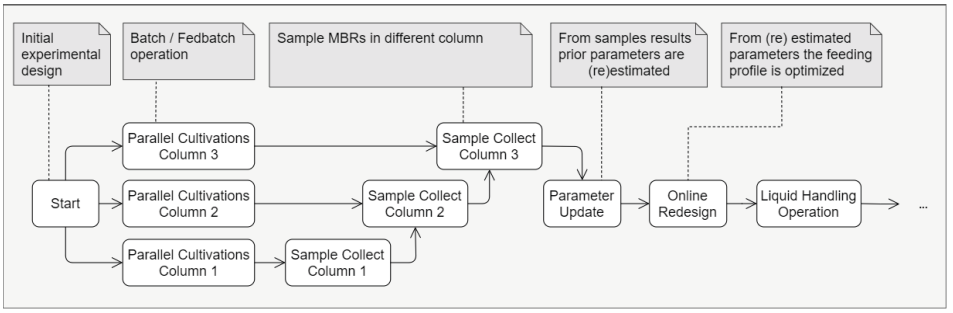


Fig. 1. Simple DAG to represent macro-tasks precedence/synchronization constraints

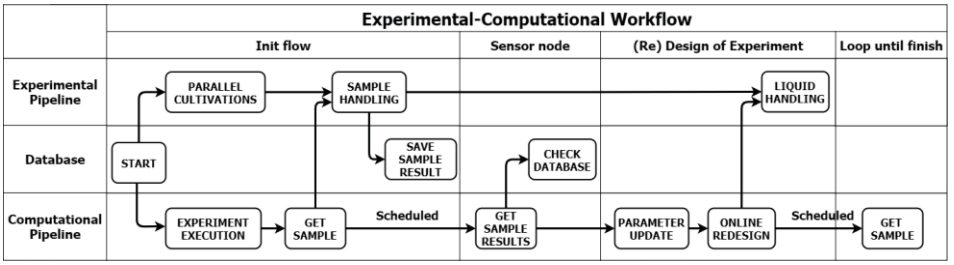


Fig. 2. An abstract DAG that integrates experimental and computational pipelines

3. Case study

To illustrate the advantages of the proposed approach, a simple robotic platform made up of nine MBRs involving the fed-batch cultivations of a microorganism are simulated using the pyFOOMB package (Hemmerich et al., 2021). More specifically, the Example 7 in the repository at https://github.com/MicroPhen/pyFOOMB will be used. As there exist uncertainty in the value of two model parameters μ_{max} and $Y_{P/X}$, the objective is to generate data which is maximally informative to define an optimal exponential feeding profile of the substrate S which is parameterized by t_F and μ_{Set} as follows:

$$F = \begin{cases} 0 & t < t_F \wedge V_L > V_{L,max} \\ \dfrac{c_{s,0} \cdot V_{L,0} \cdot \mu_{set}}{c_{S,F} - c_s} \cdot e^{\mu_{set}(t - t_F)} & t \geq t_F \end{cases} \quad (1)$$

Thus, there is an initial batch phase until t_F which is followed by a fed-batch mode of operation where the substrate exponential feeding rate is defined by μ_{Set}. The

concentration parameters in Eq. (1) are related to the initial, in the feed stream and current substrate concentrations, respectively, whose values are assumed constant as well as the initial volume of each bioreactor. The model parameters for the *in-silico* simulator used for generating sampled data are given in Table 1. Also, in Table 1 are the assumed *a priori* values for model parameters before collecting any new data in the simulated parallel experiments. The goal of the experimental design is to generate data to estimate the optimal feeding profile for maximizing the final concentration of the product once the maximum volume is reached. Initial conditions are chosen the same for all MBRs as follows: X(biomass)=0.1 g/L, S(substrate)=40 g/L, P (product)= 0 g/L, V(volume)=1.0 L.

The *a priori* parameters in Table 1 are used to define the initial design of the nine parallel experiments. Firstly, the estimated optimal values of the feeding parameters are obtained: t_F=8 min and μ_{Set}= 0.2 min^{-1} . To start with, for all MBRs the same value for t_F is used. Secondly, the feed rate parameter for each MBR will be defined based on the (re)estimated optimal μ_{Set} (in percentages of it) as detailed in Table 2. The chosen parallel experiment design aims to introduce arbitrarily a certain degree of exploration by biasing data gathering towards a region of operating conditions that are most informative for maximizing the final concentration of the product P.

No samples will be taken from any of the bioreactors during the batch phase. During the fed-batch phase, sampling is organized by columns of bioreactors. Simultaneous sampling is only feasible for MBRs in the same column. Accordingly, after four minutes in fed-batch operation, the three bioreactors in column 1 are sampled. After another minute, the MBRs in column 2 and, one minute later the MBRs in column 3 are sampled.

For the sake of simplicity, let´s assume that the time for analytical processing of the samples taken is negligible. Thus, 7 min after t_F, the two uncertain model parameters μ_{max} and $Y_{P/X}$ are re-estimated using new sampled data. On this basis, online experimental redesign is carried out and an updated value for μ_{Set} is obtained, Then, the substrate feeding profile in each MBR is changed accordingly based on the percentages of the re-estimated optimal feeding rates as detailed in Table 2. This procedure is repeated until volumetric capacities of all MBRs are full. Please refer to the repository at:

https://git.tu-berlin.de/bvt-htbd/kiwi/tf2/experimental-computational-workflows

for details about all methods used and their hyper-parameters. Note that the *a posteriori* model parameters over-estimate the product formation rate from biomass (see values for $Y_{P/X}$ in Table 1). As new data are sampled, μ_{max} and $Y_{P/X}$ have much closer values to those of the *in-silico* model. Results obtained for the optimal feeding profiles are summarized in Table 3 and Fig. 3. As shown in Fig. 3, the predicted dynamics of product formation resembles the evolution of the *in-silico* model reasonably well.

Table 1. Model parameters

Parameter	*In-silico*	*A priori*	*A posteriori*
K_S	0.02	0.02	0.02
μ_{max}	**0.4**	**0.3**	**0.398**
$Y_{X/S}$	0.5	0.5	0.5
$Y_{P/X}$	**0.2**	**0.35**	**0.203**
$C_{S,F}$ [g/L]	500.0	500.0	500.0
V_0 [L]	1.0	1.0	1.0
V_{max} [L]	2.5	2.5	2.5

Table 2. Updated feeding rates

MBR	Col 1	Col 2	Col 3
Top	102 %	101 %	103 %
Centre	96 %	100 %	104 %
Bottom	97 %	99 %	98 %

4. Airflow-based implementation environment

4.1. Architecture of the workflow manager

The overall architecture for the workflow manager in the Apache Airflow environment is shown in Fig. 4. In the schema, the interaction between the experiment execution manager (administration console) and the Airflow environment for DAGs management using projects (experiments in the robotic platform) is depicted. From this console, the execution manager completely specifies the structure and parameters of an experiment and triggers its execution. To oversee a project execution, the project manager also has at its disposal the Airflow ecosystem made up of three main components. Firstly, a PostgreSQL database to archive execution details of DAGs scheduled in the project alongside with results obtained, and the defined user preferences. Secondly, the "scheduler" which is the core of the Airflow environment that manages the resources available for parallel processing of tasks and enforces precedence and synchronization constraints. Finally, the web services which provide a user-friendly interface to monitor the execution of DAGs and micro-services to archive results obtained in a local memory.

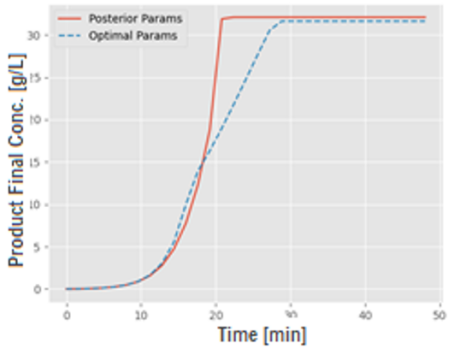


Fig. 3. Final product concentrations compared

Table 3. Optimized policies compared

	In Silico	*A Posteriori*
μ_{Set}	**0.150**	**0.230**
t_F [min]	**8.0**	**8.0**
P_{final} [g/L]	**31.59**	**32.07**

4.2. Project workflow manager

Each project in the Airflow DAGs execution environment is organized using nodes as shown in Fig. 5. The structure of files (left) is composed using Docker (centre) which are eventually managed as specific nodes in the DAGs used to execute an experiment. Each node is individually executed using a Docker container. At the end of each node script execution, its corresponding container is automatically destroyed.

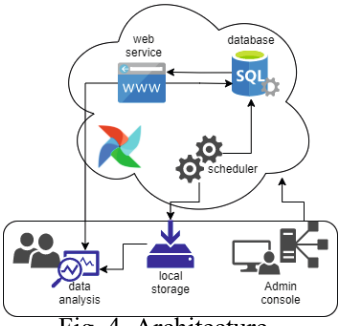


Fig. 4. Architecture

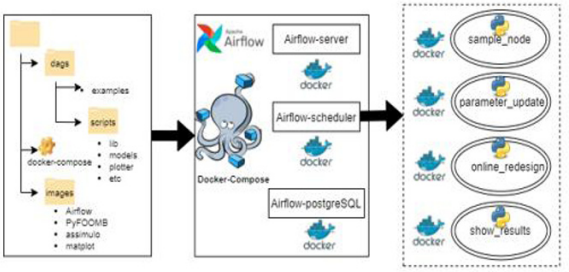


Fig. 5. Project workflow manager

4.3. Airflow ecosystem

Each project in the Airflow DAGs execution environment is organized using nodes as shown in Fig. 5. The structure of files (left) is composed using Docker containers (centre) which are eventually managed as specific nodes in the pipelines used to execute an experiment and build models from data (see Fig. 6). Each node is individually executed using a Docker container. This is necessary to transform a detailed experimental protocol into a workflow which can be understood by other modelers and experimenters. At the end of each node script execution, its corresponding container is automatically destroyed.

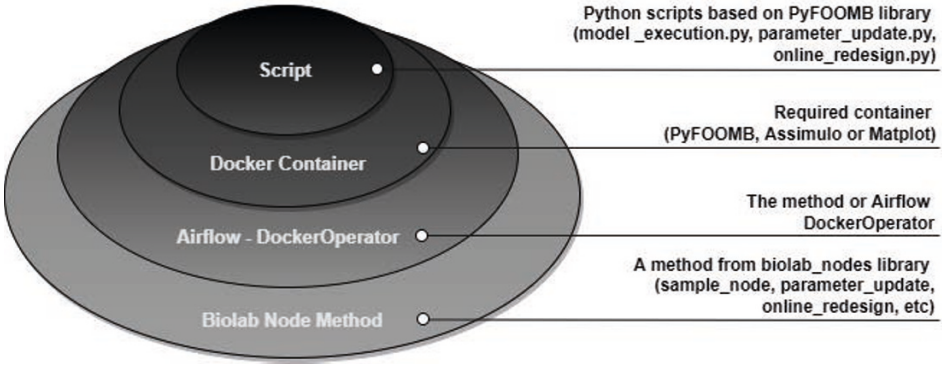


Fig. 6. Layered structure of the Airflow ecosystem

5. Concluding remarks

Simulation results demonstrate that DAGs and the Apache Airflow ecosystem are key enablers for enforcing FAIR principles in bioprocess development. The open-source nature of the software used for modeling and optimization makes experimental-computational workflows reproducible, methods transparent, and data accessible.

Acknowledgments−This work was supported by the German Federal Ministry of Education and Research through the Program “International Future Labs for Artificial Intelligence” (Grant number 01DD20002A).

References

M. Baker, 2016, 1,500 scientists lift the lid on reproducibility, Nature, 533, 7604, 452–454.

R. Celebi, J. R. Moreira, A. Hassan, S. Ayyar, L. Ridder, T. Kuhn, M. Dumontier, 2020, Towards FAIR protocols and workflows: the OpenPREDICT use case. Peer J Comput. Sci. 6:e281.

B. Haby, S. Hans, E. Anane, A. Sawatzki, N. Krausch, P. Neubauer, M. N. Cruz B., 2019, Integrated Robotic Mini Bioreactor Platform for Automated, Parallel Microbial Cultivation With Online Data Handling and Process Control, SLAS Technology, 24, 6, 569–582.

B. Harenslak, J. de Ruiter, 2021, Data Pipelines with Apache Airflow. Manning Publications.

J. Hemmerich, N. Tenhaef, W. Wiechert, S. Noack, 2021, pyFOOMB: Python framework for object oriented modeling of bioprocesses, Engineering in the Life Sciences, Eng Life Sci., 1–16. DOI: 10.1002/elsc.202000088.

M. D. Wilkinson, et al., 2016, The FAIR guiding principles for scientific data management and stewardship. Nature 3:160018 DOI 10.1038/sdata.2016.18.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50250-5

Development of cyber incident exercise to be widely adopted in supply chains

Yukiya Saito[a*], Erika Mizuno[a], Tetsushi Miwa[a], Koki Watarai[a], Yukino Suzuki[a], Midori Sumi[a], Takashi Hamaguchi[a], Yoshihiro Hashimoto[a]

[a]*Nagoya Insititute of Technology, Nagoya 466-8555 Japan*
clf19018@nitech.jp

Abstract

Industrial control systems (ICSs) are used for various kinds of social infrastructure, and these ICSs play important roles in achieving their control functions and in ensuring safety. Therefore, the cyber security of the ICSs needs to be improved. To mitigate the success of cyber-attacks on ICSs, we need to take into account suitable cooperation not from just one organization but from multiple organizations, including the companies in their supply chains. Therefore, incident response exercises require cooperation among multiple organizations. However, some hurdles related to participation in exercises are a major obstacle to improving the resilience of incident response exercises. The hurdles for participating in such exercises must be lowered. This paper focuses on this point and on the online IMANE Card we developed.

Keywords: Cyber security; Exercise; Control systems; Card game; Incident response

1. Introduction

Industrial control systems (ICSs) are used for various kinds of social infrastructure, and these ICSs play important roles in achieving their control functions and in ensuring safety. Therefore, the cyber security of the ICSs needs to be improved. Ordinarily, ICS networks have been isolated in factories. However, the ICS networks tend to connect to external networks, because a connection between the information technology (IT) system and operational technology (OT) system is required for business. Therefore, the number of cyber threats to the ICSs is increasing, and successful cyber-attacks against the ICSs can directly affect physical objects. Blout (2021) testified, a top oil company in the United States was hit by a ransomware attack in May 2021. The attack caused an incident in which the pipeline was shut down for six days, and it had a great impact on various types of fuel transport and storage. Discontinuation of services by such incidents can cause a great deal of damage to citizens' lives, companies' business, government services, and so on [1]. Cyber security for supply chains is important for our lives. However, eliminating system vulnerabilities is difficult. In addition, cyber-attacks are caused by maliciousness. Therefore, the effects of cyber-attacks are difficult to prevent even when using countermeasures. When an incident occurs due to cyber-attacks, the entire organization is required to take prompt action under limited time, cost, and resources. Moreover, safety management, business continuity management, and issue solutions must be undertaken to respond with organizational cooperation, too. Cyber-attacks are cleverly hidden and quickly spread through IT-OT networks. Therefore, identifying the intruded and attacked zone is very difficult. Even if the malware can be removed once, reinfection can occur when the countermeasures are insufficient. Suitable cooperation from not only one organization but multiple organizations, including the companies in their supply chains,

is necessary to counter such cyber-attacks. Therefore, incident response exercises require cooperation among multiple organizations, and Hashimoto et al. (2019) have developed systems for incident response exercises. They are called the IMANE (short for Incident MANagement Exercise) series, and they were designed to improve resiliency against cyber-attacks. The IMANE series has been used for exercises in workshops with various people. To improve the resiliency of multiple organizations, we need to assume that various people, such as those from group companies including superiors and subordinates, will participate. As a result, another problem may arise regarding participation in the exercises. For example, people in organizations that may have their supply chain contracts terminated may be considered risky and may hesitate to participate in the exercises. Such hurdles related to participation are major obstacles to improving resilience. The hurdles for participating in such exercises must be lowered. This paper focuses on this point and improvement one of the exercises we developed, the IMANE Card.

2. IMANE CARD

2.1 Components of the IMANE Card

The IMANE Card exercise is card-based and discussion-based. We positioned the exercise as an introduction to an incident response exercise for cyber-attacks. The participants are divided into several groups with facilitators. The participants surround one worksheet and conduct an incident response to cyber-attacks while discussing the matter with all the participants in each group.

In the IMANE Card exercise, exercise planners prepare a situation such as a virtual company and business content, plants, an organization chart and contact network, an incident scenario, etc., for the incident response exercise. In this paper, we use an example situation where one electric power company uses a hot water supply system. All the participants fully grasp the situation as a constraint and work together as a member of the company. The worksheet, milestones, and action cards of the IMANE Card are as shown in the Figure 1. This worksheet is divided into swim lanes for the actors' roles in the incident scenario. The swim lanes are presented vertically in chronological order. This example includes the computer security incident response team (CSIRT) staff and managers, who are people working in the head office, and the staff and managers of the operation section and information section, who work in the factory.

In the swim lanes, two types of boxes, called milestones, are printed such as events and target situations related to the incident scenario. This example includes the following three events as incident scenario as constraints. CSIRT checks for an increase in junk email at the head office. The information section detects suspicious signal communication on the factory. The operation section notices a change in the water level of the tank on a supervisory control and data acquisition (SCADA) display in the factory. This example features the following two target situations in the incident scenario as constraints: the recovery of abnormal water level in the tank and the resumption of the hot water supply. The participants put action cards on the worksheet. As shown in Figure 2, the action cards have card a number and name on both sides. The front side of the action card describes the concreate action as to what the card is for and the card user as to which roles the card can be used for, and some action cards have requirement states listed such as a plant stoppage. The back side of the card describes the results of the card usage in the incident scenario. The participants can imagine what kinds of roles will do what in these cards.

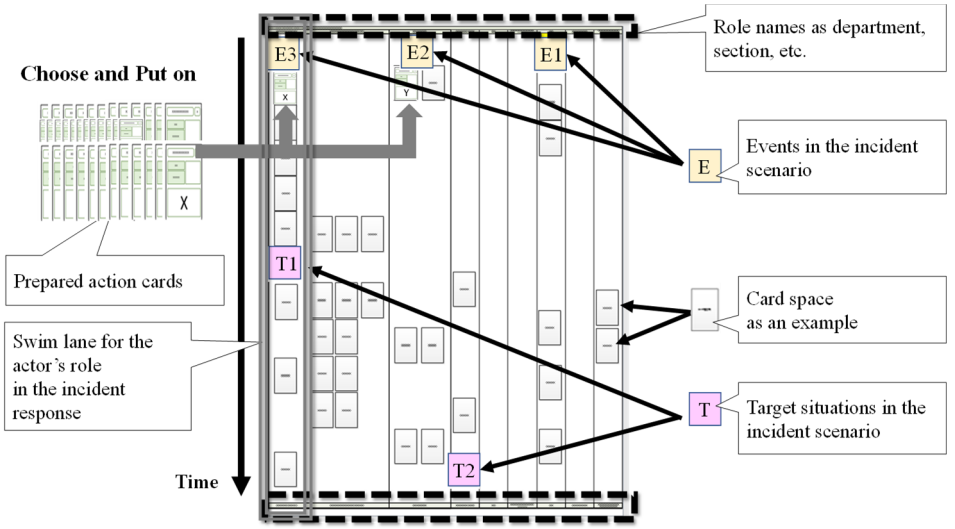


Figure 1: Worksheet, milestones, and action cards of the IMANE Card

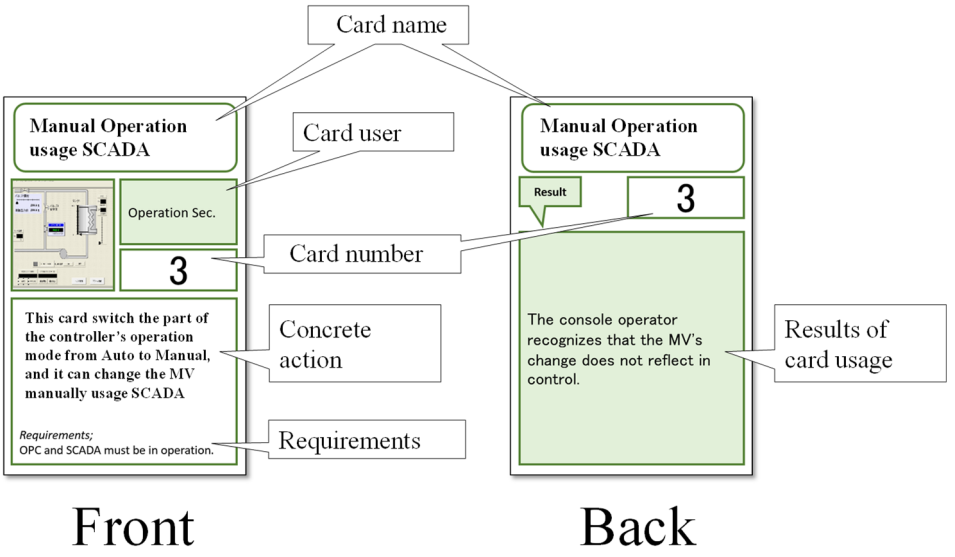


Figure 2: Action card content

For example, when the participants want to use the 17th card, the 17th card describes the situation when the 28th card is used as a prerequisite. The card arrangement order is as shown in Figure 3. As mentioned, the participants choose the suitable incident response from the prepared action cards while discussing the situation within the group. Therefore, the participants can work on exercises even if they do not have specialized knowledge about the activities of roles and cyber-attacks.

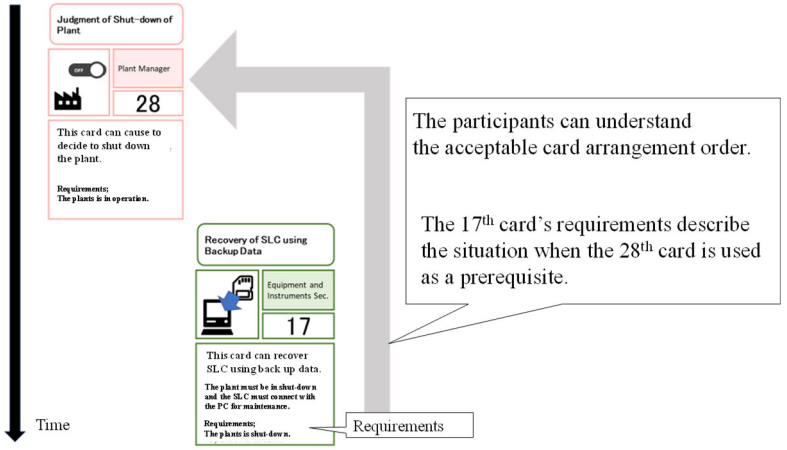


Figure 3: Card arrangement order

2.2 Procedure for the IMANE Card exercise

The IMANE Card exercise is carried out in the following phases.

Phase 1: The participants from each organization gather at the exercise place.
Phase 2: They receive some explanations about the exercise during a briefing.
Phase 3: They make a workflow and start a discussion during the exercise.
Phase 4: They review the exercise during the debriefing.
Phase 5: The exercise is declared over. The participants go back to work.

In phase 1, the participants gather at the exercise place. In phase 2, they must learn various information and knowledge about the exercise system and procedures, the virtual company, and so on prior to the exercise in phase 3. In phase 3, they make a workflow and start a discussion. The participants fully grasp the worksheet and the incident scenario as constraints. The participants browse the action cards, check the role of each card and the person in charge of execution in the virtual company, flip the cards over, read the new situation from the back side as needed, and put the suitable action cards in the correct position of the suitable swim lanes. In this way, they think about what to do next as part of the incident response. Moreover, they also consider the information linkage with other roles as suitable information for the appropriate timing as necessary, and they must show the communications with arrows and description across the swim lanes. In this way, they can discuss and consider what communications between different roles are needed to enable suitable organizational cooperation. Each facilitator supports the progress by asking the participants what they have decided and why and by answering questions from the participants. In phase 4, they review and discuss their own worksheet as a deliverable. After that, for the debriefing, the participants compare the worksheet and other groups' worksheets and discuss the incident response to deepen their awareness. The feedback is very important to improve the resiliency. In phase 5, they return to each workplace.

3. Online IMANE Card

3.1 The hurdles to improve for the incident response exercise

The following three hurdles are the focus to make it easier for various people to participate. The first hurdle is caused from a loss to the own organization due to failure in the incident

response exercise. For example, the contract may be terminated by the contractor as mentioned. Therefore, the planner must improve the exercise to enable participants to disclose as little information as necessary about themselves online in order not to have a negative impact on the working relationships when participating in the exercise. The second hurdle is the time and cost needed for the incident response exercise. For example, the physical distance to the venue can also be a major factor of this hurdle. Therefore, the planner must reduce the total time for the incident response exercise. The final hurdle is the anxiety about the lack of expertise in cyber-attacks and so on. For example, the participants may be unfamiliar with the terminology. Therefore, the planner must improve to supply the needed information to the participants on demand. In this paper, we present the online IMANE Card we developed to mitigate the impact of these three hurdles.

3.2 Components of the online IMANE Card

The online IMANE Card is based on a cloud-based online whiteboard. The online whiteboard can update the changes in real time among the participants. The necessary information required for the exercises, worksheet, card storage area, mouse pointers of other participants, overall view, and so on is in the online whiteboard of Figure 4.

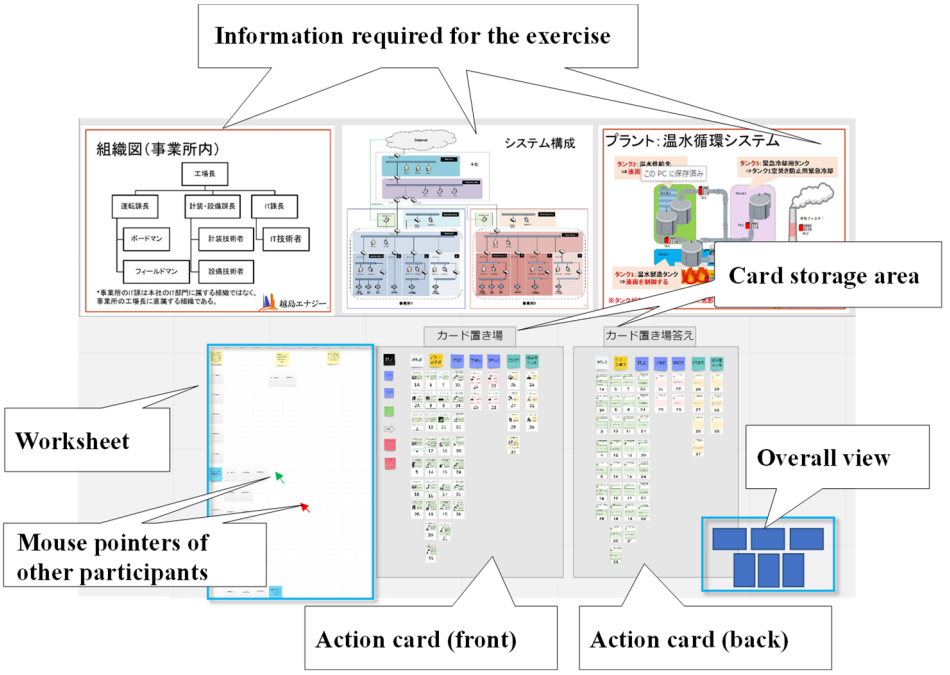


Figure 4: Online IMANE Card

The information required for the exercises such as a virtual company, business contents, the plant, organization chart and contact network, and so on are included. Action cards with a front and back are located separately in the card storage area. These are the master cards. The participants must copy the front action card and put it in the swim lane, respectively. The icons of the mouse pointers of other participants show the movement of the other participants. Each participant and facilitator can see what the other

participants are focusing on, which cards they are controlling, etc. The overall view shows the participants screen's area in the overall whiteboard. This view should make it easy to move and zoom in and out for the participants and facilitator. In addition, an online meeting tool is used for discussions that are necessary during the exercises.

3.3 Procedure for the online IMANE Card exercise

In phase 1 and 5, the participants do not have to gather in one place. The participants need to prepare a PC and have an environment where they can connect to the online whiteboard. If the participants prepare such an environment, they are able to participate in the exercise from anywhere. In phase 2, the exercise planner can use a variety of information to explain the exercises as a feature of the online whiteboard, e.g., URLs, images, and videos. The planner can attach to the online whiteboard not only the actual content needed to perform the exercises, such as the swim lane and action cards, but also the additional information that the participants feel they need such as the imagination of the equipment that appears in the exercise. Also, how to use the online tools is explained in this phase, and icebreaking among the participants is done after that. In phase 3, the participants basically only need to copy and paste and write additional characters when the participants make a worksheet. Each participant can request information that they feel necessary at any time without being aware of the other participants. After putting the card there, the facilitator and the participants read the results of the back card from the card storage area. Exercise planners may have to add new rules when conducting exercises online. The same action card may be put on more than once because the participants place cards using copy and paste. If multiple same action cards are in the swim lanes, the participants have room for further discussion, and only one of them is used at the end of the exercise. In phase 4, not only has the hurdle of participating in the exercise been lowered, but effective aspects to the output have also been revealed. The output is stored as data, making it easier to share, so the participants can expect to get more notices from some other groups and past outputs.

4. Conclusions

In this paper, we focused on the hurdles and presented the online IMANE Card we developed. It can enable more efficient conduct of discussion-based exercises where distance, cost, time, and knowledge associated with exercises for multi-organization participation may be a barrier. To improve resiliency, we need to make a system that enables more people to participate in incident response exercises as many times as possible. As in the proposed method, introducing an exercise that lowers the hurdles to participants also helps to increase the frequency of exercises.

References

J. Blout, 2021, Hearing before the United States Senate Committee on Homeland Security & Governmental Affairs, https://www.hsgac.senate.gov/imo/media/doc/Testimony-Blount-2021-06-08.pdf

Y. Hashimoto, et al., 2019, Proposal for an exercise system for achieving highely resilient organization collaboration for response to unexpected and unavoidable cyber incidents, https://www.nedo.go.jp/content/100907027.pdf

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50251-7

Requirements to a digital twin-centered concept for smart manufacturing in modular plants considering distributed knowledge

Jonathan Mädler[a*], Isabell Viedt[a], Julius Lorenz[a], Leon Urbas[a]

[a]*Chair of Process Control Systems & Process Systems Engineering Group, TU Dresden, 01062 Dresden, Germany*
jonathan.maedler@tu-dresden.de

Abstract

Modular plants are an essential step towards agile production via smart manufacturing in the process industry. An important challenge, however, is that they represent an ecosystem with distributed knowledge between process equipment assembly (PEA) manufacturers and owner/operators (O/Os). This complicates the application of familiar simulation methods. To address this issue, this paper reviews typical simulation tasks in the plant life cycle and investigates the distribution of knowledge in modular plants. While the O/O typically has extensive knowledge about the physical property system of a process, a PEA manufacturer can be expected to develop high equipment related expertise. It is suggested to use a digital twin (DT) to integrate these knowledge parts considering intellectual property concerns. The DT will be provided by the PEA-manufacturer, must be customizable through standardized interfaces, and should be applicable to different simulation tasks in the plant life cycle. To realize this concept, quality requirements specifications for DTs related to these simulation tasks must be developed. In addition, standardized interface specifications for the integration of physical property packages, the communication between different DTs and the simulation of DTs must be matured. Furthermore, hierarchical and horizontal DT structures are required to foster re-use of sub-models for different simulation tasks.

Keywords: digital twin, smart manufacturing, cyber-physical systems, modularization

1. Introduction

The process industry is facing increasing requirements to product individualization and time-to-market. Smart manufacturing is recognized as a concept to promote agile production across manufacturing and process industries and can therefore provide a solution to these challenges. Lu et al. (2020) distinguish four key features of smart manufacturing systems: (1) context- and self-awareness, (2) modularity, (3) self-organization, and (4) data-driven decision-making. Modular plants (MPs) built from process equipment assemblies (PEAs) (VDI, 2020) provide a promising solution for the feature (2) 'modularity' of smart manufacturing systems for the process industry. However, the features (1), (3) and (4) have virtually not yet been addressed for the MP concept.

The process systems engineering (PSE) community developed powerful methods and tools for modelling, simulation and control which are suitable to promote these features but due to the expectable knowledge distribution between PEA manufacturers and owner/operators (O/O) in smart MPs, adaptations to these methods and tools will be

required. Current publications of the process engineering community like Radatz et al. (2019) and Schindel et al. (2021) did not acknowledge this challenge.
Thus, the remainder of this paper is structured as follows: firstly, simulation tasks in the life cycle of conventional and MPs are reviewed; afterwards the digital twin (DT) and the distribution of knowledge in MPs are investigated. Based on the gained insights, requirements to a DT-based concept for smart manufacturing in MPs are derived, and finally a conclusion is provided discussing further research needs.

2. Simulation in the life cycle of conventional and modular plants

Modelling, simulation, and control are key tools to provide the features (3) 'self-organization' and (4) 'data-driven decision-making' for smart manufacturing systems. Oppelt et al. (2015) define four typical simulation tasks: 'design simulation and optimization', 'virtual commissioning', 'operator training' and 'plant and online optimization', which can be assigned to different phases in the life cycle of conventional plants. The goal and utilized methods differ between these tasks. The goal of (1) **design simulation and offline optimization** is to design and layout a process plant. In conceptual design, abstract unit operations are used to derive requirements to the equipment. In basic engineering, concrete equipment is designed based on these requirements. Finally, their models are combined for process simulation and optimization to find the best operating point. Applying (2) **virtual commissioning**, the automation system including the control code is tested. For this purpose, the automation system is coupled to a dynamic simulation model that mimics the behavior of the real plant. (3) **Operator training** is used to train the plant's operating staff. Simulation models are applied to mimic the real plants behavior in regular operation and failure scenarios. Finally, (4) **plant and online optimization** summarizes a broad variety of model-based support methods and tools applied during the operation phase. E.g., model predictive controllers (MPCs) are used to calculate the best operation trajectory.
In MPs, the life cycle is divided into a PEA and a MP lifecycle (see figure 1). PEA design and engineering should rely on a PEA-type specification predefined by one or more O/Os. In this phase, **design simulation & offline optimization** as well as **virtual commissioning** can be applied similar to conventional plants. Afterwards, the PEA will be fed into the plant life cycle. On the plant level, the basic engineering and detailed engineering are combined under the term orchestration by the modularization community. **Design simulation and offline optimization** change due to the transition of plant design from a sequence of equipment design tasks to a sequence of PEA selection tasks.

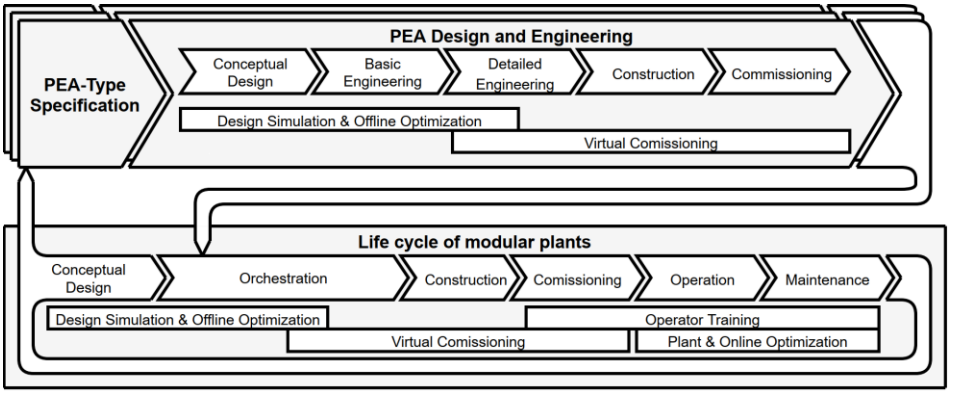


Figure 1: Simulation tasks in the life cycle of modular plants

Schindel et al. (2021) describe a general approach for technology and PEA matching in MPs. In their workflow, simulation is suggested to be used to support technology and PEA match as well as more detailed process design after choosing relevant PEAs. Radatz et al. (2019) formulate the design problem as a multi-objective superstructure optimization problem. Both papers concur in the assumption that a PEA database with simulation models of sufficient fidelity is available to the O/O. The **virtual commissioning** of MPs has been discussed e.g. by Schenk et al. (2019). The authors suggest the virtual PEA as a container packed with all virtual commissioning relevant artifacts which are needed to support the plant engineering of MPs. These models shall be used to test recipes and interlocks on MP layer. **Operator training** and **plant & online optimization** have not been addressed in the literature before.

3. The digital twin and the distribution of knowledge in modular plants

Boschert et al. (2018) define the DT as a 'semantically linked collection of the relevant digital artefacts including design and engineering data, operational data and behavioral descriptions'. Such a collection is mandatory to provide feature (1) 'context- and self-awareness' of smart manufacturing systems. In MPs, a distribution of knowledge between PEA manufacturers of high expertise and O/Os must be expected. This distribution can be illustrated applying the **product-process-resource-model** of the digital twin by Bamberg et al. (2021) (see figure 2). In this model, the **product model** represents process engineering knowledge about physical properties of pure components and mixtures, product specification, etc. (Bamberg et al., 2021). From a process simulation point of view, the product model can be related to the physical property data. In MPs, O/Os will own this knowledge and will strive to protect it from disclosure due to intellectual property (IP) concerns. The **process model** does represent physical laws, thermodynamic models, etc. and therefore equations used for the modeling (Bamberg et al., 2021). Major parts of the process model base on publicly available knowledge and are used by O/Os and PEA manufacturers alike. The **resource model** contains the 'mechanical, functional, and operational attributes of […] all equipment elements' (Bamberg et al., 2021). Hence, it characterizes the equipment with parameters like the volume of a vessel diameter. In MPs, this knowledge will be held by PEA manufacturers. Product and process model can be combined to form the **transformation model** which is a plant-independent description of the process and can be applied to derive theoretical limits e.g. of separation steps (Bamberg et al., 2021). Furthermore, it can be used to derive requirements to a concrete plant and PEAs like residence times, through puts etc. (Bamberg et al., 2021). This type of model is strongly related to the conceptual design phase of the plant life cycle of conventional and modular plants. The **capability model** is created combining the process and the resource model. It forms a product independent description of the dynamic

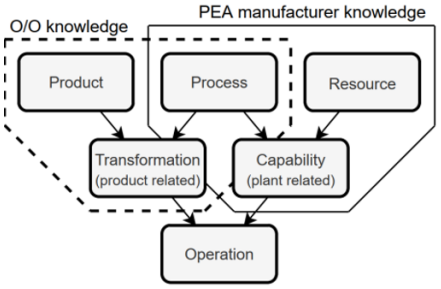


Figure 2: Knowledge distribution in MPs based on the DT structure by Bamberg et al. (2021)

behavior of a plant, equipment (Bamberg et al., 2021) or hence also a PEA. Capabilities can be estimated for ranges of material property parameters or default material system. One example are water run models, which can be applied for the purpose of virtual commissioning. Finally, the combination of transformation and capability model forms the **operation model** (Bamberg et al., 2021). This type of model is required for the optimization of process parameters and recipes for a specific product and MP. In general, this combination of information is required for the more sophisticated sub-steps in every simulation task during the plant life cycle. Therefore, a concept is required which permits to utilize the prior knowledge of both stakeholders without necessarily disclosing it.

4. A digital twin-centered concept for smart manufacturing in MPs

4.1. The smart PEA as cyber-physical system for smart manufacturing in modular plants

To address features (1), (3) and (4) of smart manufacturing systems, the modular plant concept must be extended to permit knowledge exchange and support the standard simulation tasks. At the same time, IP concerns must be considered. Thus, we suggest combining a real PEA, a digital twin of the PEA (PEA-DT), and related algorithms into a so-called smart PEA (sPEA). PEA-DTs will be composed of several information and simulation models and integrated with use case specific algorithms. In the following, requirements to PEA-DTs are discussed und R&D challenges are derived.

4.2. Simulation in the life cycle of smart modular plants

4.2.1. Design simulation and offline optimization in smart modular plants

Considering the knowledge distribution between sPEA-manufacturer and O/O, design simulation and offline optimization in smart MPs can be envisioned to split into four steps. Firstly, in a conceptual design phase the O/O will apply **transformation models** independently of specific equipment to derive requirements to the former. Based on these requirements, the O/O can find different sPEA candidates from several manufacturers or define a new PEA-type specification. The sPEA manufacturers then provide customizable **capability models** in form of downloadable or remotely accessible PEA-DTs. Hence, the PEA-DT will be used as specification of the real PEA. The O/O than must be able to customize the PEA-DTs with physical property packages to derive **operation models** from the capability model. Carrying out further process simulations using the PEA-DTs the O/O will be able to derive informed decisions which sPEAs to purchase or rent. After the physical PEA arrived, an experimentation, identification, and validation step will be necessary to reduce uncertainty of the models and validate the choice of sPEA. Finally, the identified PEA-DTs might be used to derive optimal operating points of the system. To realize this vision, PEA-DTs need to provide a sub-model describing process behavior of the PEA based on MESH equations. The external control inputs to the model should be chosen in line with the adjustable service parameters of the real PEA. The automation and control system of the sPEA itself can usually be assumed to behave ideally. The PEA-DT must provide standardized interfaces to integrate custom physical property packages, exchange mass, energy and information flows, and control the solution process. Considering model quality, the PEA-DT should provide a good accuracy, good validity around typical operation points and a high robustness. As described above, internals of simulation models might be confidential. Therefore, simulation and optimization algorithms which do not require the disclosure of the model are necessary.

4.2.2. Virtual commissioning in smart modular plants

The goal of virtual commissioning is to test the recipe with service timings and parameters, and interlocks on MP level. For this purpose, a system of PEA-DTs must be

combined in a simulation environment and coupled to the real process orchestration layer (POL), which operates the real MP afterwards. Two PEA-DTs with different levels of sophistication can be envisioned to do so. Firstly, a **capability model** considering a default material system (e.g.: water+air) and the automation system with device and I/O-models, a representation of the controller in a software-in-the-loop or model-in-the-loop configuration (cf. VDI/VDE 3693) and the OPC UA server could be used to test service sequences and interlocks. To gain a more precise picture off timings and service parameters for control loops an **operation model** will be required. Therefore, the capability model must be extendable with a physical property package. Interfaces for property packages, simulation control, and mass and energy flows can be handled like the design simulation case. The interface for information flows must be reconfigured to operate with the same service commands and data assemblies as the real PEA. The PEA-DT must provide a wide validity range, medium accuracy, and a robust numerical behavior. The simulation environment must support real time and accelerated, hybrid continuous- and discrete-time simulation since the O/O will want to test recipes and interlocks as fast as possible in a virtual environment. Accelerated simulation will require the POL to operate in a synchronized, accelerated mode as well.

4.2.3. Operator training in smart modular plants

Since, a higher fluctuation of plant configuration must be expected in MPs, operator training simulation could become a valuable tool to allow the operating staff to keep up with the changing production scenarios. Operator training does require very sophisticated process system and automation system models. In extension to 4.2.1 and 4.2.2, models must consider behavior in failure cases, which can be triggered actively and passively. Such behavior must be built in by the PEA manufacturer. The models must be combined with the POL including an operator training environment (e.g.: an app), which can be used to provide process scenarios. The interfaces of the PEA-DT are the same as in the virtual commissioning case for the most part, but an interface to actively trigger failures and provide additional feedback on failure behavior will be required. The PEA-DTs must provide a great validity range, high accuracy, high robustness, and high performance for this scenario. Simulation algorithms remain like in the virtual commissioning case.

4.2.4. Online optimization in smart modular plants

Similar to conventional plants, several different online optimization methods can be applied in smart MPs. They can be implemented on the PEA or MP level. On PEA level, the PEA-DT can be used to adjust control loops, provide MPCs, etc. In the opposite direction, the PEA-DT shall be adjusted to the state of the real PEA applying identification methods. On the plant level, PEA-DTs can be combined to e.g. provide plant wide MPCs. In most cases, the interface configuration of design simulation will be suitable for online optimization on the plant level as well. The required model quality will vary between applications. Therefore, model reduction and surrogate modeling shall be used to trade-off validity range, accuracy, robustness, and performance. The required simulation and optimization algorithms are diverse as well.

4.3. R&D challenges

Four major R&D challenges can be derived based on these requirements. Firstly, the standard simulation tasks must be related to specifiable and testable quality requirements to DTs. This is a precondition to allow certification of DTs by independent instances for specific simulation task. A promising approach is the application of quality models relying on factors, criteria and metrics (FCM) adapted from software development (Mädler et al., 2021). Secondly, standardized and open interfaces for physical property packages, flows and simulation control must be matured to allow vendor independent

exchange and application of DTs. The CAPE OPEN initiative developed promising interface specifications which are a first step but currently lack support. Another promising standard is the functional mock-up interface (FMI). Thirdly, simulation and optimization algorithms which do not require disclosure of the DT internals should be further developed. Current equation-based system simulation approaches relying on equation system manipulation will sometimes be prohibited since model internals are protected due to IP concerns. Modular sequential approaches and co-simulation will be needed to simulate process systems with distributed knowledge. Surrogate modeling and simulation offer another solution approach since freely accessible surrogates could be generated of protected PEA-DTs. Finally, hierarchical and horizontal DT structures and their semantic description should be matured to foster re-use of sub-models for different simulation tasks. Currently, multiple simulation models for different purposes are developed independently during the plant life cycle. The toolchains do not support a smooth transition of valuable sub-models from phase to phase. Matured structuring and semantic enrichment could help to overcome these issues.

5. Conclusion

In this paper, we suggest a sPEA-based approach to smart manufacturing in MPs and derive requirements to this concept. Standard PSE simulation tasks can provide self-organization and data-driven decision-making capabilities to smart MPs. Context- and self-awareness can be provided by PEA-DTs. Due to the distribution of knowledge between sPEA-manufacturers and O/Os new requirements to methods and workflows arise. We identify (1) quality assurance and specification of simulation models, (2) open and standardized interfaces, (3) simulation and optimization algorithms considering IP issues and (4) model structuring and semantic description as important R&D challenges.

References

A. Bamberg, L. Urbas, S. Bröcker, M. Bortz, and N. Kockmann, 2021, The Digital Twin–Your Ingenious Companion for Process Engineering and Smart Production, Chem Eng Technol, 44, 6, 954–961

S. Boschert, C. Heinrich, and R. Rosen, 2018, Next generation digital twin, Proc. TMCE, 7–11

M. Liu, S. Fang, H. Dong, and C. Xu, 2021, Review of digital twin about concepts, technologies, and industrial applications, J Manuf Syst, 58, 346-361

J. Mädler, I. Viedt, and L. Urbas, 2021, Applying quality assurance concepts from software development to simulation model assessment in smart equipment. In Computer Aided Chemical Engineering, 50, 813-818, Elsevier

M. Oppelt, G. Wolf, and L. Urbas, 2015, Towards an integrated use of simulation within the life-cycle of a process plant, IEEE 20th Conference on Emerging Technologies & Factory Automation (ETFA), 1–8

H. Radatz, M. Schröder, C. Becker, C. Bramsiepe, and G. Schembecker, 2019, Selection of equipment modules for a flexible modular production plant by a multi-objective evolutionary algorithm. Comput Chem Eng, 123, 196–221.

T. Schenk, A. Botero Halblaub, R. Rosen, T. Heinzerling, J. Mädler, A. Klose, S. Hensel, and L. Urbas, 2019, Co-Simulation-based virtual Commissioning for modular process plants - Requirements, Framework and Support-Toolchain for a Virtual Automation Testing Environment, Automation 2019, 229–242

A.-L. Schindel, M. Polyakova, D. Harding, H. Weinhold, F. Stenger, M. Grünewald, and C. Bramsiepe, 2021, General approach for technology and Process Equipment Assembly (PEA) selection in process design, Chem Eng Process, 159, 108223

VDI 2776-1:2020-11, 2020, Modular plants – Fundamentals and planning modular plants

VDI/VDE 3693:2016-08, 2016, Virtual commissioning – Model types and glossary

14th INTERNATIONAL SYMPOSIUM ON PROCESS SYSTEMS ENGINEERING

VOLUME 3

COMPUTER-AIDED CHEMICAL ENGINEERING, 49

14th INTERNATIONAL SYMPOSIUM ON PROCESS SYSTEMS ENGINEERING

VOLUME 3

Edited by

Yoshiyuki Yamashita
Professor and Chair of Department of Chemical Engineering
Tokyo University of Agriculture and Technology (TUAT), Tokyo, Japan
yama_pse@cc.tuat.ac.jp

Manabu Kano
Professor, Department of Systems Science
Kyoto University, Kyoto, Japan
manabu@human.sys.i.kyoto-u.ac.jp

ELSEVIER

Amsterdam – Boston – Heidelberg – London – New York – Oxford
Paris – San Diego – San Francisco – Singapore – Sydney – Tokyo

Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, USA

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

ISBN (Volume 3): 978-0-443-18726-1
ISBN (Set) : 978-0-323-85159-6
ISSN: 1570-7946

Publisher: Susan Dennis
Acquisition Editor: Anita Koch
Editorial Project Manager: Lena Sparks
Production Project Manager: Paul Prasad Chandramohan
Designer: Greg Harris

Typeset by STRAIVE

Contents

Contributed Papers: Machine Learning and Big Data

Contributed Papers: Energy, Food and Environmental Systems

Contributed Papers: Pharma and Healthcare Systems

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50252-9

Equivalence Judgment of Equation Groups Representing Process Dynamics

Chunpu Zhang, Shota Kato, Manabu Kano*

Department of Systems Science, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, Japan

manabu@human.sys.i.kyoto-u.ac.jp

Abstract

Digital twins are expected to play a key role in digital transformation of the process industry. Although process informatics, i.e., process data analytics, has attracted a lot of attention, physical models are essential to realizing the digital twins. Building a physical model of a complex industrial process is toil. We aim to free the engineers from physical model building by developing an automated physical model builder, AutoPMoB. AutoPMoB performs five tasks: 1) retrieving documents regarding a target process from literature databases, 2) converting the format of each document to HTML format, 3) extracting information necessary for building a physical model from the documents, 4) judging the equivalence of the information extracted from different documents, and 5) reorganizing the information to output a desired physical model. In this study, we propose a method of judging the equivalence of two equation groups to accomplish task 4. The proposed method first converts two equation groups in mathematical markup language (MathML) format into a format that can be manipulated in a computer algebra system (CAS). Then, the variables not shared between the groups are eliminated using the CAS. The method judges whether the two equation groups are equivalent. The results of several case studies demonstrated that the proposed method accurately judged the equivalence, including physical models of a continuous stirred-tank reactor. We also developed a web application that can easily judge the equivalence of MathML-formatted equation groups. This application is expected to reduce the effort required to find out the different models contained in multiple documents, and become an important part of AutoPMoB.

Keywords: Artificial intelligence, Equation groups equivalence, First principle model, Process modeling, Natural language processing

1. Introduction

The use of digital twins has become popular during the digitization of machinery and production systems in the manufacturing industry (El Saddik, 2018). Digital twins are virtual representations of physical entities and can be used to simulate the states inside plants under various conditions. Physical models are indispensable to enable digital twins to represent real-life phenomena accurately.

To build a physical model, experts need to survey the literature and build a physical model by trial and error until the model meets all requirements. The process of building the physical model is time-consuming.

This research aims to develop an automated physical model builder (AutoPMoB) that can automatically build a physical model. AutoPMoB first extracts information of variables, formulas, and experimental data from the literature and then combines the information to build a new physical model. Different documents may use different symbols to express the same variable and different ways to express the same formula. To combine information from multiple documents, AutoPMoB needs to recognize variables and formulas and judge their equivalence accurately.

Formulas are inherently hierarchical and can be represented as symbol layout trees (SLTs) or operator trees (OPTs) (Mansouri et al., 2019). SLTs capture the placement and nesting of symbols on writing lines, while OPTs capture the mathematical semantics of the application of operators to operands. The SLT and OPT of a formula $x - y^2 = 0$ are shown in Figure 1.

There has been no work on equivalence judgment of two equation groups. It seems possible to judge the equivalence by setting a threshold on the similarity of the two equation groups. Several similarity measures are available. Zhong and Zanibbi (2019) defined similarity between formulas using paths from OPTs. Mansouri et al. (2019) built an embedding model of formulas using SLTs and OPTs and defined the similarity as their cosine similarity. These similarities are based on appearance, and similar-looking formulas do not necessarily perform the same calculations; thus, the similarities do not properly work for the equivalence judgment.

In this work, we propose a rule-based method of judging the equivalence of two equation groups. In our proposed method, equation groups in mathematical markup language (MathML) format are converted into a format that can be manipulated using a computer algebra system (CAS). Then, the variables not shared between the groups are eliminated using the CAS. The equivalence of the two equation groups is judged by checking whether any equation in one equation group appears in the other equation group. We assume that variables having the same meaning are represented by the same symbol.

2. Proposed Method

This research aims to develop a system that can judge the equivalence of two equation groups not by their appearance but by the calculations they perform. Formulas in MathML format are just strings and cannot be manipulated following calculation rules. Thus, the

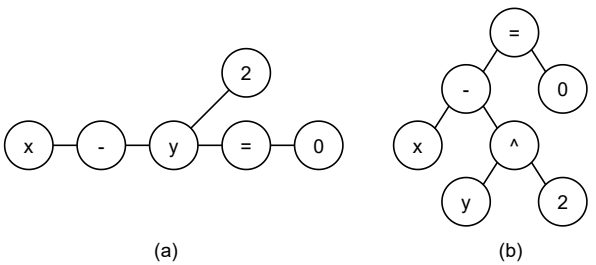


Figure 1: (a) SLT and (b) OPT of the formula $x - y^2 = 0$.

proposed method converts formulas in MathML format into another format that can be manipulated mathematically using SymPy (Meurer et al., 2017), which is a CAS written in Python. Finally, the proposed method judges whether the equation groups are equivalent using the algorithms explained in sections 2.2. and 2.3.

2.1. Conversion of Equations

There are two types of MathML: presentation markup and content markup. Presentation markup captures notation structure while content markup captures mathematical structure. This research uses presentation markup since it is mainly used on websites. We convert a formula in MathML format into tangents-SLT by using tangent-s, a formula search system developed by Davila and Zanibbi (2017). Tangents-SLT is an SLT in which each symbol is assigned a certain type. A list holding the hierarchical information of the formula is created by parsing the tangents-SLT. The list is converted into a SymPy object. For example, the tangents-SLT of a formula $\frac{a+b}{c+d}$ is first converted into [tangents-SLT 1, tangents-SLT 2, 'frac'], where tangents-SLT 1 and tangents-SLT 2 refer to the tangents-SLT of the formula $a + b$ and the formula $c + d$, respectively. Tangents-SLT 1 and tangents-SLT 2 are then converted into a list of ['a', '+', 'b'] and ['c', '+', 'd']. Finally, the list [['a', '+', 'b'], ['c', '+', 'd'], 'frac'] is converted to a SymPy object.

2.2. Equivalence Judgment of Equations

In the proposed algorithm, two equations must have the same variables to be judged equivalent. The two equations are judged equivalent if the solutions for any variable of the equations are the same; otherwise, they are judged non-equivalent. As for equations that contain a derivative term such as $\frac{\mathrm{d}V}{\mathrm{d}t}$, the whole term is regarded as one variable. For example, the solutions of the equations $\frac{\mathrm{d}V}{\mathrm{d}t} = w_i - w$ and $\frac{\mathrm{d}V}{\mathrm{d}t} + w - w_i = 0$ for w_i are both $\frac{\mathrm{d}V}{\mathrm{d}t} + w$, which means they are judged equivalent. The solution for V is not calculated since the term $\frac{\mathrm{d}V}{\mathrm{d}t}$ is regarded as one variable. We use CAS to solve equations.

2.3. Equivalence Judgment of Equation Groups

In this research, we judge not only the equivalence between equations but also the equivalence between two equation groups consisting of multiple equations.

Two equation groups have the same degree of freedom (DOF) when they are equivalent; therefore, the DOF of the two equation groups are compared at first. If two equation groups have the same equations, they are obviously equivalent, and both groups have the same set of variables. Hence, we eliminate the variables not shared between the groups. If there remain variables not shared between the groups, the two equation groups are judged non-equivalent. Finally, for each equation in one equation group, our proposed method seeks the equation performing the same calculation in another equation group using the algorithm introduced in section 2.2. If such an equation does not exist, the two equation groups are judged non-equivalent; otherwise, they are judged equivalent.

3. Case Studies

We evaluated our proposed method through several case studies and confirmed that the proposed method could accurately judge the equivalence of equation groups. In this section, we introduce one case study.

In Figure 2, three equation groups are physical models of a continuous stirred-tank reactor (CSTR), where an exothermic and irreversible reaction (A→B) takes place. The two equation groups (a) and (b), which are used by Manzi and Carrazzoni (2008) and Nekoui et al. (2010), are equivalent but written in different ways. The equation group (c) is non-equivalent to the equation groups (a) and (b). The proposed method accurately judged that the equation groups (a) and (b) are equivalent while the equation groups (a) and (c) are non-equivalent.

We developed a web application for equivalence judgment. Figure 3 shows a screenshot to judge the equivalence of the two equation groups (a) and (b). To use this web application, we first choose two files, in which equation groups are written in MathML format (Part 1). Our web application then shows the selected equation groups (Part 2) and the equivalence judgment result (Part 3). Equation groups in MathML format can be easily

$$\begin{aligned} \frac{\mathrm{d}C_\mathrm{A}}{\mathrm{d}t} &= \frac{F}{V}(C_{\mathrm{A,in}} - C_\mathrm{A}) - k_0 \exp\left(-\frac{E}{RT}\right) C_\mathrm{A} \\ \frac{\mathrm{d}T}{\mathrm{d}t} &= \frac{F}{V}(T_\mathrm{in} - T) + \frac{h_\mathrm{r}}{\rho c_\mathrm{p}} k_0 \exp\left(-\frac{E}{RT}\right) C_\mathrm{A} - \frac{UA_\mathrm{r}}{V\rho c_\mathrm{p}}(T - T_\mathrm{j}) \end{aligned}$$

(a)

$$\begin{aligned} \frac{\mathrm{d}C_\mathrm{A}}{\mathrm{d}t} &= \frac{F}{V}(C_{\mathrm{A,in}} - C_\mathrm{A}) - k_0 \exp\left(-\frac{E}{RT}\right) C_\mathrm{A} \\ \frac{\mathrm{d}T}{\mathrm{d}t} &= \frac{F}{V}(T_\mathrm{in} - T) + \frac{h_\mathrm{r}}{\rho c_\mathrm{p}} k_0 \exp\left(-\frac{E}{RT}\right) C_\mathrm{A} - \frac{Q}{V\rho c_\mathrm{p}} \\ Q &= UA_\mathrm{r}(T - T_\mathrm{j}) \end{aligned}$$

(b)

$$\begin{aligned} \frac{\mathrm{d}C_\mathrm{A}}{\mathrm{d}t} &= \frac{F}{V}(C_{\mathrm{A,in}} - C_\mathrm{A}) - k_0 \exp\left(-\frac{E}{RT}\right) \\ \frac{\mathrm{d}T}{\mathrm{d}t} &= \frac{F}{V}(T_\mathrm{in} - T) + \frac{h_\mathrm{r}}{\rho c_\mathrm{p}} k_0 \exp\left(-\frac{E}{RT}\right) - \frac{UA_\mathrm{r}}{V\rho c_\mathrm{p}}(T - T_\mathrm{j}) \end{aligned}$$

(c)

Figure 2: Physical models of a CSTR; (a) and (b) are equivalent while (c) is non-equivalent to (a) and (b).

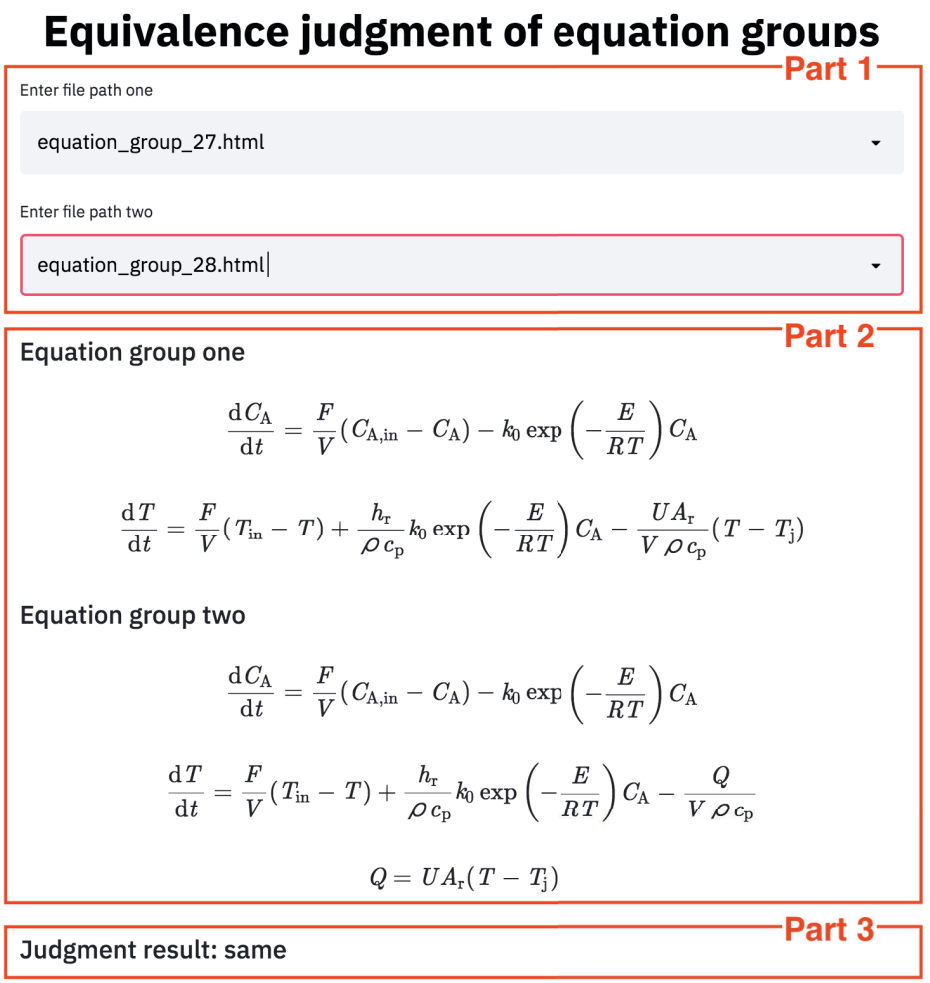


Figure 3: The screenshot of our web application; file selections for equivalence judgment (Part 1); selected equation groups (Part 2); the equivalence judgment result (Part 3).

obtained from PDF or TeX files using existing tools such as InftyReader (Suzuki et al., 2003) or LaTeXML (Miller, 2018). By using this web application, the equivalence of two equation groups can be easily judged.

4. Conclusion

In this work, we proposed a rule-based method for equation group equivalence judgment. In the proposed method, we convert formulas in MathML format into SymPy objects, which we can manipulate mathematically. We judge the equivalence of equation groups by checking whether any equation in one equation group performs the same calculation

as the equation in the other equation group. The results have shown that the proposed method accurately judges whether two equation groups are equivalent or not. Currently, the proposed method cannot be applied to equation groups, including several calculations such as summation and infinite product. For future work, we plan to make our system support more types of calculations.

Acknowledgments This work was supported by JSPS KAKENHI Grant Number JP21K18849 and JST SPRING Grant Number JPMJSP2110.

References

K. Davila, R. Zanibbi, 2017. Layout and semantics: Combining representations for mathematical formula search. In: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 1165–1168.

A. El Saddik, 2018. Digital twins: The convergence of multimedia technologies. IEEE MultiMedia 25 (2), pp. 87–92.

B. Mansouri, S. Rohatgi, D. Oard, J. Wu, C. Giles, R. Zanibbi, 2019. Tangent-cft: An embedding model for mathematical formulas. In: Proceedings of the 2019 ACM SIGIR international conference on theory of information retrieval. pp. 11–18.

J. Manzi, E. Carrazzoni, 2008. Analysis and optimization of a cstr by direct entropy minimization. JOURNAL OF CHEMICAL ENGINEERING OF JAPAN 41 (3), pp. 194–199.

A. Meurer, et al., 2017. Sympy: symbolic computing in python. PeerJ Computer Science 3, p. e103.

B. Miller, 2018. LaTeXML The Manual—A LaTeX to XML/HTML/MathML Converter, Version 0.8.3. `https://dlmf.nist.gov/LaTeXML/`, (Accessed on 11/03/2021).

M. A. Nekoui, M. A. Khameneh, M. H. Kazemi, 2010. Optimal design of pid controller for a cstr system using particle swarm optimization. In: Proceedings of 14th International Power Electronics and Motion Control Conference EPE-PEMC 2010. IEEE, pp. T7–63.

M. Suzuki, F. Tamari, R. Fukuda, S. Uchida, T. Kanahori, 2003. Infty: an integrated ocr system for mathematical documents. In: Proceedings of the 2003 ACM symposium on Document engineering. pp. 95–104.

W. Zhong, R. Zanibbi, 2019. Structural similarity search for formulas using leaf-root paths in operator subtrees. In: European Conference on Information Retrieval. Springer, pp. 116–129.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50253-0

Data-driven operation support for equipment deterioration detection in drug product manufacturing

Philipp Zürcher[a], Sara Badr[a*], Stephanie Knueppel[b], Hirokazu Sugiyama[a]

[a]*Department of Chemical System Engineering, The University of Tokyo, Tokyo 113-8656, JAPAN*
[b]*Engineering, Science & Technology, F. Hoffmann – La Roche Ltd., Wurmisweg, 4303 Kaiseraugst, SWITZERLAND*

badr@pse.t.u-tokyo.ac.jp

Abstract

This work presents a data-driven methodology for decision-support aiming at reliability experts in drug product manufacturing. The developed tool incorporates three consecutive stages. Firstly, equipment condition monitoring is performed through principal component analysis for dimensionality reduction on the process monitoring dataset. Equipment deteriorations are visualized by shifts in the monitored principal curve giving indication about deviating equipment condition. Secondly, a localization of the underlying physical source for the detected equipment deterioration is performed. Thereby, the impact from individual sensors to the observed shifts is investigated giving additional information to decision-makers on the underlying physical phenomena and location in the unit. In the last stage, prevention, the information from the two previous stages is combined in order to perform tailored maintenance actions during the production phase in order to minimize the occurrence of unplanned downtime. The developed methodology is demonstrated in the form of a case study. Industrial process data from a sterilization unit which is part of an aseptic filling line of F. Hoffmann – La Roche Ltd. located in Kaiseraugst, Switzerland is used.

Keywords: Predictive maintenance, decision-support, industrial application.

1. Introduction

Digitalization is the central pillar for the introduction of Industry 4.0. (Diez-Olivan *et al.* 2019) However, unlike other industry sectors, sophisticated data-driven applications are yet to be explored to fully take advantage of the abundance of recorded process data that is available in pharmaceutical manufacturing. The pharmaceutical manufacturing sector is highly regulated by government agencies, such as the FDA, in order to guarantee product safety. Thereby, regulators are requiring manufacturers to store process monitoring information over multiple years for backtracking purposes. (Casola *et al*., 2019) A large historical data base is thus produced, which could potentially be used in order to gain data-driven insights into the current equipment condition.

Data-driven approaches including the application of principal component analysis (PCA) are well established in the field of process monitoring. However, few approaches have been proposed that focus on the long-time equipment related trends in the process data. (Reis *et al.*, 2017) In other industry sectors, such approaches have been presented under the term of predictive maintenance. (Bousdekis *et al.* 2019) However, in the pharmaceutical industry such approaches are scarce. One obstacle there is the costly

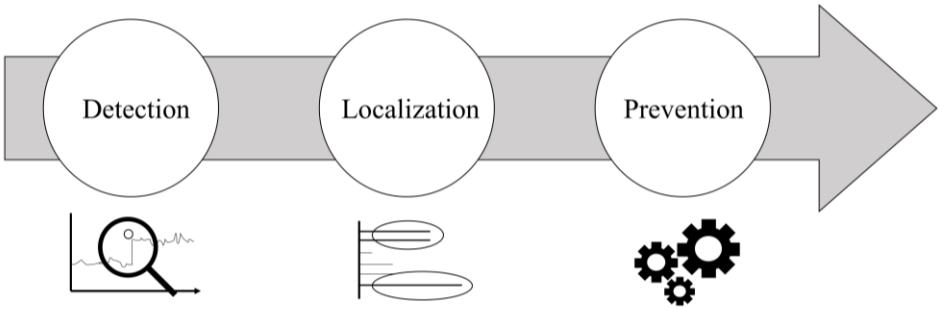


Figure 1: Overview of the individual steps of the developed methodology for equipment condition monitoring.

revalidation process associated with any changes in the production line including the installation of new sensors for predictive maintenance purposes. (Zürcher *et al.*, 2020)
Equipment reliability is of great importance in the manufacturing of highly valuable pharmaceutical products as any equipment malfunction or fault can potentially result in a loss of highly valuable drug product. Therefore, maintenance is performed on a regular basis within production facilities. It is based on a pure time-based maintenance scheme, which does not consider real-time equipment condition. Although having regular maintenance intervals, unexpected equipment faults induced by manual interventions or resulting from equipment deterioration are still frequent. Developing a method to determine the current condition of equipment is necessary, in order to minimize the occurrence of such unexpected events.
This work presents a decision-support tool for reliability experts to evaluate the current equipment condition. Thereby, aiming to provide support in early detection and localization of equipment deterioration to provide guidance for required maintenance interventions during production. Results are presented in the form of a case study for the aseptic filling process.

2. Methodology

In this work, real-time production data was used. Non-production time in the continuously collected process data was excluded. The production datasets form the basis for the developed methodology in this work. In comparison to traditional monitoring, the investigated time frame is on production campaigns and entire production phases, instead of individual batch runs.
The decision-support methodology is composed of three consecutive stages as summarized in Fig. 1. First, in the detection stage, equipment monitoring is performed, and data records are maintained. Multi-way PCA is performed for dimensionality reduction of process data. (Nomikos & MacGregor, 1994). The data is unfolded variable-wise and auto-scaled prior to the application of PCA. This is followed by the application of a noise-filter algorithm on the principal curves. (Savitzky & Golay, 1964) Long-term based deterioration becomes visible, expressed through fluctuations or shifts of the monitoring curve.
The ongoing production cycle dataset is investigated with regards to the observed shifts in any principal component which indicates a change in the equipment condition. The deterioration zone is defined based on historical analysis of the data where all shifts

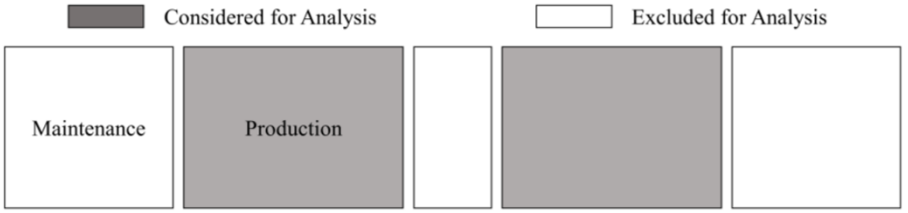


Figure 2: Overview over the production and maintenance schedule for the aseptic filling line of one year that is considered in this case study.

reaching the area ultimately led to an equipment fault within the near future (several days to weeks) after the detected shift and a restoration of the base line.
Secondly, the results from the principal component analysis are used for an in-depth loading analysis for the principal curve where a shift has been detected. Thereby, the individual contributions to the explained variance by the original variables representing measurements from installed sensors within the manufacturing unit are considered. Through the identification of the dominantly contributing sensor(s), the detected shift can be localized within the unit and the underlying physical phenomenon (e.g., leakage) is identified. Together with operator expert knowledge, this stage supports decision-makers to locate deteriorations and target maintenance actions.
Thirdly, by using the combined information from the detection and localization stage, maintenance actions can be proposed and windows where actions become necessary can be determined. Ultimately, a step towards more intelligent maintenance procedures in drug product manufacturing is achieved that supports experts in monitoring the actual equipment condition.

3. Case study

This work considers one year of industrial production data obtained from a sterilization unit that is part of the aseptic filling process. It represents a key process within drug product manufacturing.
Currently, maintenance is performed in a time-based manner in the facility twice a year. Consequently, two production phases over the course of a year are considered which are separated by maintenance actions and are combined in a data set within the scope of this work. The schematic overview is shown in Fig. 2.

4. Results & Discussion

4.1. Detection

In order to demonstrate the application of the methodology, the resulting principal curves for the sterilization unit were analysed for shifts. PC2 was selected as it showed long-term baseline shifts upwards into the deterioration zone, which persisted for several days to weeks as shown in Fig. 3.
Observed shifts having an intensity within the defined deterioration zone eventually led to an unexpected equipment fault. PC2 was identified to be suitable for long-term condition monitoring as shifts leading to an intensity within the deterioration zone always ended with an equipment fault requiring maintenance actions. The two production cycles – each representing about 3-4 months of commercial production - are combined, thereby visualizing the monitoring curve for an entire year of production. Steps in the principal curve can be observed which can be clearly separated from individual, short-

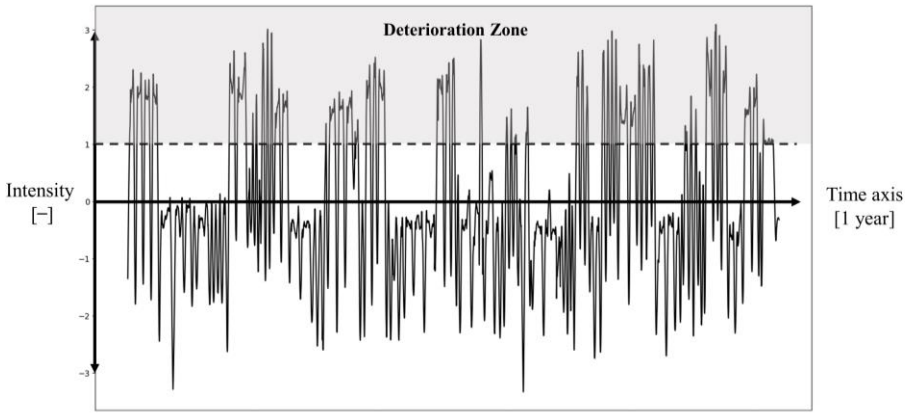


Figure 3: Representation of the principal curve (PC2) indicating a zone where increased equipment deterioration in observed which could eventually lead to an equipment fault.

term based peaks frequently occurring and resulting from intra and inter batch variance. Furthermore, observed shifts of PC2 last for up to several weeks until eventually a restoring of lower intensity is obtained. This restoration is connected to a performed maintenance action after an (unexpected) equipment fault has occurred. Therefore, an observed shift indicates that equipment deterioration is occurring which will eventually lead to a fault. Through the detection of such shifts, decision-makers can become aware of an underlying problem within the unit.

4.2. Localization

The loading plot for PC2 where shifts have been detected is shown in Fig.4. Highlighted bars represent sensors with the highest impact on the explained variance of the principal curve. Related sensors in the unit are the ventilator performance as well as the position

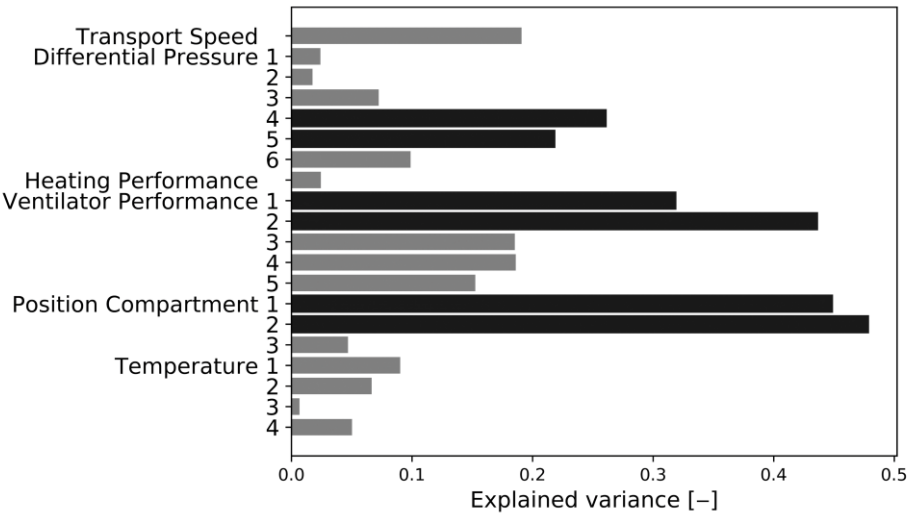


Figure 4: Overview of the loadings for principal component 2. Different numbers for a sensor type indicate different locations within the production unit. Highlighted loadings represent sensor values contributing most the monitoring curve in the detection stage.

setting of the compartments. Both are connected to the pre-heating zone (1) and cooling zone (2) respectively. Therefore, shifts observed can be located to belong to this area within the sterilization unit. Furthermore, the differential pressures between the pre-heating zone and sterilization zone (4) as well as cooling zone and sterilization zone (5), respectively, are dominantly expressed in PC2. Therefore, shifts observed for the principal curve are likely to be connected to pressure related equipment, such as valves which can suffer from deterioration eventually resulting in leakage and breakage of the sterile environment.
Results from the loading analysis show that the problem can be reduced to a small number of related sensors in specific areas of the unit. The localization stage therefore can aid decision-makers to identify the origin of detected equipment problems and increase maintenance actions through the offered possibility of better tailored actions.

4.3. Prevention

In the prevention step, the information gained from the detection and localization stage is combined. Thereby, tailored maintenance actions can be determined during phases with increased equipment deterioration represented by arrows in Fig. 6. Proposing such actions during a production cycle is an advantage for decision-makers in reliability engineering departments in comparison to the conventional practice where the condition is not monitored in real-time at all.
Phases with increased equipment deterioration eventually led to an equipment fault that is related to the identified location and underlying physical conditions that were determined in the localization stage. Therefore, the occurrence of unexpected equipment faults can be reduced by combining the information from detection and localization stages in the final prevention stage and tailored maintenance actions can be planned through the data-driven insights obtained by the proposed methodology.

5. Conclusion

In this work, a methodology for equipment condition monitoring as decision-support for reliability experts in drug product manufacturing is presented. A three-stage methodology is presented including a detection, localization and prevention stage. Thereby, equipment condition shifts due to deterioration are first identified in the detection stage

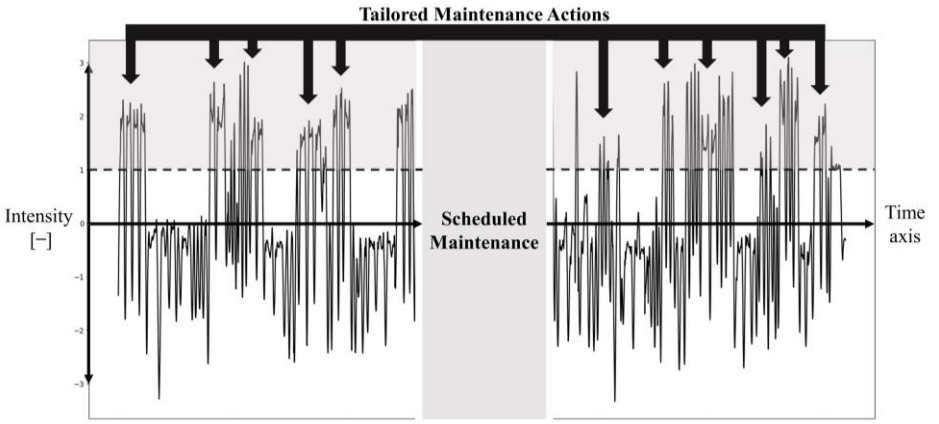


Figure 5: Schematic representation of the proposed type of maintenance after the implementation of the developed methodology, moving from scheduled to (partly) tailored maintenance.

by monitoring principal curves. Then, the deterioration is localized by loading analysis enabling both the identification of the underlying physical condition problem as well as the location within the unit.
Finally, in the prevention stage, through tailored maintenance actions, the occurrence of unexpected equipment faults from deterioration should be minimized. In phases where increased deterioration of the manufacturing equipment is detected, adequate maintenance actions have to be planned and production schedules altered accordingly. By using the information from the localization stage, tailored actions can be planned and executed thus reducing maintenance efforts.
The effective application of the methodology for monitoring the equipment condition of the sterilization unit as part of the aseptic filling process has been demonstrated. Its usefulness has been shown in both, the identification of phases with increased deterioration characterized by shifts in the monitoring curve, as well as the localization of the physical origin of the deterioration by a loading analysis. Consequently, decision-support can be provided for reliability experts on needed maintenance for special equipment. Thus, tailored maintenance actions can be performed leading to a decrease in maintenance effort, time, and resulting in a higher cost effectiveness. In order to further improve the applicability of data-driven monitoring, it is not enough to only predict changes in the underlying equipment condition. The timeframe until a failure would likely occur also needs to be addressed. Therefore, further research on how to predict the exact occurrence of such failures has to be conducted in order to further enhance more effective production and maintenance processes in drug product manufacturing.

References

Diez-Olivan, Alberto; Del Ser, Javier; Galar, Diego; Sierra, Basilio; 2019; *Data fusion and machine learning for industrial prognosis: Trends and perspectives towards Industry 4.0*; Information Fusion, 50, 92-111.

Casola, Gioele; Siegmund, Christian; Mattern., Markus; Sugiyama, Hirokazu; 2019; *Data mining algorithm for pre-processing of biopharmaceutical drug product manufacturing records*; Computers and Chemical Engineeering, 124, 253-269.

Reis, Marco S.; Gins, Geert; 2017; *Industrial Process Monitoring in the Big Data/Industry 4.0 Era: From Detection, to Diagnosis, to Prognosis;* Processes, 5, 35.

Bousdekis, Alexandros; Lepenioti, Katerina; Apostolou, Dimitris ; Mentzas, Gregoris; 2019; *Decision Making in Predictive Maintenance: Literature Review and Research Agenda for Industry 4.0*; IFAC-PapersOnLine; 52-13, 607-612

Zürcher, Philipp; Shirahata, Haruku; Badr, Sara; Sugiyama, Hirokazu; 2020; *Multi-stage and multi-objective decision-support tool for biopharmaceutical drug product manufacturing: Equipment technology evaluation*; Chemical Engineering Research and Design, 161, 240-252.

Nomikos, Paul; MacGregor, John F.; 1994; *Monitoring Batch Processes Using Multiway Principal Component Analysis;* AIChE Journal; 40; 8; 1361-1375.

Savitzky, Abraham; Golay, Marcel J.E.; 1964; *Smoothing and Differentiation of Data by Simplified Least Squares Procedures;* Analytical Chemistry; 36, 8, 1627-1639

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50254-2

Pilot Plant 4.0: A Review of Digitalization Efforts of the Chemical and Biochemical Engineering Department at the Technical University of Denmark (DTU)

Mark Nicholas Jones, Mads Stevnsborg, Rasmus Fjordbak Nielsen, Deborah Carberry, Khosrow Bagherpour, Seyed Soheil Mansouri, Steen Larsen, Krist V. Gernaey, Jochen Dreyer, John Woodley, Jakob Kjøbsted Huusom, Kim Dam-Johansen*

Department of Chemical and Biochemical Engineering, Technical University of Denmark, Building 228A, 2800 Kgs. Lyngby, Denmark
kdj@kt.dtu.dk

Abstract

The pilot plant at the Chemical and Biochemical Engineering Department at DTU (DTU Kemiteknik, abbrv. DTU KT) serves as a facility for research & education with access to various process equipment, commonly employed in up-and down-stream processes. Among the available equipment are fermenters, membranes, distillation columns, absorbers, desorbers, extractors, crystallizers, chromatography columns and all kind of high-temperature reactors and process equipment for particulates. The equipment is supplemented by mobile demonstration units for use at industrial sites and a large-scale maritime test station. These units are perfectly suited in combination with laboratory facilities to perform scale up studies together with the capabilities of a modern digital infrastructure. Some of the units are only operated manually while other units can be operated through human machine interfaces (HMI). In line with DTU's strategic objectives, DTU KT focuses on the development and application of an Industry 4.0 framework for its research and educational activities. Therefore, the pilot plant and laboratory facilities are going through a digital transformation, creating a suitable infrastructure that provides remote accessibility to all research and operational data. These efforts are presented in this work.

Keywords: Digitalization, Database, Process Control, Digital Twin, Machine Learning

1. Introduction

The pilot plant facilities of DTU KT were established more than 50 years ago, and have been under constant development since then. A first big step after its establishment dates back to the 1990's with the construction of a large high-temperature facility followed by a general continuing update from 2005 to 2020, adding all kind of up- and down-stream equipment for the chemical, biochemical and energy related industries. Further, large pyrolysers and gazifiers for biomass and waste are located at the DTU Risø Campus and have recently been supplemented by advanced equipment for fermentations based biomanufacturing along with mobile units for industrial

demonstrations of CO2 capture processes and a large-scale maritime test facility located at Hundested harbour.

The selection of units available in the pilot plant serves to recreate complete production lines in e.g. extensive scale-up studies. The available unit operations are however equipped with different degrees of automation and localized data storage. Some unit operations are only operated manually while other units are operated through a LabView interface and/or simple PID control loops. Thus, the current digital framework for the department's interaction with the raw process data generated in the pilot plant can be described as fragmented without connection to the internal network. This limits the ability of the department to test developed high-level control and optimization schemes such as digital twin implementations that embed various modelling and simulation environments with real-time data. A second limitation is the optimal use of the collected data. Until now, a minor subset of valuable simulation models, also referred to as digital twins, have been connected with the physical units available in the pilot plant. The framework presented in this contribution has the aim to develop a consistent guidance for researchers to easily integrate their simulation models at the department.

2. Overall Framework and Infrastructure

In the present strategy period of DTU (2020-2025), the department will extend the capabilities of the pilot plant. The current efforts involve equipping the pilot plant with Industry 4.0 capabilities. In-house and industrial surveys were conducted and identified improved data handling as key needs for academia and industrial collaborators. To assurethe possibility of coupling developments with external partnership from industry, DTU KT has selected to modernize the data infrastructure (Udugama et al., 2021; Bähner et al., 2021; Lopez et al., 2021).

Currently, DTU KT is implementing the modernization efforts to overcome the previously described boundaries (Gargalo et al., 2021). The integration of the unit operations into automation platforms via SCADA, OPC-UA and API communication to digital twins are shown in Figure 1. The automation software is deployed on dedicated servers (Windows Server and Kubernetes cluster). Jupyter Hub, Kubeflow and git repositories will allow researchers to easily implement real-time optimization methods such as advanced process control and scheduling of operations (Ziaei-Halimejani et al., 2021). The developed framework establishes a digitalized research and education environment. By describing the framework from the bottom up, starting with the database and the connection to digital twin models, sensors and actuators, a prototype implementation has been established to evaluate and further improve the framework and will described in the fourth section.

Process scale-up and optimizationare complex, multi-faceted disciplines that engage several domains within chemical engineering. Knowledge and research within all of these areas must be combined to achieve a feasible large-scale design of an industrial process. Disciplines or aspects which are covered in scaling-up a process include: (I) design of experiments (DoE), (II) computational fluid dynamics (CFD), (III) kinetic studies, (IV) steady-state and dynamic simulations for mass and energy balances (V) quantification of disturbances (VI) thermodynamic modelling of multi-phase and multi-component systems (VII) control system design and implementation.

Process systems engineering (PSE) has advanced from previously solving scale-up problems by addressing each of the above-mentioned disciplines individually. Now, focus has shifted towardsadopting a hybrid multi-layer decomposition approach that makes extensive use of the operational data, to gain more insight and leverage the modelling capabilities with e.g. machine learning and data analysis during process characterization in laboratory and pilot scale experimentation.

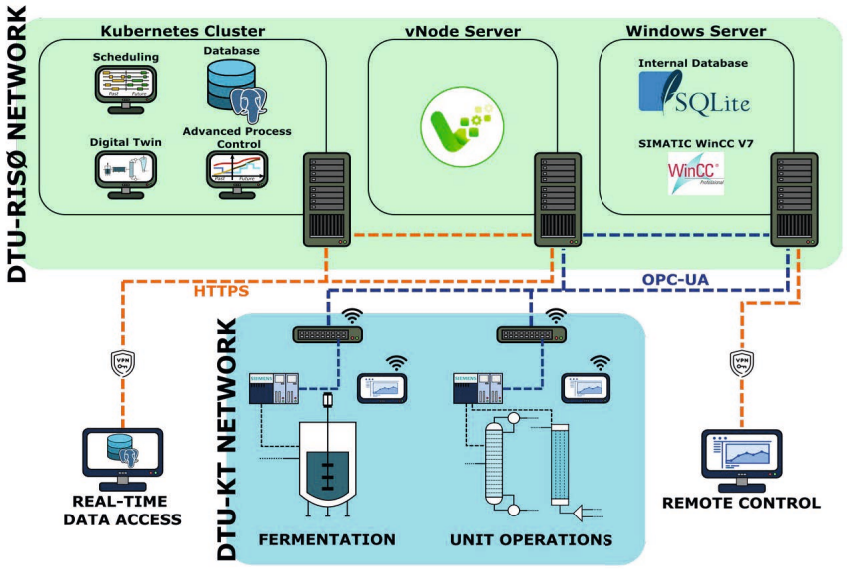


Figure 1: Overall framework for digitalizing the DTU KT pilot plant infrastructure

With the newly implemented framework, these information streams are being sent through various platforms and environments. Consolidating the whole infrastructure of such a multi-platform system has become a new task for engineers who must now spend considerable time to pipeline the data streams between process simulators, experimental equipment and various other computational tools.

3. Database Design

This project is fundamentally about integrating data and data structures from multiple different sources. Accessible via well-defined interfaces, these new features and functions will provide for a diverse range of use cases and applications. As such, the digitalization project at DTU KT is heavily focusing on designing the database schema correctly to meet the needs of all stakeholders at the department. The database schema presented in this article will continue to be modified until a schema has been developed where all the research work of all research groups can be easily mapped to a relational database and updated accordingly.

While undertaking this bottomup approach of developing the infrastructure, the database is also being harmonized with the data delivered through the Siemens SCADA system and the vNode interface which stores the history data in a non-relational database (MongoDB).

The next step in the data pipeline is the treatment and cleaning step before the data is stored in a relational database (Postgres). As seen in Figure 1, a data processing step is performed when data is being transferred from the vNode system to the database. This allows to apply data clean up and data annotation methods on the raw data stored in the non-relational database (Perazzoli et al., 2018; Russo et al., 2021). Further, outlier treatment can be performed and other statistical metrics (e.g. means, frequencies, weights) generated to be stored in the relational database. Additionally, outlier treatment, clustering, annotation and statistical analysis can be performed to enhance the stored data. The relational mapping of the whole data allows for sophisticated data look up queries with SQL while machine learning research can more easily be performed with the treated data already stored. The database is harmonized with two other frameworks, namely with the Riffyn datalake software (Juergens et al.; 2018) which is being adopted by companies such as Novozymes and BASF. Here the Riffyn data agent allows to connect the DTU KT database to the Riffyn software. This will make it easier to join process data from ongoing experiments across different units to emulate a real process. The DTU KT database is also being harmonized with the AnIML data standard. This allows the captured data from a wide variety of laboratory equipment from different vendors to be integrated with the database. The AnIML data standard is a XML schema and is divided into AnIML Core and the AnIML Technique Definitions. For further information we refer to the reference literature (Schäfer et al.; 2004).

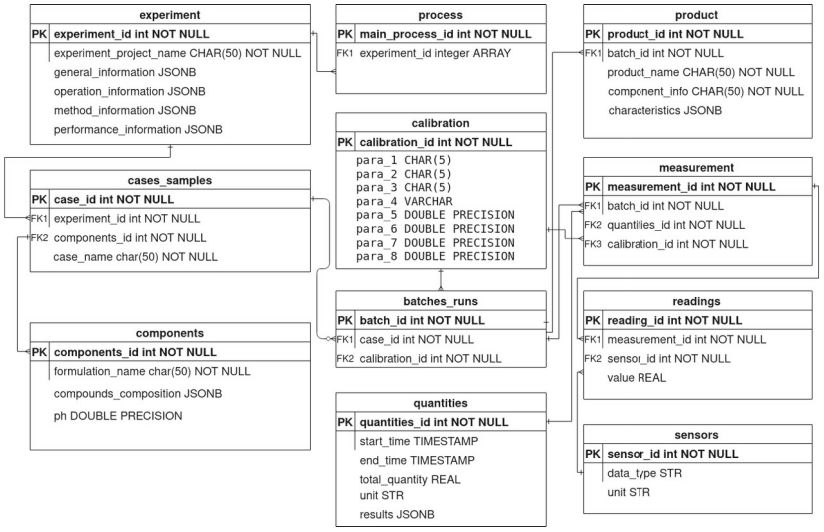


Figure 2: Current snapshot of the entity relationship (ER) diagram of the relational database under development

4. Digital Twin Integration

To facilitate the development and evaluation of digital twins for each individual unit operation, a generic digital twin framework has been developed and has been named internally at the department DelphiTwin (Nielsen et al., 2021). This framework is designed to create and test digital twins for dynamic processes, and to control the equipment during operation. The setup of DelphiTwin, requires a total of four steps where the user defines the desired level of complexity and number of modules to include. After the initial setup, DelphiTwin can subsequently be used for a number of tasks. During process operation, the digital twin will automatically collect and store sensor and actuator data in an internal local database. It will furthermore handle and send scheduled set points to low-level controllers. Based on the process model, real-time predictions of future process states can furthermore be generated. Finally, the specified control algorithms can be used to calculate and schedule future set points of low level controllers. In between process operations, DelphiTwin can also be used to investigate various design spaces for different configurations.

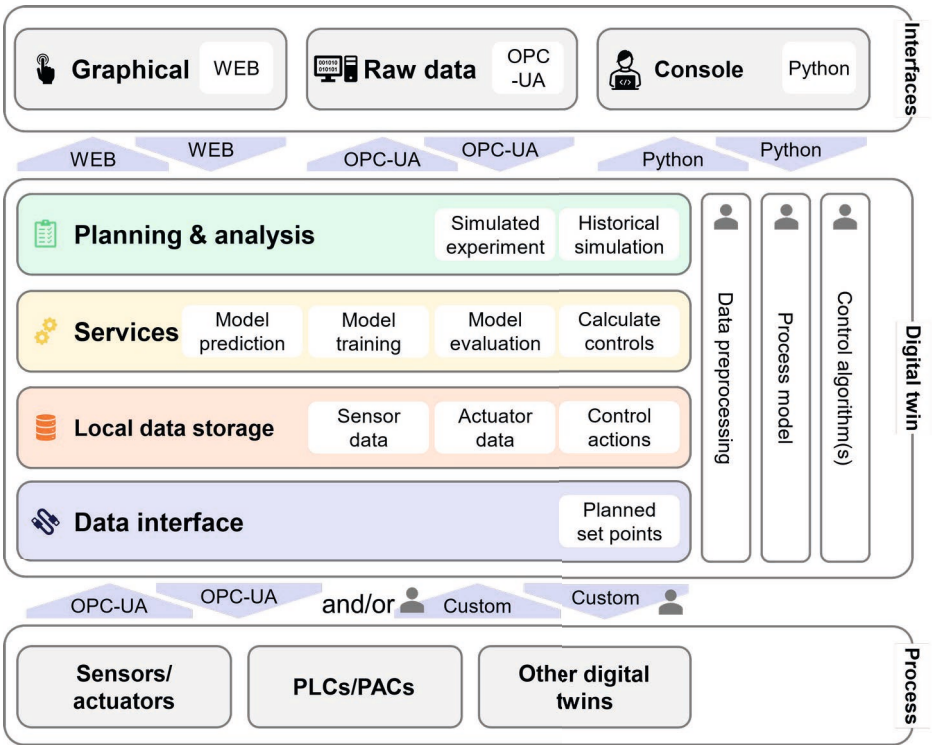


Figure 3: Digital twin framework of DelphiTwin

To allow for a high degree of flexibility when constructing a digital twin, DelphiTwin is implemented in Python with a modular architecture (as illustrated in Figure 3). Users can create custom modules to include system specific implementations of data interfaces, process models and control algorithms. DelphiTwin comes with a set of standard modules, including generic PID and MPC control algorithms and data-interfacing modules supporting OPC-UA, OPC-DA and TCP/IP. It is the plan to embed

this framework in the digital infrastructure to better utilize pilot plant equipment in future plant scale-up, design and optimization research.

5. Conclusions

This paper gives an overview of the current developments with respect to the digital transformation journey of an academic chemical and biochemical engineering department with a strong binding to industrial research. We present how a pilot plant can be retrofitted to accommodate the changes brought by the transition towards Industry 4.0. The efforts described in this article will provide a broad and foundational basis to perform advanced research tasks in a highly digitalized environment leveraging researchers' capabilities. From the educational point of view, the digital transformation of the department will allow to provide courses for students to equip them with skills for the digital era of chemical and biochemical engineering.

References

F. D. Bähner, O. A. Prado-Rubio and J. K. Huusom, Industrial & Engineering Chemistry Research, 2021, 60 (42), 14985-15003

C. L. Gargalo, S. C. de las Heras, M. N. Jones, I. Udugama, S. S. Mansouri, U. Krühne, K. V. Gernaey, Towards the Development of Digital Twins for the Bio-manufacturing Industry, Digital Twins: Tools and Concepts for Smart Biomanufacturing, 2021, 1-34.

H. Juergens, M. S. Niemeijer, L. D. Jennings-Antipov, R. Mans, J. Morel, A. J. A. van Maris, J. T. Pronk, T. S. Gardner, Evaluation of a novel cloud-based software platform for structured experiment design and linked data analytics, Scientific Data 5, Article number: 180195 (2018)

P. C. Lopez, I. A. Udugama, S. T. Thomsen, C. Roslander, H. Junicke, M. M. Iglesias, K. V. Gernaey., Transforming data to information: A parallel hybrid model for real-time state estimation in lignocellulosic ethanol fermentation, Biotechnology & Engineering, Volume 118, Issue 2, February 2021, Pages 579-591

R. F. Nielsen, N. Nazemzadeh, M. P. Andersson, K. V. Gernaey, S. S. Mansouri, An uncertainty-aware hybrid modelling approach using probabilistic machine learning, Computer Aided Chemical Engineering, Volume 50, 2021, Pages 591-597

S. Perazzoli, J. P. de Santana Neto, H. M. Soares, Prospects in bioelectrochemical technologies for wastewater treatment, Water Sci Technol (2018) 78 (6): 1237–1248.

S. Russo, M. D. Besmer, F. Blumensaat, D. Bouffard, A. Disch, F. Hammes, A. Hess, M. Lürig, B. Matthews, C. Minaudo, E. Morgenroth, V. Tran-Khac, K. Villez, The value of human data annotation for machine learning based anomaly detection in environmental systems, Water Research, Volume 206, 1 November 2021, 117695.

B. A. Schäfer, D. Poetz, G. W. Kramer, Documenting Laboratory Workflows Using the Analytical Information Markup Language, Journal of the Association for Laboratory Automation, Volume: 9 issue: 6, page(s): 375-381

I. A. Udugama, M. Öner, P. C. Lopez, C. Beenfeldt, C. Bayer, J. K. Huusom, K. V. Gernaey, G. Sin, Towards Digitalization in Bio-Manufacturing Operations: A Survey on Application of Big Data and Digital Twin Concepts in Denmark, Front. Chem. Eng., 16 September 2021.

H. Ziaei-Halimejani, N. Nazemzadeh, R. Zarghami, K. V. Gernaey, M. P. Andersson, S. S. Mansouri, N. Mostoufi, Fault diagnosis of chemical processes based on joint recurrence quantification analysis, Computers & Chemical Engineering, Volume 155, December 2021, 107549.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50255-4

Identification Method of Multiple Sequential Alarms that Occurred Simultaneously in Plant-operation Data

Ai Yanaga[a*], Masaru Noda[a]

[a]*Department of Chemical Engineering, Fukuoka University, 8-19-1, Nanakuma Jonan-ku, Fukuoka 814-0180, Japan*
mnoda@fukuoka-u.ac.jp

Abstract

Advances in distributed control systems in the chemical industry has made it possible to inexpensively and easily install numerous alarms in them. A poorly designed alarm system might cause nuisance alarms. One type of nuisance alarm is a sequential alarm, which reduces the capability of operators to cope with plant abnormalities because critical alarms are hidden in them. We propose an identification method of sequential alarms that occurred at the same time. With this method, similarities of all combinations of an alarm subsequence are compared using the Smith-Waterman algorithm (Smith *et al*., 1981). We introduced a new objective function considering the time differences between alarms into this algorithm. We applied the proposed method to the simulation data of an extractive distillation column, and the simulation results indicate that the method can extract sequential alarms that occurred simultaneously in plant-operation data.

Keywords: Big data, Operation, Alarm system, Chemical plants

1. Introduction

1.1. Plant alarm systems

A plant-alarm system notifies an operator of plant-state deviations. An alarm is issued when the process variable deviates from the range set in consideration of safety. Advances in distributed control systems (DCS) in the chemical industry has made it possible to inexpensively and easily install numerous alarms in them. While most alarms help operators detect an abnormality and identify its cause, some are unnecessary. We called such unnecessary alarms nuisance alarms.

There are three typical types of nuisance alarms: sequential, repeating, and those without operations. Sequential alarms consist of numerous alarms in succession triggered by a single root cause. Repeating alarms occur routinely. Alarms without operations do not require corresponding operation. Nuisance alarms reduce the ability of operators to cope with plant abnormalities because critical alarms are hidden in many of them.

1.2. Previous research

Cheng *et al*. (2013) proposed a method of calculating the similarities of alarm-flood sequences in plant-operation data using the Smith-Waterman algorithm. The Smith-Waterman algorithm is a local-sequence-alignment tool for identifying common molecular subsequences (Smith *et al*., 1981). Experts can conduct a thorough analysis, such as root cause, on the basis of the clustered patterns of alarm floods. However, it

cannot be used to directly determine sequential alarms hidden in the plant-operation data. Wang and Noda (2017) proposed a mining method of sequential alarms in plant-operation data using a dot matrix method (Mount, 2004). The dot matrix method is a sequence-alignment method for identifying similar regions in deoxyribonucleic acid (DNA) or ribonucleic acid (RNA), which may be a consequence of functional, structural, or evolutionary relationships between the sequences. Proposed method can identify sequential alarms from the operation data of chemical plants but occasionally fails to detect sequential alarms between two related sequential alarms because the time information when alarms occurred is not used for evaluating the similarities between them. The above methods also occasionally fail to identify sequential alarms when two types of sequential alarms occurred at the same time.

1.3. Objective

A poorly designed alarm system triggers nuisance alarms, which might lead to oversight of critical alarms. Such oversight might cause plant accidents. We propose an identification method of sequential alarms. It takes into account multiple sequential alarms that occurred simultaneously and the time distances among alarms.

2. Proposed method

2.1. Plant-operation data

Plant-operation data consist of the alarm tag of alarm or operation and occurrence times. There are two types of alarm tags, process variable and operation variable. Table 1 shows an example of plant-operation data, where "A" indicates an alarm or operation, subscript values indicate their numbers of alarm or operation. When an alarm occurs, it is recorded in the plant database. Table 1 shows that A_1, A_2, A_3, and A_4 occurred in the order $A_1 \rightarrow A_2 \rightarrow A_4 \rightarrow A_2 \rightarrow A_3$ in about 20 minutes. A huge amount of data is accumulated every day, where nuisance alarms are hidden.

Table 1 Example of plant-operation data

Date	Time	Tag	Type
2021/1/1	0:08:53	A_1	Alarm
2021/1/1	0:09:36	A_2	Operation
2021/1/1	0:11:42	A_4	Alarm
2021/1/1	0:25:52	A_2	Operation
2021/1/1	0:30:34	A_3	Alarm

2.2. Problem formulation

Plant-operation data are characterized by these tags and the order of alarm occurrence. When an alarm sequence occurs repeatedly in plant-operation data, those alarms are considered to be sequential. The problem of identifying sequential alarms in such data is formulated as the problem of searching for a repeated subsequence of alarms in plant operation data.

2.3. Preparation

Our proposed method converts plant-operation data into an alarm subsequence by using a sliding window (Mount, 2004), which prevents the window from extracting a subsequence in which the time distances among alarms are critically different. The alarm sequence is converted into a set of windows consisting of occurrence alarms in a pre-

determined time window. By enabling overlapping and extracting of alarm subsequences, it is possible to deal with cases in which multiple types of sequential alarms are issued at the same time.

2.4. Similarity evaluation method

The proposed method uses a modified Smith-Waterman algorithm (Smith & Waterman, 1981) for searching sequential alarms. It finds a pair of segments, one from each of two long sequences, such that there is no other pair of segments with greater similarity. It is a dynamic algorithm for finding the highest scored local sequence alignment and is partially used with the method by Cheng *et al*. (2013).

Their method sums three types of scores, i.e., the match score when the corresponding alarms match the similarity between the two alarm columns, mismatch score when they do not match, and insert score of the gap when there is a gap.

The final similarity score is calculated by adding these scores. In this research, a mismatch score and gap score are considered constant, and the match score is calculated on the basis of the time distances among alarms.

For example, consider the following two sequences.

$$S_1 = a_1, a_2, \ldots a_m, \ldots, a_M, m = 1, 2, \ldots, M \quad (1)$$

$$S_2 = b_1, b_2, \ldots, b_n, \ldots, b_N, n = 1, 2, \ldots, N, \quad (2)$$

where a_m is the mth tag in aligned segment S_1 of length M, and b_n is the nth tag in aligned segment S_2 of length N. The similarity evaluation method is given a pair of contiguous subsequences, one from each of the two sequences by inserting gaps in one or both of them. The similarity score is positive for a match and negative for a mismatch. For a symbolic pair including a gap symbol, the similarity score is negative as a penalty of inserting a gap.

When the alarms do not match, the score is weighted by the time differences among the alarms, and the similarity between the sequences is evaluated. With this method, to consider the time differences among alarms, which is a problem in previous researches, the match score when alarms between two sequences are matched individually is calculated on the basis of the time differences among alarms.

It first calculates the weight of the time distances among matching alarms using the scaled Gaussian function Eq.(3).

$$w(D) = \exp(-D^2 / 2\sigma^2), \quad (3)$$

Where w is the weight based on the time distances among alarms, D, and σ is the standard deviation. The scaled Gaussian function is a normal distribution expressed from 0 to 1. The method then calculates the similarity score using Eq.(4), where s is the similarity score, μ is the mismatch score, and $s_{\max}$ is the maximum match score.

$$s(a_m, b_n) = \begin{cases} \mu(1 - w) + s_{\max} \times w & \text{if } a_m = b_n \\ \mu & \text{if } a_m \neq b_n \end{cases}, \quad (4)$$

Eq.(4) indicates that a high match score means high similarity between those subsequences. The match score is close to the mismatch score when there is a long time distance and the maximum match score is a close range.

To obtain the alignment between S_1 and S_2, the score matrix H is calculated using the dynamic programming method. Eq.(5) expresses $H_{m,n}$ between a_m and b_n, where δ is the penalty of inserting a gap.

$$H_{m,n} = \max\{H_{m\text{-}1,n\text{-}1} + s(a_m,b_n), H_{m\text{-}1,n} + \delta, H_{m,\,n\text{-}1} + \delta, 0\}, \tag{5}$$

The Smith-Waterman algorithm finds the pair of segments with maximum similarity by first locating the maximum element because it indicates an optimal alignment. The Smith-Waterman algorithm searches for pairs of maximally similar subsequences on a mathematically rigorous basis. It can extract an optimum subsequence when multiple sequential alarms occurred simultaneously.

2.5. Grouping

The smith-Waterman algorithm cannot aggregate similar sequential alarm patterns when the same sequential alarm occurred several times in the plant-operation data. This algorithm is a local-sequence-alignment tool that searches for a pair of segments, one from each of two long sequences. The proposed method creates a color map on the basis of the similarity score of extracted sequential alarm patterns, and aggregates the patterns using the single-linkage method. The color map is used to identify sequential alarms if the similarity score is any given threshold or above. The proposed method can effectively aggregate similar sequential alarm groups when sequential alarms occurred several times.

2.6. Calculation procedure

The proposed method's calculation procedure is as follows.

(1) Converts plant-operation data into an alarm subsequence by using a pre-determined time window.
(2) Calculates the similarity score by using a modified Smith-Waterman algorithm.
(3) Identifies sequential alarms in accordance with a cluster of sequential alarm patterns by using the similarity-score color map.

3. Case study

3.1. Simulation Data

We applied the proposed method to the simulation data of the extractive distillation column shown in Fig.1. There was a total of 18 alarms in the DCS, and three types of malfunctions, low flow rate of coolant, low steam pressure, and valve stiction, were artificially induced in the process simulation. A defined operation for each malfunction was carried out after each malfunction occurred. Alarm occurrences were recorded in the plant-operation data. During a process simulation of 15 days, 265 alarms and operations were recorded in the plant-operation data. The grey area in Fig.2 is region of where two types of malfunctions occurred at the same time.

3.2. Results of identified sequential alarms

We set the window size to 120 minutes, maximum match score to 1, mismatch score to -0.6, gap score to -0.4, and delta to 42.8390. The method extracted a subsequence as a sequential alarm pattern when the similarity score was equal to or higher than 10. As a

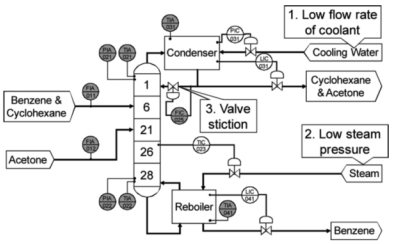


Fig.1 Process flow of extractive distillation column

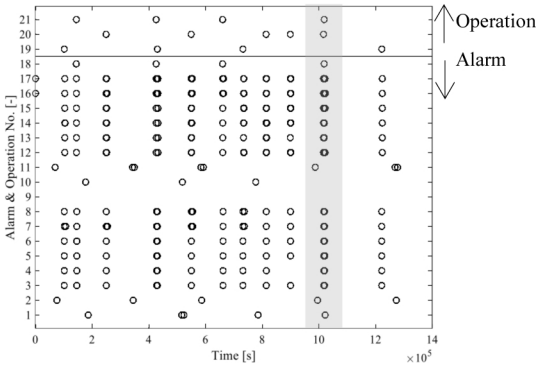


Fig.2 Simulated plant-operation data

result, it extracted 17 sequential alarm patterns and identified 3 types of sequential alarms in accordance with a cluster of patterns by using the similarity-score color map.

Fig.3 shows the identified sequential alarm No.1. X-axis is occurrence time of alarm and operation from the first alarm and y-axis is IDs of alarms and operations. There are 4 sequences identified as sequential alarm No.1. These alarms are considered to be sequential alarms issued due to a low flow rate of coolant based on the type and order of the alarms being issued. It is necessary to consider changing the alarm setting range so that this sequential alarm does not occur. All plot is overlapping, that the proposed method identifies sequential alarms correctly.

Fig.4 shows the identified sequential alarms in the grey area in Fig.2. The x-axis shows the occurrence times of alarms and operations, and the y-axis shows the IDs of alarms and operations. Because we wanted to verify if the proposed method can identify multiple sequential alarms that occurred at the same time, the grey area expanded. Between 1,018,000 to 1,019,000 seconds, 2 types of mulfanctions occurred simultaneously. Fig.4 shows that it has be identified as 2 types of sequential alarms as expected. Even though several sequential alarms occurred simultaneously, this method identified other type of sequential alarm.

4. Conclusions

We proposed a method for mining repeated similar sequential alarms. The method can identify multiple types of sequential alarms in plant-operation data even where the sequential alarms occurred simultaneously. This method can be used in more complex programs because it does not require process information.

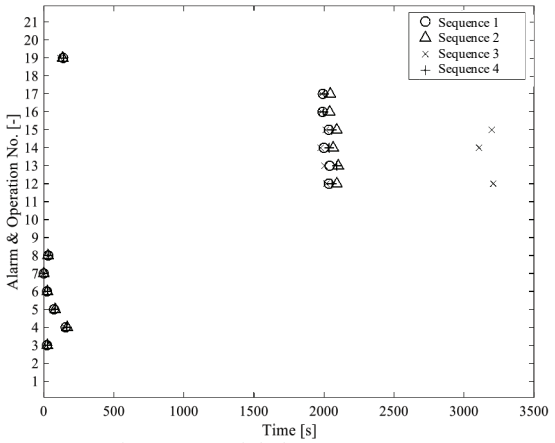


Fig.3 Sequential alarm No. 1

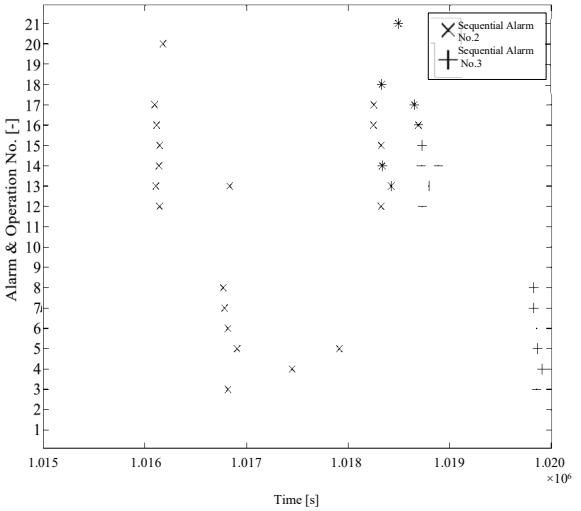


Fig.4 Results of identified sequential alarms of grey area in Fig.2

References

Cheng, Y., Izadi, I., and Chen, T., Pattern matching of alarm flood sequences by a modified Smith-Waterman algorithm, *Chemical Engineering Research and Design,* 91, 1085–1094 (2013)

Mount, D. W., Bioinformatics Sequence and Genome Analysis Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (2004)

Nishiguchi, J., and Takai, T., IPL2 and 3 Performance Improvement Method for Process Safety using Event Correlation Analysis, *Computers & Chemical Engineering,* 34, 2007–20013 (2010)

Smith, T., Waterman, M., Identification of common molecular subsequences, *Journal of Molecular Biology,* 147, 195–197 (1981)

Wang, Z., and Noda, M., Identification of repeating Sequential Alarms in Noisy Plant Operation Data Using Dot Matrix Method with Sliding Window, *Journal of Chemical Engineering of Japan,* 50, 445–449 (2017)

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50256-6

Understand how CNN diagnoses faults with Grad-CAM

Deyang Wu[a], Jinsong Zhao[a,b*]

[a]*State Key Laboratory of Chemical Engineering, Department of Chemical Engineering, Tsinghua University, Beijing 100084, China*
[b]*Beijing Key Laboratory of Industrial Big Data System and Application, Tsinghua University, Beijing 100084, China*
jinsongzhao@tsinghua.edu.cn

Abstract

CNN-based models for fault diagnosis have achieved high prediction accuracy, but the lack of explainability makes them hardly be understood by humans. In this paper, a technique used to produce visual explanations for CNN has been introduced to a CNN-based fault diagnosis model, DCNN, to make it more transparent and understandable. Experiments on Tennessee Eastman process showed variables that DCNN pays more attention to when diagnosing faults, which makes the decision making process of DCNN more explainable and understandable.

Keywords: fault diagnosis, explainable deep learning, convolutional neural network, chemical process safety

1. Introduction

Fault detection and diagnosis (FDD) is quite critical to safe operations of chemical processes to identity abnormal events that can hardly be controlled by distributed control systems (DCS). In decades, many researchers have been proposed different models of real-time FDD for processes. These models can be categories into quantitative model-based, qualitative model-based and process history-based models (Venkatasubramanian et al. 2003). Process history-based models are further classified into qualitative and quantitative models. The latter is also commonly termed as data-driven models. Data-driven models based on deep learning have drawn much attention of researchers these years, such as DBN-based models (Z. Zhang and Zhao 2017), CNN-based models (H. Wu and Zhao 2018), RNN-based models (S. Zhang, Bi, and Qiu 2020), autoencoder-based models (Cheng, He, and Zhao 2019) (Zheng and Zhao 2020) and GCN-based models (D. Wu and Zhao 2021). These models achieved high prediction accuracy, but the lack of explainablity makes them hard to understand by humans. The research of explainable deep learning models is critical to the application promotion for fault diagnosis.

In this paper, a technique used to produce visual explanation for CNN has been introduced to a CNN-based fault diagnosis model, DCNN, to make it more transparent and understandable. Grad-CAM was firstly introduced in Section 2, then Grad-CAM was applied to DCNN model with Tennessee Eastman process as the benchmark in Section 3, and the conclusio was drawn in Section 4 finally.

2. Explainable CNN and Grad-CAM

In recent years, deep learning had made great achievements in many research areas. Deep learning has many advantages such as low-cost modelling, self-directed learning from data and high prediction accuracy. However, the inference process of deep learning models cannot be fully understood by human beings. To understand how models make decisions, many researchers have been studying explainable deep learning recently.

CNN, as an important class of deep learning models, has made great success in the fields of computer vision and other engineering. Many models and methods have been proposed to make CNN more understandable. Grad-CAM (Selvaraju et al. 2020) is a technique for producing visual explanations for decisions from CNN based models, making them more transparent and explainable. As a gradient-based method, Grad-CAM uses the class-specific gradient information flowing into the final convolutional layer of a CNN in order to produce a coarse localization map of the important regions in the image when it comes to classification (Linardatos, Papastefanopoulos, and Kotsiantis 2021). Figure 1 is an example of Grad-CAM applied to an image classification task in (Selvaraju et al. 2020). It shows that Grad-CAM can highlight the key region on the image corresponding to different classification results. This localization map can be regarded as the basis for decision making of CNN, which is also reasonable and understandable to humans.

Figure 1: Example of Grad-CAM applied to an image classification task in (Selvaraju et al. 2020). (a) Original image with a cat and a dog. (b-c) Support for the cat or the dog category according to various visualizations for VGG-16.

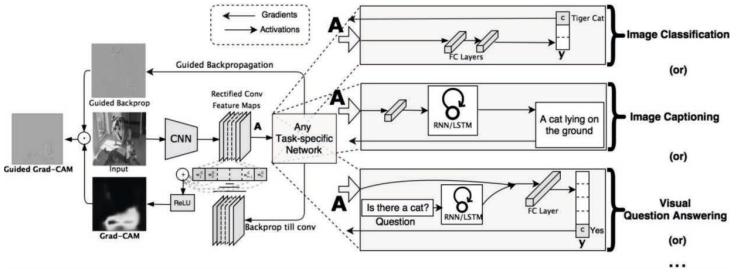


Figure 2: Grad-CAM overview

The workflow of computing the class-discriminative localization map is shown as Figure 2. Given a CNN that has been well trained for a classification task and a specific class c, the neuron importance weights α_k^c of feature maps in a convolutional layer can be calculated as Eq.(1).

$$\alpha_k^c = \frac{1}{Z}\sum_i\sum_j \frac{\partial y^c}{\partial A_{ij}^k} \tag{1}$$

y^c is the score for class c in the last fully connected layer before the softmax layer. A_{ij}^k is the activations of the k^{th} feature map (indexed by i and j respectively over the width and height dimensions) in the last convolutional layer. Eq.(1) means the gradients propagated from outputs back to feature maps, and α_k^c captures the importance of the k^{th} feature map for a target class c. The localization map Grad-CAM can be calculated as Eq.(2).

$$L_{Grad-CAM}^c = ReLU\left(\sum_k \alpha_k^c A^k\right) \tag{2}$$

The localization map is finally obtained by weighted summing all the k feature map activations and applying the ReLU function.

3. Understand how CNN diagnoses faults

3.1. CNN-based fault diagnosis

Since CNN-based models have been applied to fault diagnosis successfully, Grad-CAM is also suitable for the interpretability of these models to study how CNNs diagnose faults for processes. In these models, fault diagnosis of a chemical process is generally regarded as a classification problem. Given an observation $\boldsymbol{X}_t \in \mathbb{R}^{v\times w}$ at time t, the current operating state $y \in \mathbb{N}$ of the process should be identified out of a set of operating states consisting of normal state and different types of faults. The observation $\boldsymbol{X}_t$ is a matrix with the time window and process variable dimensions. If $w > 1$, it means the observation includes data before t in a time window w as shown in Eq.(3).

$$\boldsymbol{X}_t = \begin{bmatrix} x_{1,t} & x_{1,(t-1)} & \cdots & x_{1,(t-w+1)} \\ x_{2,t} & x_{2,(t-1)} & \cdots & x_{2,(t-w+1)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{v,t} & x_{v,(t-1)} & \cdots & x_{v,(t-w+1)} \end{bmatrix} \tag{3}$$

The operating state or class y is integer between 0 and N_c (the number of different fault states). $y = 0$ means the process is under normal operating state, and $y = c$ ($1 \le c \le N_c$) means the fault c has occurred. With enough pairs of observations and fault labels, a dataset $\left\{\boldsymbol{X}_t^{(i)}, y^{(i)}\right\}_{i=1}^N$ is obtained for training and testing a CNN-based fault diagnosis model.

In this paper, the DCNN model proposed by (H. Wu and Zhao 2018) is used to show how Grad-CAM can be applied to CNN-based fault diagnosis. The chosen architecture is 'Conv(64)-Conv(64)-Pool-Conv(128)-Pool(2×1)-FC(300)*-FC(21)' (model 7).

3.2. Tennessee Eastman process

Tennessee Eastman (TE) process (Downs and Vogel 1993) is a simulation process model modified based on an actual industrial process of Eastman Chemical Company in Tennessee, USA. In decades, TE process has been utilized for research in different

fields such as process control, process monitoring and so on. TE process mainly consists of five unit operations and defines 20 different process disturbances. In the simulation program of TE process, 52 variables can be observed, and 20 different disturbances can be inserted at any time to make the process operate in a fault state.

3.3. Data preparation and model training

The process of data preparation and model training keeps the same with (D. Wu and Zhao 2021). To obtain enough data for training the DCNN model, the simulation program of TE process ran for 3000 h in normal state. Under every different fault, the simulation program ran for 20 h after the fault inserted, and 10 sets of parallel simulations were carried out. All the data have been normalized with the mean and standard deviation calculated from data in the normal state. The normalized data were then cut to slices with the time widow of 1 h, and the data are sampled every 3 min. Data from 8 sets of parallel simulations were used for training and the rest 2 sets were used for testing.

The DCNN model was trained for 50 epochs with Adam optimizer, the mini-batch size was set to 128 and the learning rate was set to 0.001. The trained DCNN model finally got an average classification rate of 0.93.

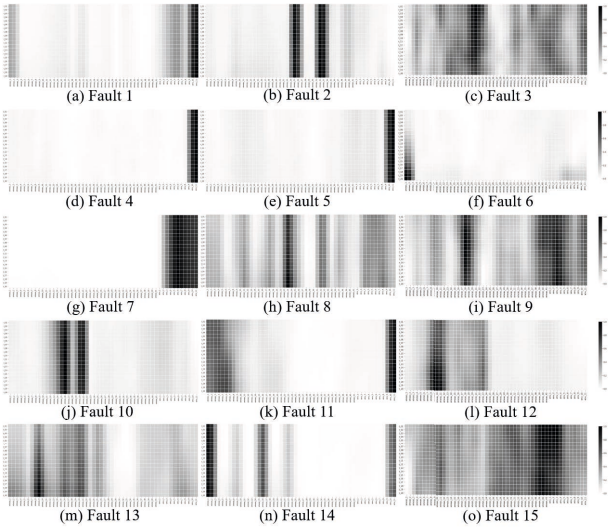


Figure 3: Average Grad-CAM localization map of 15 different faults of TE process

3.4. Explain fault diagnosis results with Grad-CAM

After the DCNN model was well trained, Grad-CAM was applied to the model following Eq.(1) and Eq.(2). The localization map of Grad-CAM is calculated based on a certain instance or observation i. In this paper, we care more about the variables but not time windows that DCNN focuses on when diagnosing faults. Thus, the average localization map for a certain type of fault c was calculated with all the maps obtained from all observations of the fault c as Eq.(4).

$$\bar{L}^{c}_{Grad-CAM} = \frac{1}{N}\sum_{i=1}^{N} L^{c}_{(i)} \tag{4}$$

$\bar{L}^{c}_{Grad-CAM}$ was then normalized using its maximum and minimum values to make every value falls into [0,1]. $\bar{L}^{c}_{Grad-CAM}$ of the first 15 faults with known causes in Table 1 is shown as Figure 3. When values changes from 0 to 1, the colour in Figure 3 changes from white to black. Figure 3 shows that what variables DCNN mainly focused on when diagnosing different faults. The deeper the colour is, the more important the variable is. Grad-CAM explains the basis of decision making of DCNN for fault diagnosis in a visual and understandable way.

Table 1: Comparison between variables affected by disturbances firstly and variables CNN focused on when diagnosing faults. T, P, L, Q, x, u mean temperature, pressure, level, flow rate, composition, control signal respectively. Subscript means unit operations or streams. Superscript means characteristics or components.

Number	Variables affected firstly	Variables CNN focused on
IDV (1)	$x^{A}_{feed}, x^{C}_{feed}, Q_{s1}, Q_{s4}, u^{Q}_{s1}, u^{Q}_{s4}$	$u^{cwr}_{cond}, u^{cwr}_{reac}, u^{steam}_{strip}, u^{recy}_{comp}, u_{purge}$
IDV (2)	$x^{A}_{feed}, \boldsymbol{x^{B}_{feed}}, \boldsymbol{x^{C}_{feed}}, Q_{s1}, Q_{s4}, u^{Q}_{s1}, u^{Q}_{s4}$	$x^{D}_{purge}, \boldsymbol{x^{C}_{feed}}, x^{E}_{purge}, x^{D}_{feed}, \boldsymbol{x^{B}_{feed}}$
IDV (3)	$T_{reac}, P_{reac}, L_{reac}, \boldsymbol{T^{cwr}_{reac}}, u^{cwr}_{reac}$	$\boldsymbol{T^{cwr}_{reac}}, W_{comp}, T^{cwr}_{cond}, Q^{steam}_{strip}, u^{cwr}_{cond}$
IDV (4)	$T^{cwr}_{reac}, \boldsymbol{u^{cwr}_{reac}}, T_{reac}, P_{reac}, L_{reac}$	$u^{cwr}_{cond}, \boldsymbol{u^{cwr}_{reac}}, u^{steam}_{strip}, Q_{s1}, Q_{s2}$
IDV (5)	$T^{cwr}_{cond}, \boldsymbol{u^{cwr}_{cond}}, T_{sep}, P_{sep}, L_{sep}$	$\boldsymbol{u^{cwr}_{cond}}, u^{cwr}_{reac}, u^{steam}_{strip}, u^{Q}_{s2}, x^{H}_{product}$
IDV (6)	$\boldsymbol{Q_{s1}}, u^{Q}_{s1}, Q_{feed}, x^{A}_{feed}$	$\boldsymbol{Q_{s1}}, Q_{s2}, Q_{s3}, u^{recy}_{comp}, u_{purge}$
IDV (7)	$Q_{s4}, u^{Q}_{s4}, T_{strip}, P_{strip}, L_{strip}$	$u_{purge}, u^{under}_{sep}, u^{recy}_{comp}, u^{under}_{strip}, u^{steam}_{strip}$
IDV (8)	$\boldsymbol{x^{A}_{feed}}, \boldsymbol{x^{B}_{feed}}, x^{C}_{feed}, Q_{s1}, Q_{s4}, u^{Q}_{s1}, u^{Q}_{s4}$	$\boldsymbol{x^{A}_{feed}}, \boldsymbol{x^{B}_{feed}}, T^{cwr}_{cond}, x^{D}_{purge}, x^{E}_{purge}$
IDV (9)	$T_{reac}, P_{reac}, L_{reac}, T^{cwr}_{reac}, u^{cwr}_{reac}$	$T_{strip}, u^{Q}_{s1}, u^{Q}_{s3}, u^{Q}_{s4}, Q^{steam}_{strip}$
IDV (10)	$T_{strip}, \boldsymbol{P_{strip}}, \boldsymbol{L_{strip}}, u^{steam}_{strip}, Q^{steam}_{strip}$	$\boldsymbol{P_{strip}}, T^{cwr}_{reac}, W_{comp}, \boldsymbol{L_{strip}}, Q^{under}_{strip}$
IDV (11)	$T^{cwr}_{reac}, \boldsymbol{u^{cwr}_{reac}}, T_{reac}, P_{reac}, L_{reac}$	$u^{cwr}_{cond}, \boldsymbol{u^{cwr}_{reac}}, Q_{s4}, Q_{s3}, Q_{recy}$
IDV (12)	$T^{cwr}_{cond}, u^{cwr}_{cond}, \boldsymbol{T_{sep}}, P_{sep}, \boldsymbol{L_{sep}}$	$T_{reac}, Q_{purge}, \boldsymbol{T_{sep}}, L_{reac}, \boldsymbol{L_{sep}}$
IDV (13)	$\boldsymbol{T_{reac}}, P_{reac}, L_{reac}, \boldsymbol{T^{cwr}_{reac}}, u^{cwr}_{reac}$	$\boldsymbol{T_{reac}}, W_{comp}, \boldsymbol{T^{cwr}_{reac}}, L_{reac}, P_{strip}$
IDV (14)	$T^{cwr}_{reac}, \boldsymbol{u^{cwr}_{reac}}, T_{reac}, P_{reac}, L_{reac}$	$Q_{s1}, Q_{s2}, P_{strip}, u^{cwr}_{cond}, \boldsymbol{u^{cwr}_{reac}}$
IDV (15)	$T^{cwr}_{cond}, u^{cwr}_{cond}, T_{sep}, P_{sep}, L_{sep}$	$x^{G}_{product}, x^{H}_{product}, x^{F}_{product}, u^{Q}_{s2}, u^{Q}_{s3}$

It should be noted that the important variables in Figure 3 are not necessarily the disturbance variables. Table 2 concludes the differences between the variables affected by disturbances firstly (observable root causes) and the first 5 variables that have highest values in Grad-CAM localization maps for every type of fault. For some faults, Grad-CAM localized some of the root causes or observable variables affected directly by root causes. For other faults, DCNN focused none of the root causes. This indicates that DCNN pays more attention to those sensitive variables that changes intensely with disturbances but not disturbance variables themselves. For those faults that are hard to identity (such as fault 3, 9 and 15), all variables only change slightly after the fault has occurred. Thus, DCNN can hardly localize some certain variables to determine the type of fault. This also reflects on the Grad-CAM localization maps that the variance of all the values in is small and they have little differences in colour.

4. Conclusions

In this paper, Grad-CAM, a technique for producing visual explanations for CNN, is introduced. Grad-CAM uses the gradient information flowing back into the last convolutional layer of the CNN to assign importance values to each neuron for a particular decision of interest (Selvaraju et al. 2020). Grad-CAM was applied to the DCNN model for fault diagnosis to make the inference process transparent. With a well trained DCNN model, Grad-CAM generates a localization map for an input matrix to show the importance of different neurons. This can help humans understand what variables DCNN pays more attention to when diagnosing the type of faults of the process. The results of experiments on TE process shows that the DCNN model pays more attention to those sensitive variables that changes intensely with disturbances but not disturbance variables themselves. When all variables only change slightly after faults happen, it'll be difficult for DCNN to identity the type of fault. In this case, the localization maps have only little differences in colour in different region. Future work will focus on construct more explainable deep learning model for fault diagnosis.

References

Cheng, Feifan, Q. Peter He, and Jinsong Zhao. 2019. "A Novel Process Monitoring Approach Based on Variational Recurrent Autoencoder." Computers & Chemical Engineering 129 (October): 106515. https://doi.org/10.1016/j.compchemeng.2019.106515.

Downs, James J, and Ernest F Vogel. 1993. "A Plant-Wide Industrial Process Control Problem." Computers & Chemical Engineering 17 (3): 245–55.

Linardatos, Pantelis, Vasilis Papastefanopoulos, and Sotiris Kotsiantis. 2021. "Explainable AI: A Review of Machine Learning Interpretability Methods." Entropy 23 (1): 18. https://doi.org/10.3390/e23010018.

Selvaraju, Ramprasaath R., Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. 2020. "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization." International Journal of Computer Vision 128 (2): 336–59. https://doi.org/10.1007/s11263-019-01228-7.

Venkatasubramanian, Venkat, Raghunathan Rengaswamy, Kewen Yin, and Surya N Ka. 2003. "A Review of Process Fault Detection and Diagnosis Part I: Quantitative Model-Based Methods." Computers and Chemical Engineering, 19.

Wu, Deyang, and Jinsong Zhao. 2021. "Process Topology Convolutional Network Model for Chemical Process Fault Diagnosis." Process Safety and Environmental Protection 150 (June): 93–109. https://doi.org/10.1016/j.psep.2021.03.052.

Wu, Hao, and Jinsong Zhao. 2018. "Deep Convolutional Neural Network Model Based Chemical Process Fault Diagnosis." Computers & Chemical Engineering 115 (July): 185–97. https://doi.org/10.1016/j.compchemeng.2018.04.009.

Zhang, Shuyuan, Kexin Bi, and Tong Qiu. 2020. "Bidirectional Recurrent Neural Network-Based Chemical Process Fault Diagnosis." Industrial & Engineering Chemistry Research 59 (2): 824–34. https://doi.org/10.1021/acs.iecr.9b05885.

Zhang, Zhanpeng, and Jinsong Zhao. 2017. "A Deep Belief Network Based Fault Diagnosis Model for Complex Chemical Processes." Computers & Chemical Engineering 107 (December): 395–407. https://doi.org/10.1016/j.compchemeng.2017.02.041.

Zheng, Shaodong, and Jinsong Zhao. 2020. "A New Unsupervised Data Mining Method Based on the Stacked Autoencoder for Chemical Process Fault Diagnosis." Computers & Chemical Engineering 135 (April): 106755. https://doi.org/10.1016/j.compchemeng.2020.106755.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50257-8

A Comprehensive Framework for the Modular Development of Condition Monitoring Systems for a Continuous Dry Granulation Line

Rexonni B. Lagare[a*], M. Ziyan Sheriff[a], Marcial Gonzalez[b], Zoltan Nagy[a], Gintaras V. Reklaitis[a]

[a]*Davidson School of Chemical Engineering, Purdue University, West Lafayette, IN 47907, USA*
[b]*School of Mechanical Engineering, Purdue University, West Lafayette, IN 47907, USA*
rlagare@purdue.edu

Abstract

The development of condition monitoring systems often follows a modular scheme where some systems are already embedded in certain equipment by their manufacturers, and some are distributed across various equipment and instruments. This work introduces a framework for guiding the modular development of monitoring systems and integrating them into a comprehensive model that can handle uncertainty of predictions from the constituent modules. Furthermore, this framework improves the robustness of the modular condition monitoring systems as it provides a methodology for maintaining quality assurance and preventing unnecessary shutdowns in the event of some modules going off-line due to condition-based maintenance interventions.

Keywords: Condition Monitoring, Probabilistic Programming, Modular, Machine Learning, Bayesian

1. Introduction

The challenges in modeling pharmaceutical powder processes, as outlined in Rogers and Ierapetritou (2014), has put an emphasis on the use of data-driven models as the basis for developing condition monitoring (CM) systems. While this approach offers a practical solution, as Webb and Romagnoli (2021) recently demonstrated for the Tennessee Eastman Process (TEP) case study, it ignores the modular nature of process control system development.

Since the data-driven models often require data spanning multiple unit operations in order to maximize the usage of process data, CM applications are likely to be on levels 1 and 2, which are distributed control systems in the process control implementation hierarchy introduced by Su *et al.* (2019). These modules are also likely to be focused on process faults, which differ from the CM modules at level 0, which are directly embedded into more advanced equipment. Because embedded modules are developed by the vendors, whose priority is on the safe and reliable operation of the equipment, they tend to focus on safety-related faults like electrical and mechanical faults.

All the aforementioned fault types need to be considered holistically, especially since they are likely correlated with each other. However, the varying levels at which these modules are installed in the process control implementation hierarchy, and the difference in their goals, create an integration challenge that needs to be addressed in order to have a safe and reliable operation of a continuous processing plant.

2. Condition Monitoring Framework Development

A natural framework for addressing this integration issue would be the probabilistic graphical modeling methodology, which is commonly used to implement hierarchical Bayesian models. As recently demonstrated by Radcliffe and Reklaitis (2021), this methodology is effective in systems where data is limited and there is significant uncertainty in the model parameters. Since this method is fundamentally based on modularity, where complex physical systems are constructed from simpler parts, it is sensible to utilize it for the CM module integration problem.

Under this methodology, the basic parts of the system are random variables with uncertain values, which are depicted as nodes in the graphical model. Conditional dependencies may be assigned based on expert knowledge on the system, and arcs can be drawn between one or more nodes to capture these relationships. Altogether, the nodes and the arcs comprising the probabilistic graphical model (PGM) form a compact representation of joint probability distributions where probability theory can be used to model the uncertainty in these variables and to make inferences on variables of interest.

2.1. A Probabilistic Condition Monitoring Model for Continuous Dry Granulation

The continuous dry granulation line of the Purdue University Pilot Plant comprises several unit operations that can blend pharmaceutical excipients and active pharmaceutical ingredients (API), granulate them, and then compress them into tablets using a rotary tablet press (see Figure 1). At the heart of this process is the granulation step, which takes place in an Alexanderwerk WP-120 roller compactor (RC). In the RC unit, the pharmaceutical blend is compacted into a ribbon, cut into flakes, and subsequently broken down into the desired granule size distribution in a classifier mill. For clarity in presenting basic concepts, the remaining discussion will focus on the RC.

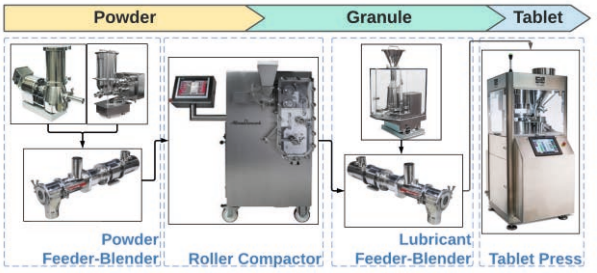


Figure 1. Dry granulation line at the Purdue University pilot plant.

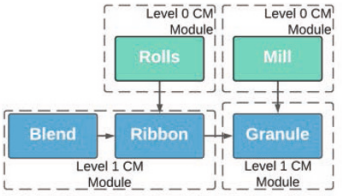


Figure 2. Roller compactor condition monitoring model.

A condition monitoring model can be constructed by considering two types of condition variables, the material condition and the equipment condition, and then forming appropriate relationships between them. For the RC, this model is shown in Figure 2. The roll and mill variables (green nodes) represent the condition of the respective roller compactor components. The WP-120 RC has a built-in condition monitoring system for each of these components, so these variables also represent “embedded” or “level 0” CM modules. The blend, ribbon, and granules variables do not have condition monitoring systems by default; so, these need to be developed as “distributed” or “level 1” CM modules that require the integration of additional PAT tools. Discussing the development of these modules is beyond the scope of the current paper, so they will be

assumed to provide uncertain values for their corresponding variables, as is the case for the embedded modules.

At this stage, the model in Figure 2 serves as a useful guideline for the modular development of CM modules that supports the complete observability of variables pertinent to Quality-by-Design (QBD) principles. Under QBD, the Critical-Quality-Attributes (CQA) targets of a unit operation should be achieved by controlling the Control Process Parameters (CPP) and the CQA of the preceding unit operation. For the RC model, the ability to control the CPP is represented by the condition of the equipment (i.e., the rolls and the mill), and the CQAs are represented by the condition of the material (i.e., blend, ribbon, and granules).

Each random variable in this graph, whether it is a material condition (blue node) or equipment condition (green node), can have discrete states: normal or a faulty state. With multiple variations possible for each faulty state, each node represents a categorical distribution, which assigns a probability for each possible state. For clarity, these distributions are depicted as probability tables that are linked to its corresponding node via broken lines in Figure 3. Moreover, the variables are assumed to take only two possible states, whereas in reality, they can have up to "N" number of states, depending on the number of faulty conditions that are recognized for each node.

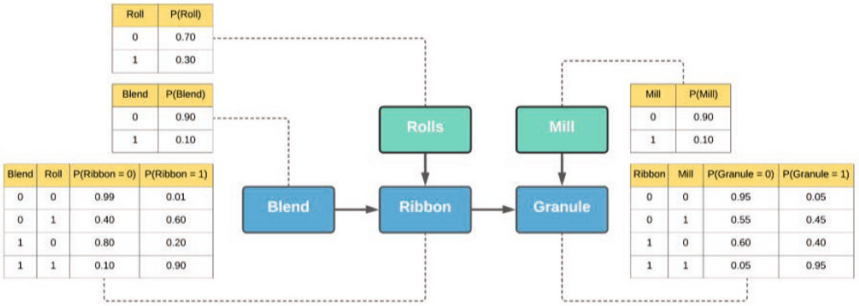


Figure 3. A probabilistic condition monitoring model for the roller compactor.

2.2. Probabilistic Programming and Inference

By the basic laws of probability, the entire graph represents the joint probability of all the condition monitoring variables per the following equation:

$$P(Blend, Rolls, Ribbon, Mill, Granule) = P(Rolls)P(Mill)P(Blend)P(Ribbon \mid Blend, Rolls)P(Granule \mid Mill, Ribbon) \quad (1)$$

With this model, interesting analysis tasks such as probabilistic inference can be performed. For example, given observations on the condition of the roll and the ribbon, e.g., both are at normal state so their values equal 0, it is possible to directly compute the posterior distribution of the condition of the blend using Bayes' Rule.

$$P(Blend \mid Rolls = 0, Ribbon = 0) = \frac{P(Blend, Rolls=0, Ribbon=0, Mill, Granule)}{P(Rolls=0, Ribbon=0)} \quad (2)$$

where: $\mathrm{P}(\mathrm{Rolls} = 0, \mathrm{Ribbon} = 0) =$ (3)

$$\sum_{Blend,Mill,Granule} P(Blend, Rolls = 0, Ribbon = 0, Mill, Granule)$$

However, with modularity in mind, this model is expected to get bigger as adjacent unit operations along the manufacturing line are integrated. The increasing number of variables will slow down the exact inference computations to a point that makes it impractical for monitoring applications. To circumvent this, the graphical model can instead be encoded in a probabilistic programming framework like Infer.NET (Minka et al., 2018), where approximate inference tasks can be quickly performed via efficient message passing algorithms.

3. Results and Discussion

3.1. Parameter Learning

As demonstrated, a fully-defined model such as shown in Figure 3 can make useful predictions on variables based on observations from other variables. However, in practice, these parameters are not always initially available. Fortunately, the graphical modeling methodology can perform parameter learning by simply adding the parameters as variables in the graph, and then using the same approximate inference techniques to infer parameter values. Figure 4 shows the modified graph that addresses parameter learning; the yellow nodes represent the prior probabilities of the CM modules, and the block arrows depict message passing during the inference of the prior probabilities. In order for the message passing algorithms to remain computationally tractable as more modules are integrated, the probabilities of the parameter variables are assigned a Dirichlet distribution, which is a conjugate prior for a categorical distribution. This conjugacy ensures that the number of distribution parameters do not increase intractably during the implementation of the message passing algorithms. (Winn, Bishop, and Diethe 2015)

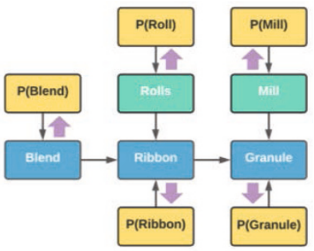


Figure 4. Parameter learning in a probabilistic graphical model.

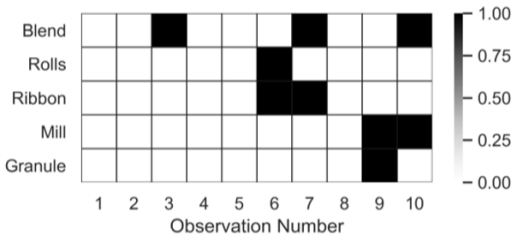


Figure 5. Dataset required for parameter learning (where all condition variables are observed).

Initially, the prior variables can be assigned either non-informative or weakly informative priors. Then, parameter learning can be performed on data acquired when all the CM modules are functional (see Figure 5), and message passing algorithms can be used to infer the posterior distributions of the parameter variables. As more data is collected, the inferred distribution of the parameter variables would be more "informed" and have less variance.

This can be observed from the results in Figure 6, which shows the inferred probabilities of the blend and mill condition. After just 100 observations, the comprehensive model was able to correctly infer the "true" probabilities of the model shown in Figure 3, which is from where the dataset was randomly sampled. Beyond this

number of observations, the mean of the inferred probability distributions barely changed, while the variance continued to decrease significantly.

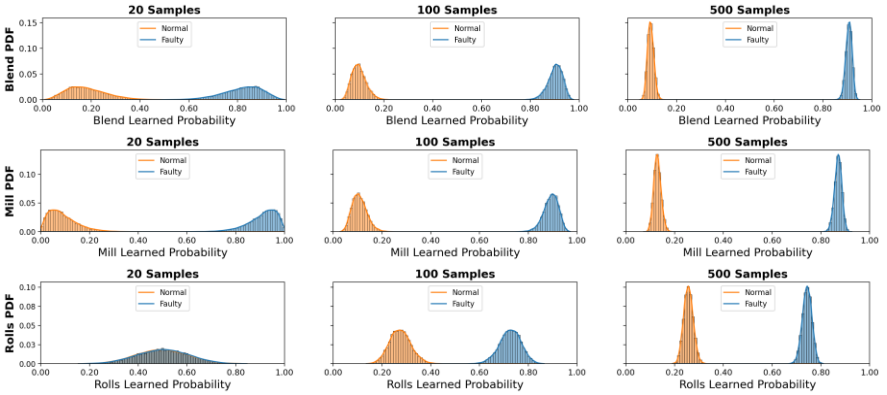


Figure 6. Learned probabilities for the blend, mill, and roll condition at varying sizes of training data.

3.2. Predictive Modeling

One of the main challenges in monitoring the CQA of the RC is the lack of current capability to measure the condition (e.g., flowability and tabletability) of the granules in real-time. Fortunately, the graphical modeling framework allows for the inference of the granule condition given observations from other CM modules. This scenario is shown in Figure 7, with the block arrows indicating message passing from observed variables that are not d-separated(Bishop 2006) from the granule variable. Results of this inference are shown in Figure 8, where based on observations from the surrounding CM modules, the probability of the granule condition changes correspondingly.

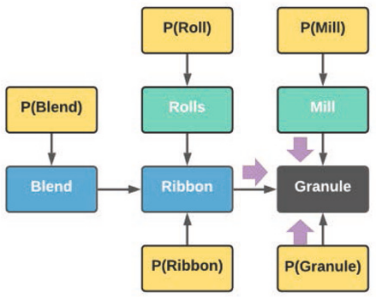


Figure 7. Granule Condition Inference Scheme

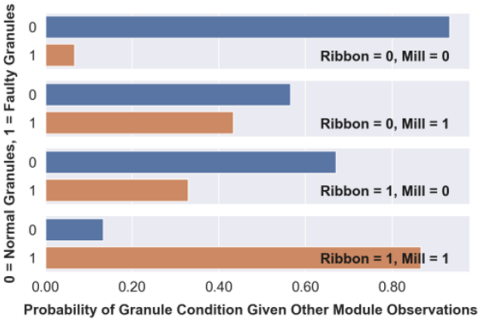


Figure 8. Inference Results on the Granule Condition

Throughout the operation of a continuous processing line, some CM modules might break down, rendering the condition of the actual material and equipment to be unobservable. For the RC example, the NIR sensor observing the ribbon could be undergoing maintenance because of fouling. While this temporary lack of observability could compromise quality assurance of the process, it should not be a reason for

shutting down. Using the message passing scheme shown in Figure 9, the condition of the granule could still be inferred from available CM modules such as the blend, rolls, and the mill condition monitoring modules. The results of these predictions are shown in Figure 10, for varying conditions of the blend, rolls, and the mill.

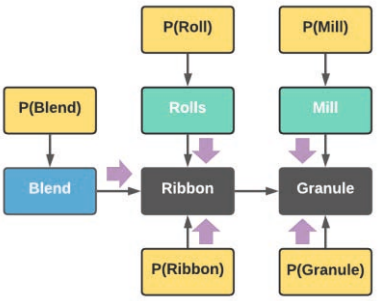


Figure 9. Inference Scheme with Multiple Unobserved Variables.

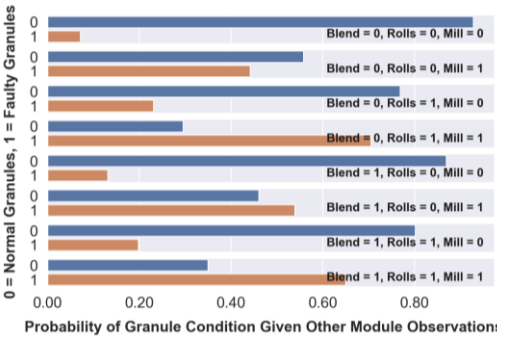


Figure 10. Inference Results on the Granule Condition with Unobserved Ribbon Condition.

4. Conclusions

A comprehensive condition monitoring model for a roller compactor was developed by first considering material and equipment condition variables that are involved, and then establishing logical relationships between them. The condition variables were assumed to be categorical variables with discrete states, and their relationships were encoded into a probabilistic programming framework. This framework was able to efficiently perform approximate inference to learn the parameters of the model, and most importantly, to infer the condition of other less-visible variables like the granule condition based on observations from other condition variables.

5. References

C.M. Bishop, 2006, Pattern Recognition and Machine Learning. Springer 1, 740.

T. Minka, J. Winn, J. Guiver, Y. Zaykov, D. Fabian, J. Bronskill, 2018. Infer.NET 0.3. Microsoft Research Cambridge. http://dotnet.github.io.infer.

A.J. Radcliffe, G. V. Reklaitis, 2021. Process Monitoring under Uncertainty: An Opportunity for Bayesian Multilevel Modelling. Computer Aided Chemical Engineering, 50, 1377-1382.

A. Rogers, M. Ierapetritou, 2014. Challenges and Opportunities in Pharmaceutical Manufacturing Modeling and Optimization. Computer Aided Chemical Engineering, 34,144–149.

S. Roweis, Z. Ghahramani, 1999. A Unifying Review of Linear Gaussian Models. Neural Computation, 11 (2), 305–45.

Q. Su, S. Ganesh, M. Moreno, Y. Bommireddy, M. Gonzalez, G.V. Reklaitis, Z. K. Nagy, 2019. A Perspective on Quality-by-Control (QbC) in Pharmaceutical Continuous Manufacturing. Computers & Chemical Engineering, 125, 216–31.

Z. Webb, J. Romagnoli, 2021. Real-Time Chemical Process Monitoring with UMAP. Computer Aided Chemical Engineering, 50, 2077–82.

J. Winn, C. Bishop, T. Diethe, 2015. Model-Based Machine Learning. Microsoft Research Cambridge. http://www.mbmlbook.com.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50258-X

Framework for Suppressing Transient Fault Alarms Online

Shu Xu[a*], Mark Nixon[a]

[a] *Emerson Automation Solutions, 1100 W Louis Henna Blvd, Round Rock, TX, 78681, USA*

richard.xu@emerson.com

Abstract

While running process monitoring and fault detection software online, it is quite common to face alarms triggered by faults resulting from non-steady state transients. Usually soft sensor models (e.g., principal component analysis (PCA) models) are updated periodically to accommodate such process changes so that similar alarms will not be prompted in the future operation. During online operation, however, alarms might overwhelm operators before model updates and suppressing the alarms online temporarily after checking upon the process or equipment accordingly is deemed a better alternative. In this paper, a framework is presented on how to adjust standard deviations of parameters online to temporarily turn off T^2 and Q alarms in the PCA implementation for transient faults while keep the product's capability to detect other types of faults. Such a framework is tested and validated in a case study using Emerson's *Continuous Data Analytics* (CDA) product.

Keywords: fault detection, alarm suppression, online, transient state

1 Introduction

In modern chemical plant, numerous process variables must be kept within specific limits, and excursions of key variables beyond those limits are often bound to have ramifications in plant safety, the environment, product quality and plant profitability. Process monitoring and fault detection plays an increasingly important role in ensuring that the plant performance meets the operating objectives. Process monitoring takes its root from uni-variate approaches including limit checking, quality control charts and the Six Sigma Approach. However, such methods fail to provide satisfactory results as quality variables increase in number and become highly correlated. To overcome such challenges faced in process monitoring, multivariate methods such as principal component analysis (PCA), project to latent structure (PLS) have been widely used. Recently, uniform manifold approximation and projection (UMAP) from the machine learning community has also gained popularity among industrial practitioners thanks to its superior performance in dimensional reduction (Joswiak et al., 2019; Webb and Romgnoli, 2021). Indeed, several decent review papers on applying state-of-art statistical and machine learning techniques to process monitoring and fault detection have been published (Qin, 2014; Chiang et al., 2017; Ge, 2017; Qin et al., 2021). Yet in this paper we focus on solving problems related to online implementation of PCA in Emerson's *Continuous Data Analytics* (CDA) product used for process monitoring. During online implementation, the life span of PCA models is limited because most processes rarely stay at the same steady state and slow changes in equipment,

feedstock and operating strategy may all compromise model performance. Although PCA models are updated periodically to accommodate process changes, operators are often overwhelmed by alarms generated by state-transient faults during online operation. In addition, simply acknowledging alarms associated with control limit violation is not desirable as the product might lose the capability to detect other faults that violate the control limits as well. After checking upon the process or equipment indicated by alarms, the operators request for tools to suppress them online temporarily before next model update. Industrial practitioners from Dow Chemical shared their experiences on online soft sensor maintenance where robust mean and variance estimators of the inputs and outputs are used (Chiang et al., 2017; Lu and Chiang, 2018) to overcome the long-term model degradation problem. Inspired by such work, a new framework is developed which automatically adjusts standard deviations of parameters online to temporarily turn off T^2 and Q alarms associated with transient faults and keeps the product's capability to detect other types of faults.

The remainder of this paper is organized as follows. Section 2 describes the proposed framework for suppressing transient fault alarms online, and it is validated by case study shown in Section 3. Section 4 draws conclusions based on results obtained in the study.

2 Algorithm for suppressing transient fault alarms online

Assuming each alarm suppressing step leads to the values of T^2 and Q falling around 0.8 of their control limits respectively, standard deviation of the ith parameter (σ_i) has to be adjusted to achieve such a goal.

2.1 T^2 alarms

If a T^2 alarm occurs at time t in a new steady state, since the standard deviation array of all parameters ($\boldsymbol{\sigma_{a,0}}$) is available, by incorporating the contribution of T^2, the updated standard deviation ($\boldsymbol{\sigma_{b,0}}$) is provided by:

$$\sigma_{b,j,0} = \left[\frac{\left|contT^2_{t,j}\right|}{\max_j \left\{\left|contT^2_{t,j}\right|\right\}} \left(\frac{T^2_t}{0.8T^2_{UCL}} - 1 \right) + 1 \right] \sigma_{a,j,0} \tag{1}$$

where T^2_{UCL} is the control limit of *Hotelling* T^2, and $\left|contT^2_{t,j}\right|$ is the contribution of jth parameter. As shown in Eq. (1), if the jth parameter makes no contribution to the T^2, i.e., $\left|contT^2_{t,j}\right| = 0$, the corresponding jth standard deviation will not be changed. After obtaining $\boldsymbol{\sigma_{a,0}}$ and $\boldsymbol{\sigma_{b,0}}$, an interactive strategy based on the bisection method is used to obtain the standard deviation array, $\boldsymbol{\sigma_c}$, which leads to updated T^2 values close to the target of $0.8T^2_{UCL}$.

2.2 Q alarms

After suppressing the T^2 alarms, there still might be Q alarms that require operators' attention. Although the Q values might have already decreased with increased standard deviations $\boldsymbol{\sigma_c}$ obtained in the T^2 alarm suppressing step, they need to be further reduced to quell the remaining Q alarms. An approach similar to the strategy used in T^2 alarm management is applied, shown as follows.

Assuming the standard deviation array of all parameters from the previous T^2 alarm suppressing step ($\boldsymbol{\sigma_{c,0}}$) is available, by incorporating the contribution of Q, the updated standard deviation ($\boldsymbol{\sigma_{d,0}}$) is provided by:

$$\sigma_{d,j,0} = \left[\frac{|contQ_{t,j}|}{\max_j \{|contQ_{t,j}|\}} \left(\frac{Q_t}{0.8Q_{UCL}} - 1 \right) + 1 \right] \sigma_{c,j,0} \tag{2}$$

where Q_{UCL} is the control limit of Q, and $|contQ_{t,j}|$ is the contribution of jth parameter.

Similar to Section 2.1, after obtaining $\boldsymbol{\sigma_{c,0}}$ and $\boldsymbol{\sigma_{d,0}}$, an iterative strategy based on the bisection method is used to obtain the standard deviation array, $\boldsymbol{\sigma_e}$, which leads to updated Q values close to the target of $0.8Q_{UCL}$.

3 Case study

In this section, the continuous process of mixing two ingredients (A and B) is studied, as shown in Figure 1. A process fault caused by changes in *Main Flow A* occurred during online operation, as indicated by soaring T^2 and Q values in Figure 2 which violate the control limits and trigger alarms. It is worth noting that in Emerson's product, T^2 and Q are normalized by their control limits T^2_{UCL} and Q_{UCL} respectively so that they share an identical normalized control limit of 1.

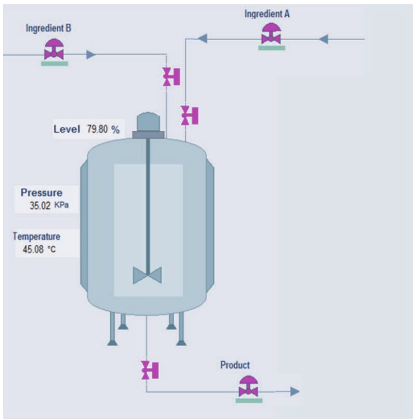


Figure 1: Mixing process of two ingredients

Variables with top contributions to the T^2 fault, together with associated changes in standard deviations are summarized in Table 1. Comparing the T^2-adjusted standard deviations $\boldsymbol{\sigma_c}$ with the original $\boldsymbol{\sigma_a}$, it is observed that variables with large contributions show significant increases in their standard deviations, such as "Main Flow A "and "Main Flw Valve ".

After suppressing the T^2 alarm, results are obtained (see Figure 3) with T^2 values fall below the control limit. Q values have also been drastically reduced from 68.8 to 12.9, yet they are still above the control limit.

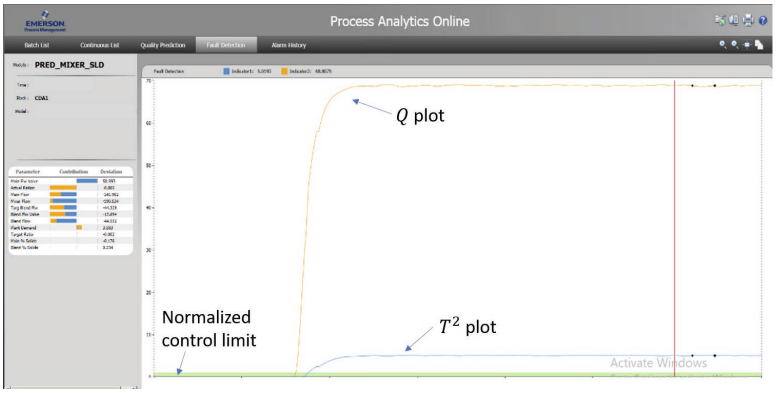


Figure 2: Process analytics online GUI with both T^2 and Q alarms

Table 1: Variables with top T^2 fault contribution and associated standard deviation changes

Variable Tag	T^2 contribution	σ_a	σ_c
Main Flow A	-18.74	45.08	81.74
Main Flw Valve	37.78	7.494	19.78
Targ Blend Flw	-14.89	21.95	36.13
Mixer Flow	-28.25	59.63	132.7
Blend Flow	-24.27	21.65	44.45

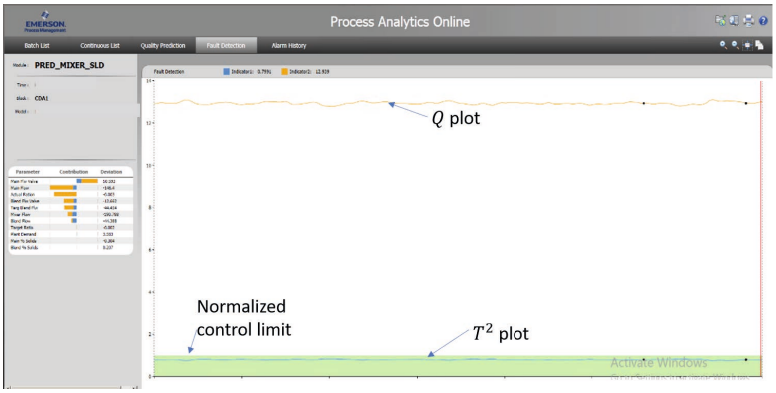


Figure 3: Process analytics online GUI with Q alarm only

In this case, further steps are taken to suppress the Q alarm. Variables with top contributions to the updated Q fault, together with updated standard deviations are summarized in Table 2. Similar to the results shown before, variables with the lion share of contribution enjoy drastic rises in standard deviations ($\boldsymbol{\sigma_c} \rightarrow \boldsymbol{\sigma_e}$), such as "Main Flow A "and "Main Flw Valve ". The resulting Q value drop (below the control limit) is demonstrate in Figure 4.

Table 2: Variables with top Q fault contribution and associated standard deviation changes

Variable Tag	Q contribution	$\boldsymbol{\sigma_c}$	$\boldsymbol{\sigma_e}$
Main Flow A	-1.602	81.74	382.6
Main Flw Valve	6.555	19.78	317.7
Actual Ration	-1.424	0.0622	0.2655
Targ Blend Flw	-0.596	36.13	85.60
Blend Flw Valve	-0.643	9.859	24.42

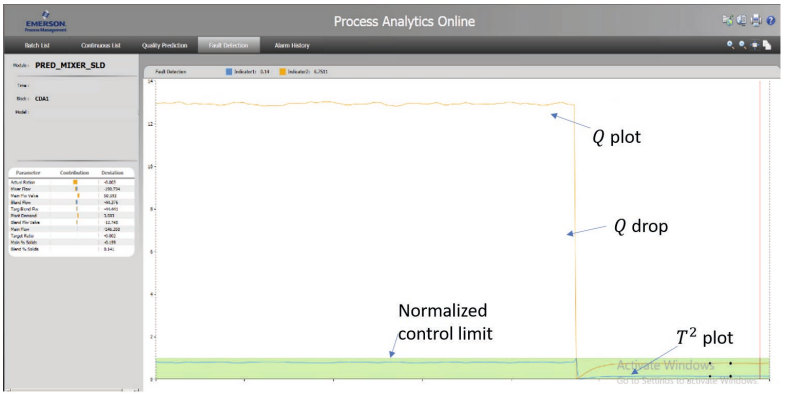


Figure 4: Process analytics online GUI with both T^2 and Q alarms suppressed

After suppressing the alarms caused by *Main Flow A* fault, the system is still able to sound alarms if a new fault occurs. As shown in Figure 5, later on the process is suffering a load disturbance where "Blend% Solid"decreases from 65 to 50 (the black line in the bottom figure), which leads to a quality deviation indicated by a jump of T^2.

4 Conclusions

In this paper, a new framework to suppress transient fault alarms online is proposed, and it is developed for Emerson's DeltaV *Continuous Data Analytics* product where PCA-based process monitoring approach is implemented. By adjusting standard deviations of parameters online based on variable contribution to faults, both T^2 and Q alarms can be turned off. Such an approach keeps the product's capability to detect other types of faults, as shown by the promising results in an industrial case study. It is worth pointing out that the new method only acts as a temporary solution and it works better when new steady

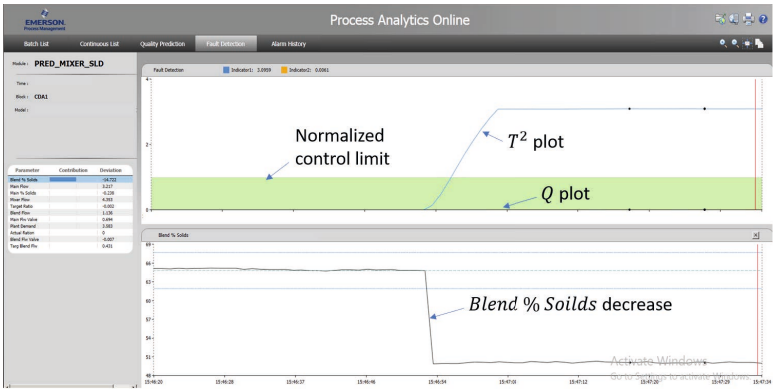


Figure 5: Process analytics online GUI with a new T^2 alarm

states are reached before the next model update. To further improve PCA monitoring performance, future work will be focused on incorporating filtering techniques to remove outliers.

References

L.H. Chiang, E.L. Russell, R.D. Braatz, 2000, Fault detection and diagnosis in industrial systems, Springer-Verlag London Ltd.

L.H. Chiang, L.F. Colegrove, 2007, Industrial implementation of on-line multivariate quality control, Chemometr. Intell. Lab. Syst., 88, 143-153

L.H. Chiang, B. Lu, I. Castillo, 2017, Big Data analytics in chemical engineering, Annu. Rev. Chem. Biomol. Eng., 8(1), 63-85

Z.Q. Ge, 2017, Review on data-driven modeling and monitoring for plant-wide industrial processes, Comput. Chem. Eng., 171, 16-25

M. Joswiak, Y. Peng, I. Castillo, L.H. Chiang, 2019, Dimensionality reduction for visualizing industrial chemical process data, Control Eng. Pract., 93, 104189

B. Lu, L.H. Chiang, 2018, Semi-supervised online soft sensor maintenance experiences in the chemical industry, J. Process Control, 67, 23-34

S.J. Qin, 2014, Process data analytics in the era of Big Data, AIChE J., 60(9), 3092-3100

S.J. Qin, L.H. Chiang, 2019, Advances and opportunities in machine learning for process data analytics,Comput. Chem. Eng., 126, 465-473

S.J. Qin, S. Guo, Z. Li, L.H. Chiang, I. Castillo, B. Braun, Z. Wang, 2021, Integration of process knowledge and statistical learning for the Dow data challenge problem, Comput. Chem. Eng., 153, 107451

Z. Webb, J. Romagnoli, 2021, Real-time chemical process monitoring with UMAP, Comput. Aided Chem. Eng., 50, 2077-2082

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50259-1

Using Reinforcement Learning in a Game-like Setup for Automated Process Synthesis without Prior Process Knowledge

Quirin Göttl[a], Dominik G. Grimm[b,c,d] and Jakob Burger[a,*]

[a] *Technical University of Munich, Campus Straubing for Biotechnology and Sustainability, Laboratory of Chemical Process Engineering, Uferstraße 53, 94315 Straubing, Germany.*
[b] *Technical University of Munich, Campus Straubing for Biotechnology and Sustainability, Bioinformatics, Schulgasse 22, 94315 Straubing, Germany.*
[c] *Weihenstephan-Triesdorf University of Applied Sciences, Petersgasse 18, 94315 Straubing, Germany.*
[d] *Technical University of Munich, Department of Informatics, Boltzmannstr. 3, 85748 Garching, Germany.*
burger@tum.de

Abstract

The present work uses reinforcement learning (RL) for automated flowsheet synthesis. The task of synthesizing a flowsheet is reformulated into a two-player game, in which an agent learns by self-play without prior knowledge. The hierarchical RL scheme developed in our previous work (Göttl et al., 2021b) is coupled with an improved training process. The training process is analyzed in detail using the synthesis of ethyl tert-butyl ether (ETBE) as an example. This analysis uncovers how the agent's evolution is driven by the two-player setup.

Keywords: Automated Process Synthesis; Flowsheet Synthesis; Artificial Intelligence; Machine Learning; Reinforcement Learning.

1. Introduction

RL is a frequently used machine learning approach in the process engineering community. Besides many applications in process control, it is also employed for problems that require forward planning. For example, Wang et al. (2020) designed synthetic pathways for organic chemistry with an RL approach combined with Monte-Carlo tree search. Khan and Lapkin (2020) showed that RL can identify processing routes for hydrogen production.

We presented an approach called SynGameZero, which enables training an agent without prior knowledge to synthesize entire process flowsheets using RL (Göttl et al., 2021a). Thereby, flowsheet synthesis is transformed into a turn-based two-player game. Both players start with an empty flowsheet. In their turns, they add unit operations or recycles to their flowsheet while also seeing what their opponent does. After each turn, a flowsheet simulator generates stream tables and provides them to the players as a base for their next moves. The winner is determined by the net present value, calculated after flowsheet completion. In a tied game, the player who finishes the flowsheet first is the winner. The reward $r \in \{1, -1\}$ is a binary value indicating win or loss. The agent consists of an artificial neural network (ANN) combined with a tree search and is trained by playing

this game against itself. The ANN's output contains a suggestion for the next move of the current player. It also includes an estimate of the chances of the current player to win the game. This output guides the tree search, which considers several moves in advance, imitating a typical human planning process. The search is adaptive regarding depth and explores only promising options selected based on the ANN's output. The tree consists of nodes (representing flowsheet states) connected by branches (possible moves of adding units or recycles). No prior knowledge is required to initialize the agent. During the training process, the parameters of the ANN are optimized to improve its ability to suggest good moves and estimating the chances to win the game. Recently, we have improved the two-player framework by structuring the agent's decision in three hierarchy levels and introducing hierarchical RL (Göttl et al., 2021b). Recasting the problem into this game framework enables powerful agent structures and training methods from the literature outside chemical engineering. Many authors proved that RL serves as a powerful tool to master complex problems like winning the board games of Go and Chess (Silver et al., 2018).

The present work builds upon the hierarchical RL scheme as presented in Göttl et al. (2021b) and modifies the agent's structure and the training process resulting in a slightly improved performance. For the first time, the agent's evolution during training is analyzed in detail. The analysis uncovers the importance of the two-player game setup for the success of the SynGameZero method.

2. Methodology

2.1. Example Process and Flowsheet simulation

For comparability, we adopt the process design problem of ETBE synthesis and the flowsheet simulator from Göttl et al. (2021b). Here, we give a brief summary; the reader is referred to the original paper for details. A quaternary system consisting of ethanol (EtOH), isobutene (IB), n-butane (nBut), and ETBE is considered. Two feed streams are sampled randomly at the start of the game (the first one containing EtOH and the second one a mixture of IB and nBut). The goal of the two-player game is to create the flowsheet that maximizes the net present value of the process. Thereby, the following idealized unit operations may be used:

Reactor (R): The following reversible reaction occurs in the reactor, which always reaches equilibrium.

$$\mathrm{EtOH} + \mathrm{IB} \leftrightarrow \mathrm{ETBE}. \tag{1}$$

Distillation columns (D_L and D_H): Distillation columns are assumed with infinite height and total reflux for faster simulation. Due to binary azeotropes, a distillation boundary separates the quaternary system into two distillation regions. It is possible to choose between two different product splits. The split D_L obtains the light-boiler of the feed's distillation region with the highest possible yield, and the split D_H obtains the respective heavy-boiler with the highest possible yield.

Mixer / Recycle (M): any open stream can be admixed to another open stream or an already used stream. The latter results to a recycle in the process.

2.2. Agent structure and training procedure

The present work utilizes a similar framework as described by Göttl et al. (2021b), and we refer again for a detailed description. The agent's actions are structured hierarchically

into three levels. At Level 1, the agent chooses an open stream in the flowsheet or the termination of the synthesis process. If an open stream was chosen at Level 1, the agent selects a unit operation from the above list as the destination of the open stream at Level 2. If the agent decides for Mixer / Recycle (M), then Level 3 is used. There, the agents select another already existing stream as the destination to admix/recycle the open stream chosen at Level 1. In contrast to Göttl et al. (2021b), the agent does no longer differentiate between a mixer and a recycle. They are conceptually the same. The flowsheet simulator determines automatically if the chosen action leads to a mixer or a recycle and simulates the respective option.

The ANN is structured according to the three hierarchy levels and explained using Figure 1 (for the specifications and the setup of the ANN, we refer to GitHub: https://github.com/grimmlab/SynGameZero/tree/ETBE_synthesis2.0). Its first part is a convolutional block, which processes the stream tables of both players. Each hierarchy level is represented by an actor-critic network, which receives data processed by the previous networks. At Level 2, information on the open stream chosen at Level 1 is also provided to the actor-critic network. Each actor-critic network i generates two kinds of outputs $\boldsymbol{\pi}_i$ and v_i. $\boldsymbol{\pi}_i$ represents a probability distribution, which is a suggestion for the decision at this level, while v_i is an estimate of the expected reward for the current player. Those outputs are used to guide the tree search by exploring promising flowsheet alternatives (promising is quantified by $\boldsymbol{\pi}_i$ and v_i). The agent's actions in the game are determined based on the results of the tree search.

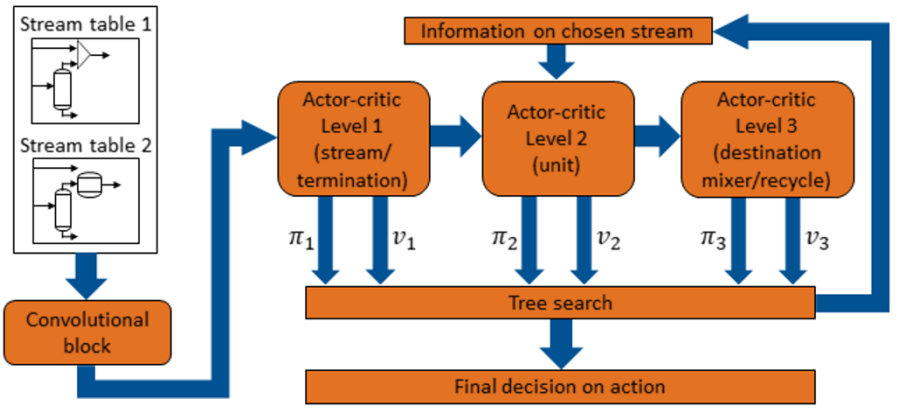


Figure 1. The agent's hierarchical structure. The arrows indicate flow of information. Artificial neural networks with trainable parameters are shown in rounded boxes.

During the training process, the agent plays $N_{\text{steps}} = 10{,}000$ games against itself with randomly sampled feed streams. In previous work (Göttl et al., 2021b), we did not consider feed streams with equimolar rates of EtOH and IB, although those configurations are particularly interesting. In the present work, equimolar feed streams are included in 20 % of all training games to enable the agent to deal with these configurations. After each game, the states (stream tables of both players), the results of the tree search (the basis for the decisions on actions), and the reward $r \in \{1, -1\}$ for each player are stored in a memory with size $N_{\text{memory}} = 160$. A batch of size $N_{\text{batch}} = 64$ is sampled out of the memory to perform a stochastic gradient descent step on the ANN with a learning rate of $\beta = 0.0001$.

2.3. Evaluation

After the training procedure, the agent is evaluated by playing 1,000 games against itself with randomly sampled feed streams. Due to combinatorics, it is not possible to store the optimal flowsheet for every conceivable feed stream combination. Therefore, the agent is evaluated by comparing its flowsheets with three benchmark flowsheets created by humans (Göttl et al., 2021b). They include the industrial Oxeno-process for ETBE synthesis (Ryll et al., 2014). The following metrics are used to quantify the performance. R_1 is the fraction of games where the agent proposed a flowsheet at least as good as the best benchmark. R_2 is the average relative gain in the net present value of the agent over the best benchmark. For a precise mathematical definition of the metrics, the reader is referred to (Göttl et al., 2021b).

To uncover the importance of the two-player game setup for the success of the SynGameZero method, the agent state (i.e., the ANN's parameters) is saved at various stages during the training process. The behavior of these agent versions is studied by letting it play the two-player game against itself for a fixed set of feed stream combinations: the molar flow rate of EtOH is varied between 15 and 95 kmol/hr. The molar flowrates of IB and nBut are set equal. They are also varied between 15 and 95 kmol/hr.

3. Results

The training procedure is repeated five times. Table 1 shows the average values for the performance metrics. R_1 shows that the agent can outperform the benchmark flowsheets in almost every case. According to R_2, the benchmarks are surpassed on average by 24.4 %. Compared to the results shown in previous work, the agent shows improved performance.

Table 1. Resulting performance metrics.

	R_1	R_2
Present work	0.9996	0.2438
Göttl et al. (2021b)	0.9864	0.2279

Figure 2 illustrates the evolution of the agent during training by showing its behavior after different numbers of training steps. The various feed stream combinations are depicted as cells in the matrix. The winning player of every combination is indicated with a color code. For one feed stream combination, the winning flowsheet (marked red) is displayed.

Without any training ($N_{\text{steps}} = 0$), the agent consists of a randomly initialized ANN and the tree search. The behavior of the agent is the same for all shown combinations of feed streams. In the role of Player 1 it terminates the synthesis right away. In the role of Player 2, it sets up the shown flowsheet and wins the game. After 100 training steps, Player 1 has copied this tactic (for all shown feeds) and wins all games (if both players do the same, the game is tied, and Player 1 wins because he/she finishes first).

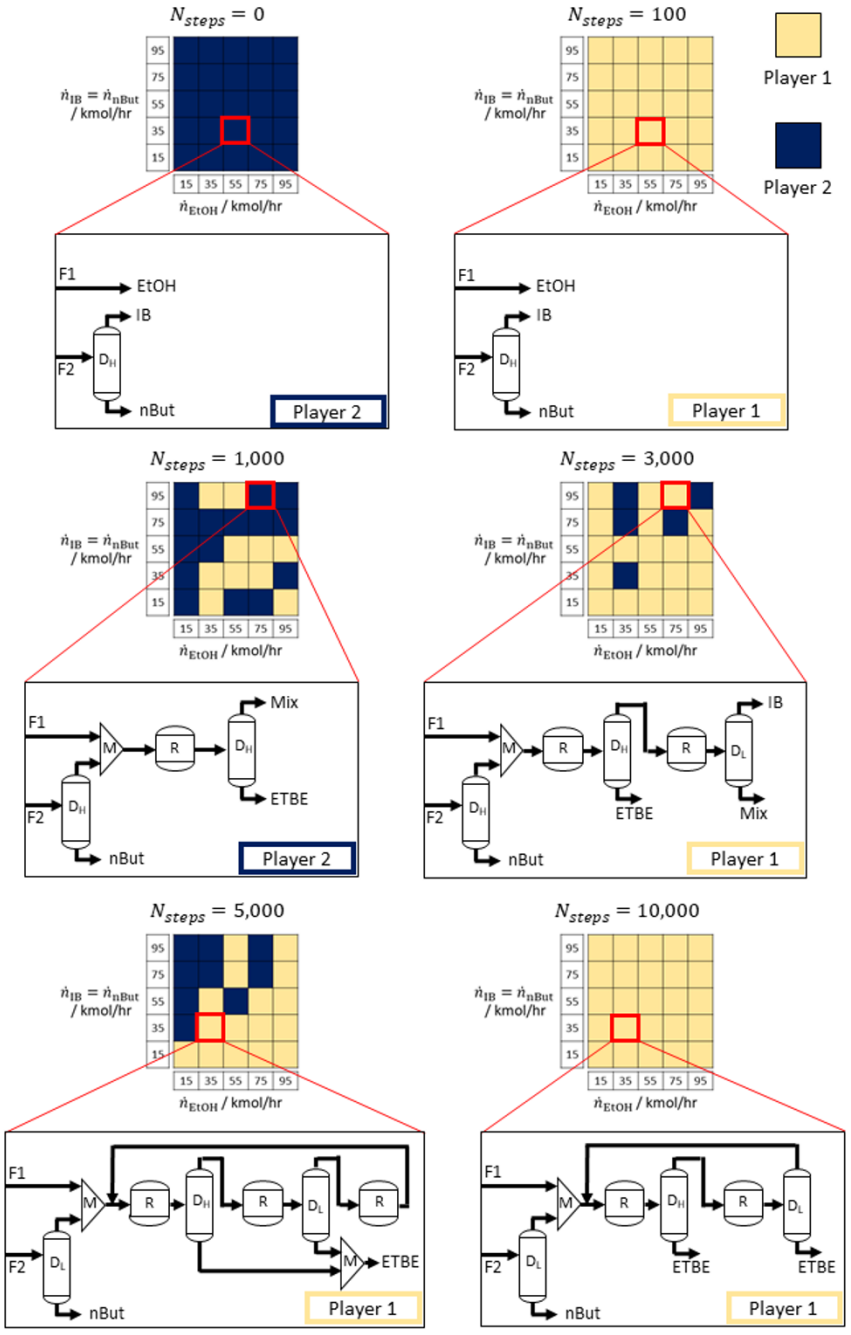


Figure 2. Illustration of the agent's evolution after different numbers N_{steps} of training steps. The matrix field represent different feed stream combinations. The color code marks the winning player. The red box shows the winning flowsheet for the respective feed streams.

After 1,000 training steps, the game is more balanced. Both players can win some situations. From the shown flowsheet generated by Player 2, it is visible that the agent has learned to use a reactor to synthesize ETBE, clear progress. After 3,000 training steps, Player 1 is more dominant, and the flowsheets become more sophisticated. This balance change between Players 1 and 2 winning in the game is observed many times during training. Typically, Player 1 is copying (if needed) and using the so-far best-known tactic. Player 2 has to avoid a tie and therefore forced to explore alternative tactics. It is consequently mainly Player 2 who uncovers novel improved tactics. Player 2 will afterward win more games than Player 1 for a short period during training. Eventually, Player 1 acquires the novel tactic and wins again. The bottom row in Figure 2 shows situations at the late stages of the training. After 5,000 steps, the complexity of the flowsheets further increases, while the game is still quite balanced. For equimolar feed rates of IB and EtOH (i.e., on the diagonal of the matrix), the agent has learned to generate flowsheets with complete conversion of IB and EtOH. The chemical equilibrium is overcome by using a recycle (cf. the shown flowsheet). However, the design is still not optimal. After 10,000 steps, the flowsheets are slightly more improved, and the training is completed. Player 1 wins all games. Even with further training, Player 2 is unable to find a better tactic. Such a constellation signifies that a local or maybe even global optimum for the performance has been reached.

4. Conclusions

The SynGameZero approach, which enables an agent via RL to synthesize flowsheets, is slightly improved and demonstrated for an example process with incomplete conversion in the reactor and recycles. Efficient and effective training is achieved by a hierarchical agent structure and a two-player game setup. The latter forces one player to explore novel tactics. The other player adopts them as soon as they are advantageous.

Funding: funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project-ID 466387255 - SPP 2331.

References

Q. Göttl, D. G. Grimm, J. Burger, 2021a, Automated synthesis of steady-state continuous processes using reinforcement learning, Front. Chem. Sci. Eng., DOI: 10.1007/s11705-021-2055-9

Q. Göttl, Y. Tönges, D. G. Grimm, J. Burger, 2021b, Automated Flowsheet Synthesis Using Hierarchical Reinforcement Learning: Proof of Concept, Chem. Ing. Tech., 93, 12

A. Khan, A. Lapkin, 2020, Searching for optimal process routes: A reinforcement learning approach, Comput. Chem. Eng., 141, 107027

O. Ryll, S. Blagov, H. Hasse, 2014, Thermodynamic analysis of reaction-distillation processes based on piecewise linear models, Chem. Eng. Sci., 109, 284-295

D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, T. Lillicrap, K. Simonyan, D. Hassabis, 2018, A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play, Science, 362, 1140-1144

X. Wang, Y. Qian, H. Gao, C. W. Coley, Y. Mo, R. Barzilayb, K. F. Jensen, 2020, Towards efficient discovery of green synthetic pathways with Monte Carlo tree search and reinforcement learning, Chem. Sci., 11, 10959-10972

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50260-8

Generation and Benefit of Surrogate Models for Blackbox Chemical Flowsheet Optimization

Tim Janus[a*], Felix Riedl[a], Sebastian Engell[a]

[a]*TU Dortmund University, Emil-Figge-Straße 70, Dortmund 44137, Germany*
tim.janus@tu-dortmund.de

Abstract

Commercial process simulators are widely used in process design, due to their extensive library of models and ease of use. The results obtained from these simulators can be used for global flowsheet optimization but often gradient information is not provided so that derivative-free optimization methods must be used. The process simulator is called as a black box and this is computationally expensive, thus filtering out simulations that have a low probability of providing good results by machine learning is attractive to increase the efficiency of derivative-free optimization. The surrogate models used for filtering are initially based upon small data sets. We explore the generation of these initial data sets and we investigate two alternatives and suggest a heuristic for the choice of the decision function for inequality constraints.
Keywords: Aspen Plus, Process Optimization, Surrogate Models, Machine Learning, Evolutionary Algorithms

1. Introduction

Chemical process design usually is performed as an iterative process of comparing the performance and the costs of alternative process configurations and parameterizations by an interdisciplinary team of experts. In many cases, such design studies for chemical processes are performed interactively using commercial block-oriented flowsheet simulators e.g. Aspen Plus. Asprion et al. (2018) give an overview on process simulators. While this can lead to satisfactory solutions, optimal designs cannot be expected. On the other hand, the rigorous optimization of a chemical process is a very challenging task as the models that represent the process units are generally nonlinear and the problems are large and often nonconvex. Further, discrete decisions as e.g. the number of stages and the feed stage of a distillation column enlarge the complexity of the optimization problem. In research, usually mixed-integer non-linear program (MINLP) solvers and equation-based models are used. This approach has the disadvantage that specific model formulations and expert knowledge to set up the models are needed. In contrast, commercial process simulators offer a large model library, and different design alternatives can be simulated easily. It is therefore promising to combine the ease of modelling provided by commercial flowsheet simulators with the power of optimization. But the lack of interfaces to internal information like sensitivities is a big obstacle. An approach to combine the use of commercial simulators with the power of optimization is to use derivative-free methods that are not as efficient as the derivative-based optimization of equation-based models with mathematical programming. In our previous work, we have proposed to use evolutionary algorithms together with flowsheet simulators for optimization-based process design, for details see Urselmann et al. (2016). This however requires a large number of calls to the simulator of which a significant fraction do not converge or result

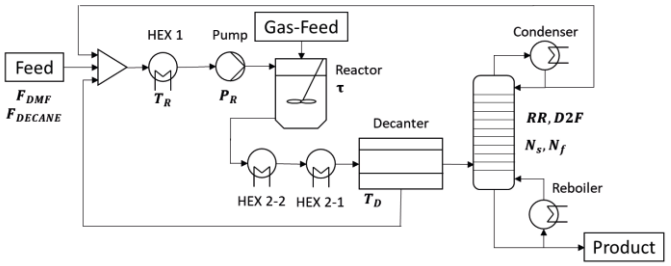


Figure 1 - Flowsheet of the case study of the hydroformylation of 1-dodecene in a thermomorphic solvent system with three major unit operations and two recycle streams. The degrees of freedom are highlighted in bold font.

in large violations of the design constraints. Due to the relatively long computation times for the single simulations, there is a strong interest to filter the design proposals and to steer the optimization to promising regions such that computation time with non-converging solutions and solutions that are far from the design specifications is not wasted. Surrogate-assisted optimization is a possible approach to enhance the efficiency of derivative-free blackbox optimization. It applies methods from Machine Learning (ML) to steer the optimization and to avoid the exploration of non-promising parts of the solution space, see Haftka et al. (2016) for a review. In Janus et al. (2021) it was shown that the required time for global flowsheet optimization could be halved by such methods. In this contribution, we analyze the use of surrogate models to decide on the execution of simulations in more detail, specifically the initial generation of the data for the training of the surrogate models and the choice of the decision rules.

2. Case Study

For the numerical studies, we use the case study of the homogenously catalyzed hydroformylation of 1-dodecene to n-tridecanal in a thermomorphic solvent system (TMS). The TMS process considered here has been investigated extensively experimentally in the Collaborative Research Center DFG Transregio SFB 63 ”Integrated chemical processes in liquid multi-phase systems InPROMPT”. The process is performed in a mixture of the solvents dimethylformamide (DMF) and decane in order to recover the expensive homogeneous rhodium catalyst from the product stream. The phase behavior of the mixture of DMF and decane is temperature-dependent and a change of the temperature is used to switch between a homogenous mixture in the reactor and a mixture with two liquid phases in the decanter, see Kiedorf et al. (2014). Figure 1 illustrates the flowsheet. The feed stream is heated and pressurized to create a homogenous mixture in the reactor and H_2 and CO are added as syngas. Beside the main reaction of 1-dodecene to n-tridecanal, four side reactions occur in the reactor, see Merchan et al. (2016). A cascade of two heat exchangers cools down the mixture. Then a liquid-liquid separator splits the two phases of the mixture. The polar DMF-rich phase contains the catalyst and is recycled back into the reactor. The decane-rich phase that contains the unconverted feed, the product and the byproducts is fed into the distillation column for purification. The top stream of the column is recycled back to the reactor and a product-rich liquid stream is obtained at the bottom of the column. The bottom stream must have a product purity of more than 99 mol %. The ten degrees of freedom (DOFs) of the process design considered here are the number of theoretical stages, the feed stage, the distillate to feed ratio and the reflux ratio of the column, the temperature, pressure and residence time of the reactor, the temperature of the decanter and the flow

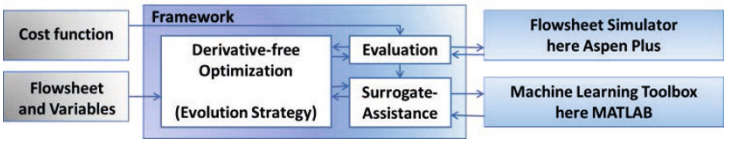


Figure 2 - Diagram of the flowsheet optimization framework that consists of three parts. The input for the framework, the optimization framework with the three modules: optimization, evaluation and surrogate-assistance, and the third party software. Here, Aspen Plus is used as a commercial flowsheet simulator and MATLAB is used as a machine learning toolbox.

rates of DMF and decane. The operating window is between 80 and 120 °C for the reactor and -5 to 25 °C for the decanter. The operating window of the reactor pressure is between 15 and 30 bar, the other vessels are operated at atmospheric pressure.

3. Framework for Flowsheet Optimization

Figure 2 shows the flowsheet optimization framework. The global derivative-free optimization is implemented by an evolution-strategy (ES) as described by Beyer and Schwefel (2002). An initial flowsheet, the degrees of freedom, and their bounds must be provided to the framework to start the design optimization. The framework calls the process simulator by setting the values of the degrees of freedom, and then uses the results, in particular the critical purities and the flow rates as well as equipment and operational parameters. If the simulation does converge, the framework calculates the cost function and stores the result and the internal process information. If a simulation does not converge, no internal process information, e.g. concentrations, is collected. The surrogate-assistance module in Figure 2 contains the candidate-rejection and the candidate-generation heuristics that guide the search of the optimizer. It generates and evaluates surrogate-models that estimate results of the flowsheet simulation, e.g. the product purity. The evaluation of a surrogate-model is several orders of magnitudes

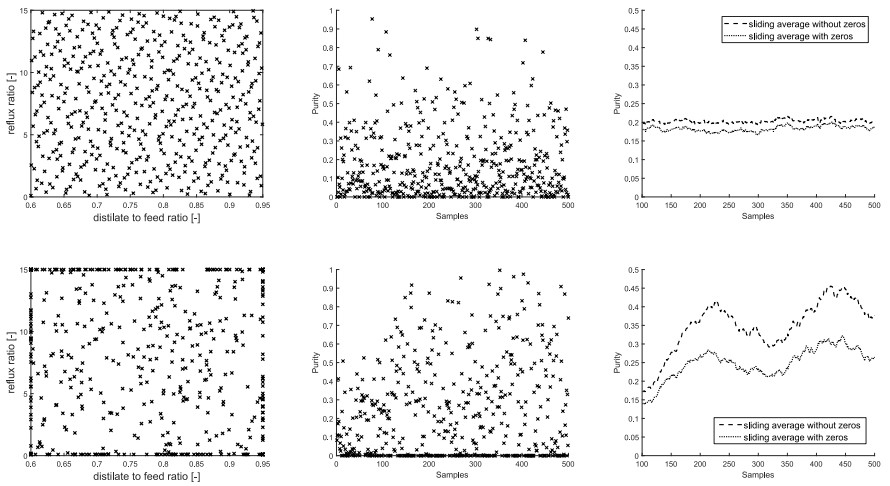


Figure 3 - Distribution plots. The first row contains plots with samples obtained from a ‘Halton set’ and the second row contains plots with samples from simulations that were triggered by the evolution strategy. The first column shows the distribution of the degrees of freedom 'distillate to feed ratio’ and ‘reflux ratio’, the second column shows the purity distribution of the samples, and the third column shows a sliding average of the purity based on the last 100 samples.

faster than a flowsheet simulation. The generation of the surrogate-models is performed by an optimization that minimizes the difference between the estimates of the surrogate-models and the result of the simulator. If an initial data set is provided, the execution of the framework starts with the generation of data for the training of surrogate-models. After each evaluation of the process model, the data is stored and after a certain number of simulations, the machine learning toolbox (MLT) retrains the surrogate-models. The MLT supports multi-shot training and hyperparameter optimization.

4. Evaluation and generation of surrogate-models

ML methods generate surrogate-models based upon training data. We employ neural networks as surrogate-models. Beside the guidance of the optimization, the surrogate-assistance controls the iterative generation and evaluation of new surrogate-models.
In our approach, two surrogate-models are used. A classifier predicts if a simulation shall not be started because the process simulator will most likely not converge, and a regression model estimates the satisfaction of crucial constraints, in this example the product purity. The candidate rejection rule discards a simulation if either the classifier predicts non-convergence or the estimated product purity is equal to or below a threshold. For this, we use a dynamic boundary b that depends on the purity constraint c_p as a heuristic, see Eq. (1). The symbol mae here denotes the mean absolute error of the neural network and the index $feas$ refers to the feasible part of the training set that consists of the data that has been generated by the optimization before the training. Y_i and P_i represent the simulated and the estimated purity for data point i. If Eq (1) is not applicable, i.e. the number of feasible points in the test set N_{feas} is less than two, the boundary b is defined as $b = c_p - 3 \cdot mae_{all}$ whereby mae_{all} refers to the mean absolute error of the training set on the neural network.

$$b = c_p - \left(mae_{feas} + 1.96 \cdot \left(\sqrt{\frac{1}{N_{feas}} \sum \left(|Y_i - P_i| - mae_{feas} \right)^2} \right) \right) \tag{1}$$

Beside the configuration of the neural network, the distribution of the training data has a strong influence on the performance of the generated surrogate-models, e.g. it should be beneficial to have a high density of samples in the region of the purity constraint. A Halton set with RR2 scrambling and data from previous simulations that were triggered by the ES were compared for the generation of the initial samples. Figure 3 shows the

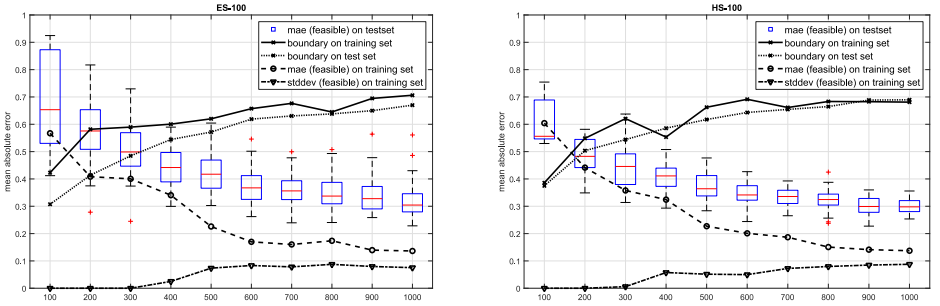


Figure 4 - Development of the mean absolute error of the neural networks that predict the product purity. On the left side, the first 100 samples origin from previous optimizations and on the right side the first 100 samples were generated as a Halton set. The test set consists of 12,000 process alternatives, 49 of them have a purity >= 99 %. The x-axis indicates the number of samples in the training set.

distribution of these two data sets for 500 samples. With samples from the Halton set, the space of the degrees of freedom is uniformly covered within the bounds. In the initial explorative phase, the evolution strategy tends to explore specifically the boundaries of the search space. After 100 samples, the average purity of the simulated process alternatives is around 20 %, but from there on the evolution strategy starts to find process alternatives with a higher purity. This is beneficial for the optimization as the goal of the surrogate-model is to predict high purities with a high accuracy, and therefore data points with higher purities should be part of the training data set.
When the surrogate models are updated, the training set contains the initial data (i.e. either the Halton set or the initial simulations generated by the ES) and all process alternatives that have been simulated during the optimization up to the start of the training procedure. The surrogate-assistance optimizes the neural networks on the entire training set, i.e. the training set equals the test set. As the heuristic boundary could be unreliable during the early stages of the optimization, we validate it with a large test set.

5. Results

We compare the use of an initial Halton set and the use of data from the initial simulations as training data for the surrogate models in the initial stage in Figure 4. Only candidate rejection was applied but not heuristic generation of promising candidates as in Janus et al. (2021). Ten runs were performed to reduce random effects. The training of the neural networks was repeated five times. The neural network architecture was chosen from previous studies as networks with two hidden layers with 45 nodes each. The neural networks were retrained every 100 simulations. When the training data consists of up to 1000 samples, the training requires approx. 20 s, while a single process simulation requires approx. 6 s, so the additional effort for the training of the surrogate models pays off quickly when later unpromising simulations are saved.
For surrogate models trained with 100 samples, the Halton set leads to more accurate models with a smaller variance than the data sequence from the calls of the optimization strategy, but this difference becomes smaller when more samples are available. Figure 4 shows the development of the mean absolute error (mae) of the networks. The x-axis shows the samples of the training set, where after the first training, the evolution strategy generated the training data. The box plots show the mae on the large test set for 50 trainings, i.e. ten optimizations with five repeats of the training. The solid line shows the values of the boundaries that were used in the purity-based candidate-rejection during the optimization which are based on the data seen so far. The dotted line shows the boundaries that were calculated based on the posteriori test set. Both lines are close to each other and have a similar behavior, which validates the proposed heuristic. The first dashed line shows the mae on the feasible part of the training set or three times the entire mae on the training set if the training data did not contain at least two feasible process alternatives. For the first 500 simulations, the median of the mean absolute error in the HS-100 variant is approximately 0.05 below the median mae of the ES-100 variant. From 500 simulations on, the difference shrinks but the variance remains lower for the HS-100 variant, i.e. the training of the neural network is more robust when it starts from a small uniformly distributed set of training data as a basis.

6. Conclusions

We discussed the data generation for and the evaluation of surrogate models that are used to guide global derivative-free flowsheet optimization, with the goal to avoid non-

promising computationally expensive calls to the process simulator based on the evaluation of process constraints. We propose a heuristic for a dynamic boundary that is calculated based on the metrics of the surrogate-model, e.g. the mean absolute error, and can be applied whenever a process variable is restricted by an inequality constraint. The surrogate models and their metrics depend on training data that may not reflect the interesting regions of the search space, especially in the early stages of optimization. We confirmed the efficiency of the heuristic by an a posteriori evaluation of the surrogate models. We compared two methods to collect training data for the initial surrogate models, data from a Halton set and data generated by the initial simulations. The robustness of the training of the surrogate-models is improved if a small number of uniformly distributed samples are used, i.e. the application of the Halton set is beneficial.

In this work, we evaluated the performance of the surrogate-models after the optimization. However, an evaluation of past surrogate models during the optimization is advisable, e.g. a revision of decisions that were based on too aggressive surrogate-models. The training data consists of the data seen by the optimization and converges in the direction of regions of a high interest, e.g. around a constraint limit. In adaptive sampling methods, surrogate model predictions are used to generate promising new sample points. Ludl et al. (2021) propose an adaptive sampling approach that balances multiple goals, finding the border between convergent and divergent simulations and exploring sample points to reduce the uncertainty of the surrogate model. Winz et al. (2021) propose an upper confidence bound acquisition function to train surrogate models that provide more information regarding the target function of the optimization. To include such approaches is a promising direction for future research.

Acknowledgement

Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) - TRR 63 ”Integrierte chemische Prozesse in flüssigen Mehrphasensystemen” (Teilprojekt D1) - 56091768. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - TRR 63 ”Integrated Chemical Processes in Liquid Multiphase Systems” (subprojects D1) – 56091768

References

N. Asprion, M. Bortz, 2018, Chemie-Ingenieur-Technik., 90, 11, 1727–1738.

H.-G. Beyer, H.-P. Schwefel, 2002, Nat. Comput., 1, 1, 3–52.

R. T. Haftka, D. Villanueva, A. Chaudhuri, 2016, Struct. Multidiscip. Optim., 54, 1, 3–13.

T. Janus, S. Engell, Chemie-Ingenieur-Technik. 2021, 93, 12, 2019-2028

G. Kiedorf, D. M. Hoang, A. Müller, A. Jörke, J. Markert, H. Arellano-garcia, A. Seidel-morgenstern, C. Hamel, 2014, Chem. Eng. Sci., 115, 31–48.

P. O. Ludl, R. Heese, J. Höller, N. Asprion, M. Bortz, 2021, Front. Chem. Sci. Eng.

V. A. Merchan, G. Wozny, 2016 Ind. Eng. Chem. Res., 55, 1, 293–310.

M. Urselmann, T. Janus, C. Foussette, S. Tlatlik, A. Gottschalk, M. T. M. M. Emmerich, T. Bäck, S. Engell, 2016, Comput. Aided Chem. Eng. 38, 187–192.

J. Winz, S. Engell, 2021, Comput. Aided Chem. Eng., 50, 2004, 953–958.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50261-X

Flowsheet Recognition using Deep Convolutional Neural Networks

Lukas Schulze Balhorn[a], Qinghe Gao[a], Dominik Goldstein[b], Artur M. Schweidtmann[a,*]

[a]*Department of Chemical Engineering, Delft University of Technology, Van der Maasweg 9, Delft 2629 HZ, The Netherlands*
[b]*Aachener Verfahrenstechnik - Process Systems Engineering, RWTH Aachen University, Aachen 52062, Germany*

a.schweidtmann@tudelft.nl

Abstract

Flowsheets are the most important building blocks to define and communicate the structure of chemical processes. Gaining access to large data sets of machine-readable chemical flowsheets could significantly enhance process synthesis through artificial intelligence. A large number of these flowsheets are publicly available in the scientific literature and patents but hidden among innumerable other figures. Therefore, an automatic program is needed to recognize flowsheets. In this paper, we present a deep convolutional neural network (CNN) that can identify flowsheets within images from literature. We use a transfer learning approach to initialize the CNN's parameter. The CNN reaches an accuracy of 97.9% on an independent test set. The presented algorithm can be combined with publication mining algorithms to enable an autonomous flowsheet mining. This will eventually result in big chemical process databases.

Keywords: Flowsheet, Data Mining, Image Classification, Deep Learning, Transfer Learning

1. Introduction

In recent years, machine learning (ML) has emerged as a popular method to solve complex problems in various domains. This popularity has predominantly been driven by (i) the increase of computational power, (ii) the improvement of ML algorithms, and (iii) the availability of big data (LeCun *et al.*, 2015). Chemical engineering has already seen many successful applications of ML (Schweidtmann *et al.*, 2021; Venkatasubramanian, 2018). However, literature on the structural synthesis of chemical processes through ML is scarce (c.f. (d'Anterroches & Gani, 2005; Zhang *et al.*, 2018; Oeing *et al.*, 2021)). While a variety of promising ML methods exist, big chemical process data is missing (Schweidtmann *et al.*, 2021; Weber *et al.*, 2021). We argue that this lack of structured chemical process data is hindering further progress of ML developments for chemical process synthesis.

The topological information about chemical processes is usually communicated through flowsheets. Flowsheets are technical drawings describing the unit operations connectivity of a process. There exists at least one flowsheet for every chemical process. Eventhough

most flowsheets are only available in internal company reports, a large number of flowsheets are also publicly available in scientific publications and patents. These flowsheets are mostly depicted on figures in PDF documents. However, searching for the flowsheet figures in scientific publications and patents can be as difficult as looking for a needle in a haystack. In particular, a manual search through the enormous amount of available literature would not only be a slow and labor-intense process, but it would also be prone to errors. Therefore, an algorithm is needed that autonomously recognizes flowsheet images.

In the previous literature, information extraction from scientific literature has mostly focused on text mining using natural language processing (Hong *et al.*, 2021; Nasar *et al.*, 2018). In the context of chemistry for example, Swain & Cole (2016) developed the ChemDataExtractor which extracts chemical identifiers, spectroscopic attributes, and chemical property attributes from scientific literature. Furthermore, information extraction from scholarly images has been performed in the past. The majority of research on the classification of scientific images has been conducted on biomedical literature pushed by the yearly ImageCLEF challenge (c.f. (Pelka *et al.*, 2020)). Furthermore, a few works exist in chemistry on information extraction from images. This works mostly focus on the recognition and digitization of structural formulas (Tharatipyakul *et al.*, 2012; Beard & Cole, 2020). Another example for chemical image analysis is the ImageDataExtractor which mines microscopy images to extract information about the particle sizes and shapes (Mukaddem *et al.*, 2019). However, to the best of our knowledge, image classification has not been applied to chemical process design literature and there exists no previous algorithm that identifies chemical flowsheet images.

In this work, we propose an algorithm that recognizes flowsheet images from chemical engineering journal articles. The proposed algorithm will contribute to our long-term vision to build a database of chemical processes. In Section 2., we provide a brief background on Convolutional Neural Networks (CNNs). In Section 3., we present our methods, data set, and pre-processing. In Section 4., we evaluate the performance of the proposed flowsheet image classification model and discuss the results. Finally, we conclude our findings in Section 5.

2. Deep Convolutional Neural Networks

Inspired by the biological visual system (O'Shea & Nash, 2015), deep CNNs have been proposed as a computational method to bridge the gap between the capabilities of humans and machines for high-level tasks such as image classification, text recognition, and speech recognition (LeCun *et al.*, 2015). The powerful performance of deep CNNs in advanced tasks is achieved through the layout of the framework, which generally consists of three parts: Convolutional layers, pooling layers, and fully-connected layers. Convolutional layers contain a set of learnable filters that will convolve over the inputs to extract the underlying features. Intuitively, simple features such as edges, corners, and blotches will be detected in the early convolutional layers. Ultimately, more complex patterns such as 'unit operations' will appear with further layers. Pooling layers are usually periodically inserted between two convolutional layers to reduce the spatial dimension and the number of parameters. Average pooling and max pooling are the most common choices. Fully-connected neural network layers play the role of mapping the learned "distributed feature representation" to the sample label space, namely, making a classification. Additionally,

to introduce nonlinearity into the output, activation functions such as sigmoid, ReLu, and hyperbolic tangent are usually included after convolutional or fully-connected layers. Furthermore, the size of the training data is an important factor for the performance of the deep CNNs and data-labeling is often expansive. Therefore, the concept of transfer learning emerged in recent years. In transfer learning, the CNN is first trained with a sufficiently big data set from one domain of interest. Afterward, the data set of the classification task from another domain of interest is used to fine-tune the CNN.

3. Method

The flowsheet recognition algorithm aims to identify flowsheets among a large number of images. We train a deep CNN for the recognition algorithm based on manually labeled images mainly from scientific journal articles.

3.1. Data Set

At present, no public data set of flowsheet images exists. To create a training data set, we automatically mine figures of scientific journal articles. First, we retrieve a list of all DOIs for a given journal ISSN from the crossref API. Then, the PDFs of the corresponding journal articles are downloaded through publisher APIs. Subsequently, all figures are extracted from the PDFs using the Python package PyMuPDF. The describe procedure is applied to the journals "Theoretical Foundations of Chemical Engineering" and "Frontiers of Chemical Science and Engineering" to generate an initial dataset. Subsequently, the extracted images are manually reviewed and labeled as being a flowsheet or not. In addition to the figures from scientific journal articles, we also add flowsheet images retrieved from a google search to our data set. In total, our data set contains about 1,000 flowsheet images and about 13,000 other images from scientific publications.

3.2. Data Augmentation and Oversampling

As a result of the data mining from journal articles, the data set is imbalanced. In particular, there exist far fewer flowsheet images than other images. This imbalance can cause the classifier to develop a bias towards the majority class. To overcome this issue, oversampling has been used in previous studies (Johnson & Khoshgoftaar, 2019). We oversample the flowsheet images by a factor of 13 to balance the data set. As this large oversampling factor can cause overfitting, we also employ a data augmentation technique (Shorten & Khoshgoftaar, 2019). Each copy of a flowsheet image is augmented by stretching it along the horizontal and vertical axis independently by a random factor between 0.7 and 1.2. Other common data augmentation techniques such as shifting, rotation, and shearing were dismissed because they are expected to destroy some key features of flowsheet images. For example, flowsheets usually include horizontal and vertical lines making image rotation pointless. The images of the negative "other" class are not augmented because of abundant data availability.

3.3. Model Training

The CNN architecture for the flowsheet recognition is based on the VGG16 network by Simonyan & Zisserman (2014). The network includes 13 convolutional layers, 5 max

pooling layers, and 3 fully-connected layers. Since our data set is limited, we use a transfer learning approach. In particular, we use the publicly available VGG16 network that has been pre-trained on the ImageNet data set including tens of millions of images and 1,000 categories. To adapt the network to the use case of this work, we reduced the number of nodes in the output layer to two. The training is conducted using the PyTorch framework which is built on the Torch library. The model takes in images with a resolution of 224 × 224 pixels. We randomly divide our data set into training (70%), validation (15%), and test (15%) data set. The model is trained on the training data in batches of 150 images. The validation set was used to validate the training progress and tune the hyperparameters of the model. The independent test data set is used for the final performance evaluation. Notably, the test set is truly independent as it does not contain any augmented images from the training or validation sets.

4. Results and Discussions

The most important performance metrics for classifiers is the accuracy as defined in Eq. 1. In the light of class imbalance, we also evaluate the precision (Eq. 2) and recall (Eq. 3):

$$Accuracy = \frac{TN + TP}{TP + FP + TN + FN}, \tag{1}$$

$$Precision = \frac{TP}{TP + FP}, \tag{2}$$

$$Recall = \frac{TP}{TP + FN}, \tag{3}$$

where TP denotes true positive, TN denotes true negative, FP denotes false positive and FN denotes false negative. The training history is shown in Fig. 1. The classifier reaches a satisfying accuracy already after the first epoch. This good initial performance can be explained by the use of a pre-trained model. After the second epoch of training, the classifier shows a validation accuracy of over 98%. The training process was ended after 10 epochs. In training runs with more epochs no further improvement was experienced. The final training accuracy after 10 epochs is 98.1% while the validation accuracy is 98.2%. Notably, we do not observe any overfitting behavior in the training process.

Overall, the flowsheet recognition algorithm shows a promising performance on the independent test set. The confusion matrix on the test set is shown in Table 1. Of all predictions on the test set, 97.9% were correct. Furthermore, the precision is 80.7% and lower than the recall with 94.4%. The high recall shows that almost all flowsheet images are retrieved while the number of false negative flowsheets is very low. Furthermore, the fairly low precision could be explained by the class imbalance. The data set contains about thirteen times more images of the class “other”. If only a small fraction of the class “other” is misclassified, these images already make up a great share of the flowsheet predictions.

Finally, the runtime of the image classification is investigated. The evaluation of an image by the trained CNN takes about 7 milliseconds on average on a personal computer. This short evaluation time allows for an online application that autonomously mines flowsheets from literature.

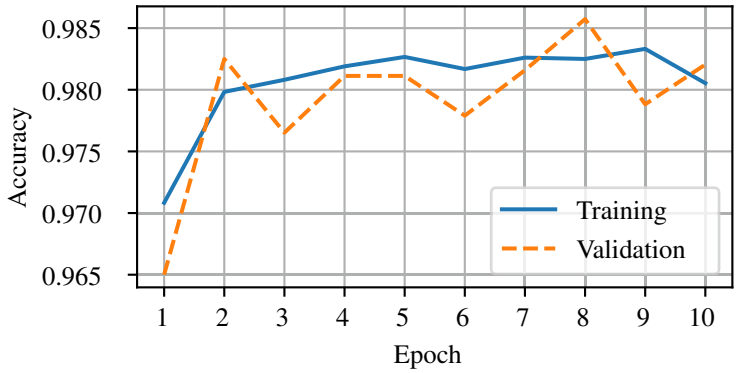


Figure 1: Training history of the CNN.

Table 1: Confusion matrix for the flowsheet recognition algorithm on the test set.

	Actual flowsheet	Actual other
Predicted flowsheet	151	36
Predicted other	9	1,976

5. Conclusions

We propose an image classification algorithm that can recognize flowsheet images. The algorithm consists of a deep CNN which classifies images with a high accuracy of 97.9%. In order to train the CNN, we mined about 1,000 flowsheet images from scientific literature and online search engines. Moreover, the transfer learning improved the prediction accuracy. The proposed tool can be used to automatically identify flowsheet images from scientific literature or other sources within a few milliseconds. In a preliminary study we applied our mining algorithm to the journal "Computers & Chemical Engineering" and identified more than 1500 flowsheets. Future work will digitize the flowsheet images to identify process topologies. This will eventually result in an open-source knowledge graph database providing chemical processes in a structured format. We believe that this database has a tremendous value for future process design because it allows the search and optimization over existing processes. In addition, our database will eventually serve as a training database for advanced ML algorithms able to design novel processes.

References

Beard, Edward J., & Cole, Jacqueline M. 2020. ChemSchematicResolver: A Toolkit to Decode 2D Chemical Diagrams with Labels and R-Groups into Annotated Chemical Named Entities. *Journal of Chemical Information and Modeling*, **60**(4), 2059–2072.

d'Anterroches, Loïc, & Gani, Rafiqul. 2005. Group contribution based process flowsheet synthesis, design and modelling. *Fluid Phase Equilibria*, **228-229**(Feb.), 141–146.

Hong, Zhi, Ward, Logan, Chard, Kyle, Blaiszik, Ben, & Foster, Ian. 2021. Challenges and Advances in Information Extraction from Scientific Literature: a Review. *JOM*, Oct.

Johnson, Justin M., & Khoshgoftaar, Taghi M. 2019. Survey on deep learning with class imbalance. *Journal of Big Data*, **6**(27).

LeCun, Yann, Bengio, Yoshua, & Hinton, Geoffrey. 2015. Deep learning. *Nature*, **521**(May), 436–444.

Mukaddem, Karim T., Beard, Edward J., Yildirim, Batuhan, & Cole, Jacqueline M. 2019. ImageDataExtractor: A Tool To Extract and Quantify Data from Microscopy Images. *Journal of Chemical Information and Modeling*, **60**(5), 2492–2509.

Nasar, Zara, Jaffry, Syed Waqar, & Malik, Muhammad Kamran. 2018. Information extraction from scientific articles: a survey. *Scientometrics*, **117**(3), 1931–1990.

Oeing, Jonas, Henke, Fabian, & Kockmann, Norbert. 2021. Machine Learning Based Suggestions of Separation Units for Process Synthesis in Process Simulation. *Chemie Ingenieur Technik*, Sept.

O'Shea, Keiron, & Nash, Ryan. 2015. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*.

Pelka, Obioma, Friedrich, Christoph M, García Seco de Herrera, Alba, & Müller, Henning. 2020 (Sept.). Overview of the ImageCLEFmed 2020 Concept Prediction Task: Medical Image Understanding. *In: Proceedings of the CLEF 2020-Conference and labs of the evaluation forum*.

Schweidtmann, Artur M., Esche, Erik, Fischer, Asja, Kloft, Marius, Repke, Jens-Uwe, Sager, Sebastian, & Mitsos, Alexander. 2021. Machine Learning in Chemical Engineering: A Perspective. *Chemie Ingenieur Technik*, Oct.

Shorten, Connor, & Khoshgoftaar, Taghi M. 2019. A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, **6**(60).

Simonyan, Karen, & Zisserman, Andrew. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Swain, Matthew C., & Cole, Jacqueline M. 2016. ChemDataExtractor: A Toolkit for Automated Extraction of Chemical Information from the Scientific Literature. *Journal of Chemical Information and Modeling*, **56**(10), 1894–1904.

Tharatipyakul, Atima, Numnark, Somrak, Wichadakul, Duangdao, & Ingsriswang, Supawadee. 2012. ChemEx: information extraction system for chemical data curation. *BMC Bioinformatics*, **13**(17).

Venkatasubramanian, Venkat. 2018. The promise of artificial intelligence in chemical engineering: Is it here, finally? *AIChE Journal*, **65**(2), 466–478.

Weber, Jana M., Guo, Zhen, Zhang, Chonghuan, Schweidtmann, Artur M., & Lapkin, Alexei A. 2021. Chemical data intelligence for sustainable chemistry. *Chemical Society Reviews*.

Zhang, Tong, Sahinidis, Nikolaos V., & Siirola, Jeffrey J. 2018. Pattern recognition in chemical process flowsheets. *AIChE Journal*, **65**(2), 592–603.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50262-1

Active learning for multi-objective optimization of processes and energy systems

Julia Granacher[a]* and François Maréchal[a]

[a]*Industrial Process and Energy Systems Engineering Group, Ecole Polytechnique Fédérale de Lausanne Valais Wallis, Rue de l'Industrie 17,Case Postale 440, 1951 Sion, Switzerland*
julia.granacher@epfl.ch

Abstract

In superstructure optimization of processes and energy systems, the design space is defined as the combination of unit considerations, process conditions and model parameters that might be subjected to uncertainty. Most of the time, decision makers are not looking for a single best solution, but rather are interested in analyzing a set of Pareto-optimal superstructure designs. The generation of Pareto-optimal solutions is computationally expensive, especially if nonlinear process evaluations or simulation is required. In our approach, we address the question of how to efficiently generate Pareto-optimal sets of solutions by applying machine learning concepts. Using the criteria of Pareto-optimality to evaluate the performance of a set of design space variables and the corresponding solution, we train our algorithm on predicting if a solution is belonging to the Pareto frontier. Following the approach presented by Zuluaga et al., (2016) and applied to the design of materials by Jablonka et al., (2021), an adaptive learning concept is used to systematically identify the next best function evaluation to improve the confidence of the Pareto-frontier definition. Gaussian process surrogate models provide a prediction of the mean and the standard deviation of the relevant objectives. Design points with high probability to of being in the Pareto-optimal domain are evaluated by the original model, increasing the confidence with which the Pareto front is predicted. Simultaneously, the design space is continuously reduced by discarding the design points for which the probability of being in the set of Pareto-optimal solutions is low. The procedure is stopped when all points are labeled as Pareto-optimal or discarded. The algorithm is applied to the design of a utility superstructure for an industrial energy system. Our algorithm is compared and benchmarked with quasi random sampling of the design space.

Keywords: Multi-objective Optimization; Active Learning; Energy System Design; Utility Superstructure; Mathematical Programming; Machine Learning; Artificial Intelligence

1. Introduction

One of the most pressing challenges our society is facing is climate change, revealing the need for efficient and reliable design methods of energy systems that are sustainable in economic, environmental and social terms. Over the last decades, the methodologies on energy system and process design have evolved drastically. Process systems engineering (PSE) has developed as a conceptual element of chemical engineering, including the definition, design, planning and control of complex chemical processes (Mencarelli et al., 2020). PSE was initially dominated by the progress made in process simulation, where algebraic equations and flowsheeting methods are applied to describe

a system's behavior. The emerging focus on quantitative descriptions of processes by means of simulation led to a more thorough analysis of the system performance (Rudd et al., 1973). Process synthesis addresses the development, simulation, and optimization of processes, where the unit operations are selected and interconnections are defined (Biegler et al., 1997; Douglas, 1988). In superstructure optimization, a network of all potential unit operations and connections is defined and translated into a mathematical programming model, which is then used for generating results by solving an optimization problem. Instead of generating one optimal design that may only be valid under certain external conditions, the generation of a set of feasible alternatives may be preferred.

Multi-objective optimization is widely applied for analyzing trade-offs between two or more objective functions. The set of optimal solutions obtained from a multi-objective optimization problem can be displayed in a Pareto-optimal curve, on which, for each point on the curve, none of the objectives can be improved without penalizing the others. However, most optimization techniques rely on the introduction of a total order in the search space, which biases the search and may introduce technical difficulties (Jablonka et al., 2021). Machine learning has gained interest recently for designing processes with complex design spaces, as they allow for the fast prediction of process performance. However, training large datasets makes the problem unnecessarily computationally expensive, especially when simulation-based approaches are included in the superstructure.

In this paper, we are addressing the question how Pareto fronts of large energy superstructure optimization problems with non-linear relations between the design space and the objectives can be generated efficiently with the assistance of machine learning algorithms, ensuring reliable predictions of the system performance along the Pareto front. Thereby, an active learning approach is applied, which iteratively improves the machine learning model where it is needed the most.

2. Methodology

A modified implementation of the of the ϵ-PAL algorithm introduced by Zuluaga et al., (2016) and implemented by Jablonka et al., (2021) is applied to the optimization of energy and process system superstructures. The ϵ-PAL algorithm iteratively reduces the effective design space by discarding those design points from which we know that they are Pareto-dominated by another design point. The design point with the highest dimensionless uncertainty from a set of possible design points predicted to be near the Pareto front is evaluated. When all points are either classified or discarded, the search ends. The approach offers the additional benefit of enabling the tuning between accuracy and efficiency, by tuning the granularity of the approximation to the Pareto front in every objective. In the following sections, the general superstructure problem as well as the ϵ-PAL algorithm is described.

2.1. Superstructure optimization

A process superstructure has the aim of describing the system's units and the way they can interact with others. By activating certain units and their connections, different system configurations are achieved. In this work, the methodology for superstructure modelling and optimization is adapted from Gassner and Maréchal, (2009). For each unit in the system, energy and mass flow balances are formulated, describing

corresponding transformations. They are derived using either flowsheeeting or simplified black-box models. Binary decision variables describe whether a unit is installed and used in a certain period. Continuous decision variables describe the installed size of the unit and the level of usage at which it is operated in each period. Parameterized bounds constrain continuous and binary variables. In our approach, all process units are connected to a heat exchange system, which allows for the exchange between the process and the hot/cold utilities to close the energy balance. Operating and investment costs are derived based on equipment size. Environmental impacts of the system are estimated using the Life Cycle Inventory Ecoinvent database (Wernet et al., 2016). Pinch analysis is applied to model heat recovery opportunities and to investigate the integration of the utility system by introducing the heat cascade constraints explained in Maréchal and Kalitventzeff, (1998).

For generating a solution, the decision variables are fixed solving a mixed-integer linear programming (MILP) problem formulated in the AMPL optimization language (Fourer et al., 2002), using the CPLEX branch-and-bound algorithm (IBM, 2017). The superstructure model in the lower-level is integrated in an upper-level framework, in which optimization problems are formulated for exploring the impact of non-linear decision variables on the results, such as operating conditions of the utility system. For a problem communicated by the upper-level framework, the lower-level generates a solution and reports it to the upper-level. For generating and communicating optimization problems to the lower-level, different approaches could be envisioned. Besides the usage of random sampling, the application of evolutionary algorithms or the hereafter presented Pareto-active learning approach can be considered.

2.2. Pareto-active learning

For efficiently and reliably identifying the Pareto front, we apply the modified version of the ϵ-PAL algorithm presented in Jablonka et al., 2021. In the ϵ-PAL algorithm, the uncertainty estimation σ of a Gaussian process is used to construct hyperrectangles for a prediction (Figure 1A). The algorithm enables the classification of Pareto-optimal samples, as well as the proposition of the next best sample for evaluation.

The ϵ-PAL algorithm starts with a set of experiments, and the desired objectives are calculated by calling the original model for a subset of the generated samples. An initial machine learning model is trained on the obtained dataset of decision variables and objectives, and predictions for unlabeled datapoints are made. For each prediction, hyperrectangles around the prediction mean are constructed, the width being equivalent to the standard deviation of the posterior of the Gaussian process. The lower and upper limits are equivalent to the best/worst performance estimates (Figure 1A). The points to be discarded with confidence can be identified from the ϵ-Pareto dominance relation, as well as the ones with a high probability of being Pareto-optimal (Figure 1B). If the pessimistic estimate of a prediction is greater than a defined tolerance above the optimistic estimate of all other predictions, it will be part of the Pareto front. For estimating the accuracy of the Pareto front, one can connect the bottom left corners of hyperrectangles associated with the current estimate of the front, which gives the most pessimistic front (dashed blue line in Figure 1B). The optimistic front is then obtained by connecting the upper right corners.

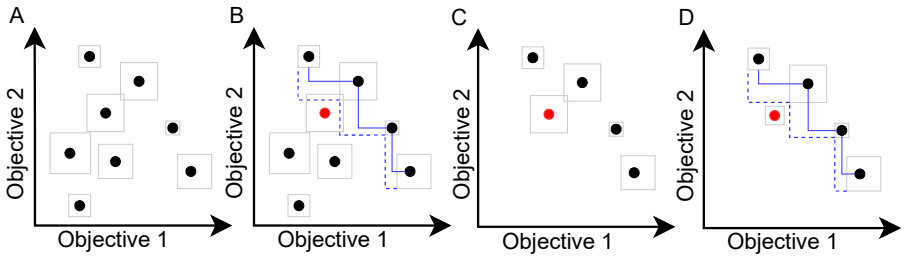


Figure 1: A: hyperrectangles based on predicted mean and variance. B: Pareto-optimal points identified. C: Relevant design space shrinks, red point identified as new point to label. D: Uncertainty in red point reduced after retraining the model with new sample, adapted from Jablonka et al., (2021).

Thus, a geometric construction is created that allows for classification whether a predicted solution is Pareto-optimal or whether it can be discarded. The next design point to be evaluated by the optimization is identified as the one that reduces the uncertainty in classifying points as Pareto-optimal. For this, it is assumed that the uncertainties are normalized by the predicted mean, so that the area of the hyperrectangles represent the relative error. The prediction model is then improved by reducing the uncertainty of the largest rectangle among the points presumed near the Pareto front (Jablonka et al., 2021). In Figure 1C, the red point is identified as assisting the model improvement most, and the more accurate estimate of the updated model is represented in Figure 1D. This procedure is repeated until the desired accuracy of the Pareto front is reached. For further information about the algorithm and its implementation, the reader may consult Jablonka et al., (2021); Zuluaga et al., (2016).

3. Application

The proposed methodology is applied to the design of a Pareto front for the optimal operation of a steam network integrated in a large-scale Kraft pulping process, producing 1000 air-dried tons of pulp per day. The pulp and paper industry is known as an energy-intensive industrial sector, consuming large quantities of water and energy. Integration techniques, including heat integration, water network optimization and steam cycle operation optimization as applied in Kermani et al., (2019) can significantly reduce water and energy consumption of the mill. In this contribution, we focus explicitly on the optimization of the operating conditions of a steam network integrated with the pulp mill by applying the ϵ-PAL algorithm. In the steam network, steam can be produced between 50 and 160 bar, while it can be consumed at 3 pressure levels. To meet the specification of combined steam and electricity production in industrial plants, steam production can only happen at the highest pressure level, and turbines are placed between the highest pressure level and the subsequent levels. The superstructure model is adapted from Kermani et al., (2019), which might be consulted for more information. Objectives selected for optimization and representation in a Pareto front are the annual operating cost (OPEX) and the annualized capital cost (CAPEX). The ϵ-PAL algorithm is applied on the upper-level, evaluating samples from the decision space for retrieving Pareto fronts. For the application of the ϵ-PAL algorithm, the decision variables in Table 1 are used. Two ϵ-PAL instances are created, one with 200 and one with 500 samples. In the first iteration, 50 samples are labeled, while in each following iteration, 10 new samples are labeled.

Table 1: Decision variables for designing the Pareto front, adapted from Kermani et al., (2019).

Decision variable	Range, Unit	Description
p_1^{st}	[50;160], bar	Boiler pressure
p_2^{st}	[9;14], bar	High-pressure steam header
p_3^{st}	[3;8], bar	Medium-pressure steam header
p_4^{st}	[0.5;2], bar	Low-pressure steam header
T_1^{sup}	[150;300], °C	Degree of superheating in the highest pressure level

Results

When running the ϵ-PAL algorithm on the small set of samples, five Pareto optimal points are identified. The learning curve in Figure 2A shows that after ten iterations, all design points are either discarded or identified as Pareto-optimal. Figure 2B shows the classification of all design points. The error bars indicate the obtained uncertainty which is used for calculating the hyperrectangles when qualifying whether a point might be Pareto-optimal or not. Overall, 140 design points are labeled, and the hypervolume obtained is 9614. Computation took 83 minutes, while labelling all 200 samples for obtaining the same Pareto front takes 118 minutes, indicating time savings of 30%.

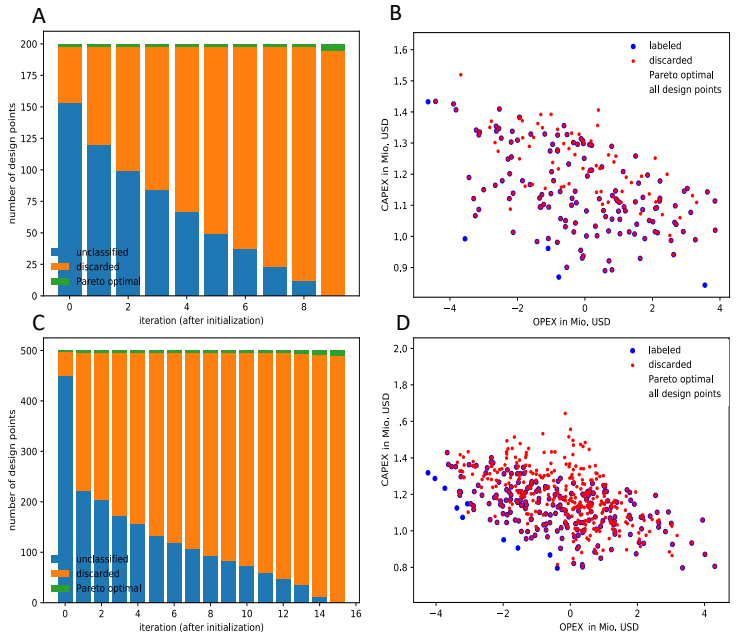


Figure 2: Results for applying ϵ-PAL algorithm. A: Learning curve for 200 samples, B: Pareto front for 200 samples, C: Learning curve for 500 samples, D: Pareto front for 500 samples.

Figure 2C shows the learning curve for 500 sampled design points. After 16 iterations, all design points are either discarded or classified as Pareto-optimal. A total of 11 Pareto-optimal points is identified, the rest is discarded. The hypervolume obtained from this Pareto front is 9651, so slightly higher than for the smaller design space. Over all iterations, a total of 200 samples were labeled by calling the original optimization

model. The computational time was 122 minutes. Compared to random sampling and evaluation of 500 samples, this accounts for time savings of approximately 60%. It is worth noting that all points identified as Pareto-optimal are also labeled, meaning that the error in this region of the design space is minimized. This ensures reliable results in the Pareto-optimal domain.

4. Conclusion and outlook

In this contribution, we demonstrated that the ϵ-PAL algorithm can be used to identify Pareto-optimal design points for process and energy system superstructures. Compared to random sampling, significant time savings were obtained, the relative savings increasing with the size of the design space. The quality of the Pareto front, measured with the hypervolume error, increases with the design space size. The algorithm manages to identify relevant regions in the design space as Pareto-optimal or near Pareto-optimal, allowing for continuous improvement of the prediction quality where it is necessary. Future work will include benchmarking the applied methodology to a genetic algorithm, as well as obtaining results for larger design spaces.

Acknowledgements

This research has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 81801.

References

L.T. Biegler, I.E. Grossmann, A.W. Westerberg, 1997. Systematic methods for chemical process design, 1st ed, Prentice Hall. Prentice Hall.

J.M., Douglas, 1988. Conceptual design of chemical processes. McGraw-Hill.

M. Gassner, F. Maréchal, 2009. Methodology for the optimal thermo-economic, multi-objective design of thermochemical fuel production from biomass. Comput. Chem. Eng., https://doi.org/10.1016/j.compchemeng.2008.09.017

K.M. Jablonka, G.M. Jothiappan, S. Wang, B. Smit, B. Yoo, 2021. Bias free multiobjective active learning for materials design and discovery. Nat. Commun. 12. https://doi.org/10.1038/s41467-021-22437-0

M. Kermani, I.D. Kantor, A.S. Wallerand, J. Granacher, A.V. Ensinas, F. Maréchal, 2019. A Holistic Methodology for Optimizing Industrial Resource Efficiency. Energies 12, 1315. https://doi.org/10.3390/en12071315

F. Maréchal, B. Kalitventzeff, 1998. Process integration: Selection of the optimal utility system. Comput. Chem. Eng., European Symposium on Computer Aided Process Engineering-8 22, S149–S156. https://doi.org/10.1016/S0098-1354(98)00049-0

L. Mencarelli, Q. Chen, A. Pagot, I.E. Grossmann, 2020. A review on superstructure optimization approaches in process system engineering. Comput. Chem. Eng. 136, 106808. https://doi.org/10.1016/j.compchemeng.2020.106808

D.F. Rudd, G.J. Powers, J.J. Siirola, 1973. Process Synthesis. Prentice-Hall.

G. Wernet, C. Bauer, B. Steubing, J. Reinhard, E. Moreno-Ruiz, B. Weidema, 2016. The ecoinvent database version 3 (part I): overview and methodology. Int. J. Life Cycle Assess. 21, 1218–1230. https://doi.org/10.1007/s11367-016-1087-8

M. Zuluaga, A. Krause, M. Puschel, 2016. e-PAL: An Active Learning Approach to the Multi-Objective Optimization Problem. Journal of Machine Learning_Research 17

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50263-3

Data-driven Stochastic Optimization of Numerically Infeasible Differential Algebraic Equations: An Application to the Steam Cracking Process

Burcu Beykal[a,b*], Zahir Aghayev[a,b], Onur Onel[c,d], Melis Onel[c,d], Efstratios N. Pistikopoulos[c,d]

[a]*Department of Chemical and Biomolecular Engineering, University of Connecticut, Storrs, CT 06269, USA*
[b]*Center for Clean Energy Engineering, University of Connecticut, Storrs, CT 06269, USA*
[c]*Texas A&M Energy Institute, Texas A&M University, College Station, TX 77843, USA*
[d]*Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX 77843, USA*
burcu.beykal@uconn.edu

Abstract

In this work, we address the data-driven stochastic optimization of the numerically infeasible differential algebraic equations (DAEs) using Support Vector Machines (SVMs) and scenario analysis. Data-driven optimization is an attractive method for optimizing systems with highly complex first-principles models using iterative sampling, surrogate modeling, and optimization steps. Yet, the numerical stability and the unknown interconnections between the initial conditions of the DAEs determine the overall performance of the data-driven optimizer. Specifically, in the sampling step where the DAE system is initialized, varying samples of initial conditions can cause premature termination of the simulation due to numerical infeasibilities, without retrieving any viable output data that is essential for the surrogate modeling and grey-box optimization steps. These challenges are further amplified when there are stochastic elements present in the system which the numerically infeasible system of DAEs needs to handle to achieve robust solutions. Using the steam cracking process as our motivating example, the SVMs are used to accurately map the feasible region of the numerically infeasible system of DAEs representing the first-principles models of the cracking reactor, while the scenario analysis allows us to handle the uncertainty in the feed composition of the natural gas liquid (NGL). The resulting modeling framework is incorporated in a data-driven optimization solver and utilized to generate the guaranteed feasible solution for the design and operation of an NGL steam cracking reactor under uncertain feed compositions.

Keywords: Data-driven optimization, support vector machines, stochasticity, scenario analysis, steam cracking process.

1. Introduction

The first-principles representations of many unsteady-state or dynamic chemical engineering systems are composed of ordinary or partial differential equations (i.e., mass, energy, and momentum balances) and algebraic expressions (i.e., rate law), creating a system of differential algebraic equations (DAEs). The optimal decision-making with

both differential and algebraic components is not straightforward since the deterministic optimization solvers cannot be directly implemented with such formulations. Typically, dynamic optimization of DAE systems is handled through: (1) Process simulation software, like gPROMS or Aspen Custom Modeler (Lang and Biegler, 2007); (2) Orthogonal collocation on finite elements to reduce a dynamic optimization problem to a constrained nonlinear problem (Biegler, 1984), or (3) Data-driven modeling and optimization (Beykal et al., 2020).

Optimization with dynamic programs can also be complicated by numerical infeasibilities, like stiffness, ill-conditioned algebraic equations, and undefined solutions. Orthogonal collocation on finite elements can handle such challenges easily with its discretization strategy, however, the resulting large-scale nonlinear program (NLP) is typically solved to local optimality (Caballero et al., 2015). Data-driven optimization enables the exploration of global solutions through sampling, surrogate modeling, and deterministic optimization steps (Beykal et al., 2018). Yet, such algorithms are not designed to handle numerical infeasibilities and implicit constraints. Especially at the sampling stage, when candidate sampling points for the decision variables (i.e., the initial conditions of the DAE system) are identified, sampling procedures like the Design of Experiments will assume that all decision variables are independent of each other. However, there might be interdependencies among variables that define the numerical stability of the solution and can only be represented in explicit mathematical forms if the analytical solution exists. As a result, without the explicit a priori knowledge on these interdependencies, data-driven optimization algorithms will return unrealistic solutions that cannot be validated by the first principles-based simulations. Recently, Beykal et al. (2020) developed a Support Vector Machine (SVM) based data-driven optimization framework to overcome numerical infeasibilities and implicit constraints by mapping the numerical feasibility boundary with a two-class classification model. This study showed that implicit constraints are accurately captured with SVMs, significant computational savings are achieved with the SVM classifier, and guaranteed feasible solutions are attained with data-driven optimization techniques for such difficult cases of DAE systems.

Despite recent efforts, the challenges in numerically infeasible DAE systems are further complicated with uncertain initial conditions. Even when the deterministic solution is retrieved by the aforementioned SVM approach, a slight deviation in the optimal input conditions due to system disturbances could still lead to failures and undesirable outcomes. For example, in a reaction system like steam cracking, the input natural gas liquid (NGL) feed composition can be uncertain, or the feed composition can be adjusted to maximize profit with changing market conditions. Hence, the steam cracker design and operation problem should consider these stochastic elements to be able to offer a solution that is flexible and robust against changing initial conditions.

Motivated by this, in this work, we further extend our previous analysis on handling numerically infeasible DAE systems using SVMs and introduce stochastic initial conditions in the problem formulation to study the effects of uncertainty on the data-driven optimization performance. We demonstrate the effectiveness of our approach on the optimal steam cracker design and operation problem which is subject to stochastic NGL feed compositions. Details on the problem formulation and the data-driven optimization results of the computational case study are provided in the following sections.

2. Methodology

2.1. Steam Cracking Optimization: Problem Formulation

The steam cracking process is modeled using a plug flow reactor (diameter = 0.108 m) with mass, energy, and momentum balances under coking effects. Steam and NGL streams are co-fed at the reactor entrance (Figure 1). The plug flow reactor is subject to constant external heat flux, Q, across the reactor length, L. The model is one-dimensional and dynamic along the spatial coordinate, z with spatial changes in the molar of species, reactor pressure, and temperature are provided as $F_j(z)$, $P(z)$, and $T(z)$, respectively.

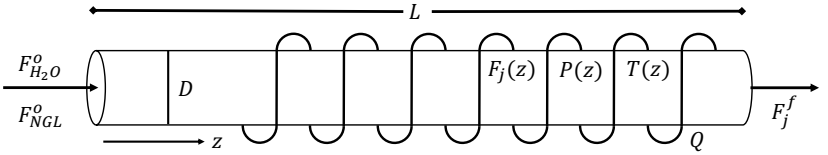


Figure 1 The plug flow reactor for modeling steam cracking of NGLs.

Our goal is to determine the optimal values of the reactor length, external heat flux, inlet pressure, inlet temperature, the inlet flowrates for the NGL and steam feeds to maximize the total stochastic profit from propylene and ethylene production subject to the stiff steam cracking model, the known constraint on the total initial flowrate, and output constraints on the reactor exit temperature and pressure (Eq.(1)). The objective function is calculated by multiplying the profit obtained from each scenario, n, and their corresponding probability of occurrence, ϕ^n. The profit at every scenario is obtained by numerically integrating the steam cracking model. Output reactor constraints are handled as grey-box constraints whereas the numerical feasibility of the mathematical models is handled through the SVM model. The detailed list of species, reaction mechanisms, and model equations considered in the formulation are available in Beykal et al., (2020).

$$\begin{aligned}
&\max\ Total\ Stochastic\ Profit = \sum_{n=1}^{N} \phi^n \cdot Profit^n \\
&s.t.\quad Mass, Energy, Momentum\ Balances \quad \forall n \\
&\qquad Rate\ Law\ \&\ Reaction\ Mechanism, Coking\ Effects \quad \forall n \\
&\qquad F^0_{H_2O} + F^o_{NGL} \le 0.05\ kmol/s \quad \forall n \\
&\qquad T^f \le 1300\ K \quad \forall n, \qquad P^f \ge 80\ kPa \quad \forall n \\
&\qquad T^{in} = [700\ K, 1100\ K], P^{in} = [290\ kPa, 500\ kPa], L = [5\ m, 100\ m] \\
&\qquad Q = [10\ kW/m^2, 1000\ kW/m^2], F^o_{H_2O}, F^0_{NGL} = [0.003\ kmol/s, 0.05\ kmol/s]
\end{aligned} \tag{1}$$

2.2. Scenario Analysis for the Stochastic Feed Compositions

Scenarios are created to represent the stochastic feed compositions in the steam cracking problem. We assume that the NGL feed is only composed of ethane and propane and create 11 scenarios with varying compositions of these compounds. The created scenarios aim to capture a variety of events that the steam cracker may encounter. For example, when ethylene demand increases, the reactor may be operated with a pure ethane feed. Likewise, when propylene demand increases, the reactor may be operated with pure

propane feed. In less extreme cases, the NGL feed composition may vary depending on the supplier or due to system disturbances. Hence, by taking the pure ethane and propane compositions as the endpoints in our scenario analysis, we create 9 other scenarios with 0.1 increments in the propane composition (Figure 2). All scenarios are assumed to have an equal probability of occurrence with $\phi_n = 1/11$, and the steam cracker model is solved for every scenario to calculate the total stochastic cost.

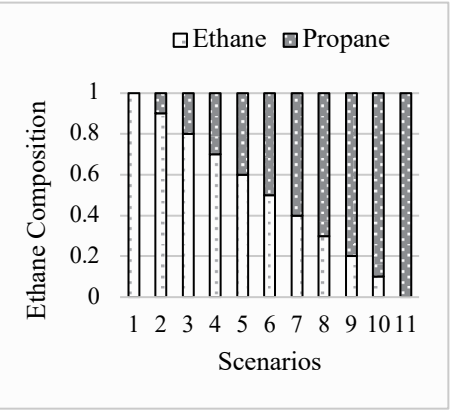


Figure 2 Representative scenarios for the stochastic NGL feed compositions.

2.3. Mapping the Feasible Region of Implicit Constraints with Support Vector Machines

The next step is to train the SVM model to map the feasibility boundary of the infeasible DAE system. SVMs are supervised learning models that are commonly used for regression, outlier detection, and classification analysis (Onel et al., 2019). Due to their highly flexible nature and ability to use nonlinear transformations, SVMs can learn highly nonlinear relationships within a dataset with high accuracy. To facilitate data-driven optimization and to ensure the validity of the final solution, the feasibility of the numerical integration of a DAE system is modeled using a two-class *C*-parametrized nonlinear SVM classification algorithm. As highlighted by Beykal et al. (2020), the main idea of this approach is two-fold: (1) to collect samples from the numerical integration and use their output information to train an SVM model of the feasible region for the DAE system (the offline phase); and (2) to incorporate this classifier within a grey-box optimization solver to eliminate infeasible solutions prior to sample collection (the online phase).

In the offline phase, we follow the recipe outlined in Beykal et al. (2020) and Onel et al. (2019) and construct a Latin Hypercube Design with 2000 points that satisfy the total maximum flowrate constraint in Eq.(1). Each sampling point is then numerically integrated across all feed composition scenarios and their discrete output information is collected as either “feasible - 0” or “infeasible - 1”. The output value of a sample is deemed feasible if the sampling point is successfully integrated without any failures in simulation across all scenarios. Otherwise, the sample is deemed infeasible and an output tag of “1” is assigned. The input data is then min-max scaled within the bounds of the decision variables and randomly split into train, validation, and test sets. 90 % of the data set is reserved for training the SVM model with 5-fold cross-validation and the remaining 10 % is reserved for blind testing the model performance. Gaussian radial basis function is used as the nonlinear kernel for the SVM model and the respective hyperparameters are tuned through an exhaustive grid search. Finally, the predictive capability of the trained classifier is assessed using several performance metrics, including the accuracy, precision, recall, F_1 score, and area under the curve (AUC).

Once this offline phase is completed, the SVM model is ready to filter any numerically infeasible combinations of decision variables for data-driven optimization prior to the simulation call. In this study, we embed the SVM model in the ARGONAUT algorithm (Boukouvala and Floudas, 2017). The key findings for the SVM model performance and the stochastic steam cracking optimization problem are provided below.

3. Results

3.1. Performance Metrics for the SVM Classifier in the Offline Phase

The results summarizing the predictive performance of the SVM classifier are presented in Table 1. As ARGONAUT operates in sessions (i.e., the first session in the original variable bounds; the second session in the tightened variable bounds), two sets of performance metrics are reported. The results show that highly accurate SVM classifiers are trained using the numerical feasibility information obtained across all scenarios. When the SVM model is retrained within the tightened variable bounds, we observe that the model accuracy, precision, and F_1 score are improved.

Table 1 SVM model performance with the blind testing set.

SVM Model	Accuracy	Precision	Recall	F_1 score	AUC
Session 1	98.5 %	96.9 %	100 %	98.4 %	100 %
Session 2	99.5 %	100 %	99.1 %	99.5 %	100 %

Here, although we achieved highly accurate classifiers, one of the biggest drawbacks of the offline phase is the computational overhead required to collect samples for modeling. Especially with computationally expensive simulations, this step is very demanding. Our future work will focus on making the offline and online phases seamless to improve the computational efficiency of the overall framework.

3.2. Optimal Solution: Reactor Design Parameters and Key Results in the Online Phase

The highly accurate SVM model classifiers are then incorporated in the ARGONAUT algorithm and the stochastic steam cracking model is optimized over 10 random runs, each starting with a different set of Latin Hypercube Design. The optimal values for the decision variables in the best-found solution are: T^{in} = 788.4 K; P^{in} = 341.4 kPa; Q = 350.9 kW/m^2; L = 39.7 m; F^{o}_{NGL} = 0.0289 kmol/s; and $F^{o}_{H_2O}$ = 0.018 kmol/s. This set of optimal decision variables provides a guaranteed feasible solution for the 11 studied scenarios and achieves a total stochastic profit of \$0.1677/s using the hybrid SVM and grey-box optimization approach. This total stochastic profit is almost 50 % less than the profit reported for pure ethane feed, but it is also almost 50 % greater than the profit reported for the pure propane feed in Beykal et al. (2020). This is an expected result as the previous study only considered pure feeds in the problem setup where the reactor parameters are fine tuned to maximize profit for the deterministic NGL feeds. However, in the current study, the optimal reactor parameters can handle a wide range of feed compositions which allows the reactor to be more flexible while achieving high profit values, even when the feed compositions change due to changing market conditions or other external factors.

In addition, Figure 3 shows the production of ethylene, propylene, and other key products under the provided optimal conditions for the two representative scenarios: (1) High ethane content (70 % ethane-30 % propane); and (2) high propane content (20 % ethane and 80 % propane). In both cases, we observe that the molar flowrate of the reactants is decreasing across the reactor length, whereas the molar flowrate of the desired products is increasing through carrying out the favorable reactions. The mean ethane conversion across all scenarios (except for the pure propane scenario) is 0.79 with a standard error of 0.015 across these scenarios. Likewise, the mean propane conversion is also high, 0.88, with a standard error of 0.006. These results indicate that guaranteed feasible solutions

can be achieved for numerically infeasible stochastic problems using data-driven modeling and grey-box optimization.

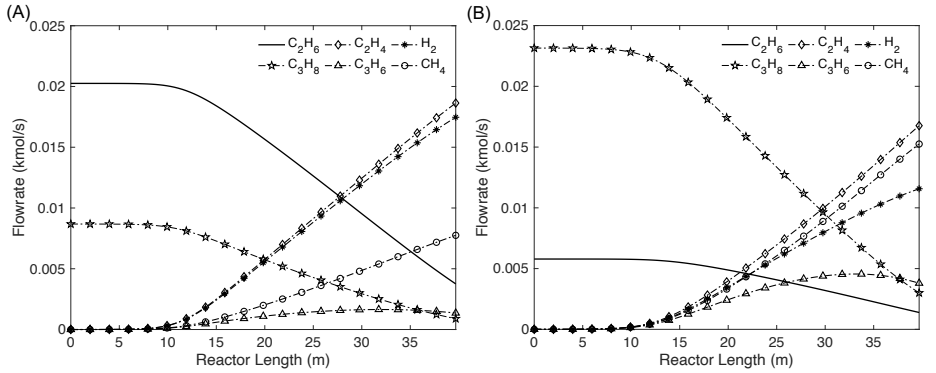


Figure 3 Molar flowrate of species for the favorable reactions in two scenarios with: (A) 70 % ethane – 30 % propane in the NGL feed; (B) 20 % ethane – 80 % propane in the NGL feed.

4. Conclusions

We present a framework to address the data-driven stochastic optimization of numerically infeasible DAE systems without the full discretization of the first-principles model. The stochastic elements in the formulation are handled via scenario analysis whereas the numerical infeasibilities are modeled using Support Vector Machines. By incorporating these two analysis methods in a grey-box optimization solver, we provide guaranteed feasible solutions to numerically infeasible stochastic problems using data-driven modeling. Results of the computational case study of the steam cracking of natural gas liquids with uncertain feed compositions show that high total stochastic profit is achieved using the proposed approach.

References

B. Beykal, F. Boukouvala, C.A. Floudas, N. Sorek, H. Zalavadia, E. Gildin, 2018, Global optimization of grey-box computational systems using surrogate functions and application to highly constrained oil-field operations, Computers & Chemical Engineering, 114, 99-110.

B. Beykal, M. Onel, O. Onel, E.N. Pistikopoulos, 2020, A data-driven optimization algorithm for differential algebraic equations with numerical infeasibilities, AIChE Journal, 66, 10, e16657.

L.T. Biegler, 1984, Solution of dynamic optimization problems by successive quadratic programming and orthogonal collocation, Computers & Chemical Engineering, 8, 3-4, 243-247.

F. Boukouvala, C.A. Floudas, 2017, ARGONAUT: AlgoRithms for Global Optimization of coNstrAined grey-box compUTational problems, Optimization Letters, 11, 5, 895-913.

D.Y. Caballero, L.T. Biegler, R.Guirardello, 2015, Simulation and optimization of the ethane cracking process to produce ethylene. Computer Aided Chemical Engineering, 37, 917-922.

Y.D. Lang, L.T. Biegler, 2007, A software environment for simultaneous dynamic optimization, Computers & Chemical Engineering, 31, 8, 931-942.

M. Onel, C.A. Kieslich, E.N. Pistikopoulos, 2019, A nonlinear support vector machine-based feature selection approach for fault detection and diagnosis: Application to the tennessee eastman process, AIChE Journal, 65, 3, 992-1005.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50264-5

Tensor-Based Autoencoder Models for Hyperspectral Produce Data

Charlotte Cronjaeger[a], Richard C. Pattison[b], Calvin Tsay[a]*

[a]*Department of Computing, Imperial College London, London SW7 2AZ, United Kingdom*
[b]*Apeel Sciences, Santa Barbara, CA 93117, United States of America*

c.tsay@imperial.ac.uk

Abstract

Effectively monitoring and controlling product quality is critical in produce supply chain management. Hyperspectral imaging has emerged as a promising technique for monitoring food products, but the size of hyperspectral datasets complicates storage and processing. This work develops a novel architecture for autoencoder models that is well-suited for nonlinear subspace learning on tensorial, hyperspectral data. In particular, separate submodels are used to (de)compress each mode of the data tensor, preserving spatial locality information and greatly reducing the number of autoencoder parameters. The approach enables memory-efficient training, nonlinear dimensionality reduction, and multi-task learning, as demonstrated by a real-world case study.

Keywords: Tensorial Data, Supply Chain Management, Hyperspectral Images, Dimensionality Reduction, Subspace Learning.

1. Introduction

Degradation of perishable products is highly dependent on storage/transport conditions and represents a considerable challenge to manage; approximately one-third of food produced each year is lost or wasted, at a cost of nearly $1 trillion (USD) (World Food Program, 2020). Modeling and monitoring of product quality and degradation play a vital role in addressing these issues. For instance, degradation models can account for waste in supply chain optimization (Rong et al., 2011; Tsay and Baldea, 2019), while monitoring product quality can provide closed-loop "feedback" (Lejarza and Baldea, 2020). To this end, hyperspectral imaging is a promising technique for food products, bridging spectroscopy and computer vision (i.e., spectral and spatial information). Hyperspectral images can both reveal internal characteristics, such as firmness, dry matter, and sugar content, and detect external contaminants/defects. However, the size of hyperspectral data (often $>$100 MB/image) complicates storage and processing (Feng and Sun, 2012).

Owing to their size, many techniques can be applied to extract features from hyperspectral data (Huang et al., 2014). For example, linear subspace learning techniques such as principal component analysis (PCA) are widely applied for dimensionality reduction and pattern recognition. Analogous nonlinear techniques have since been proposed; autoencoders are often used as a form of nonlinear PCA, as they can be trained using methods tailored for large datasets (Kramer, 1991). However, hyperspectral data involve three *modes*: length, width, and spectral band. Such multi-modal data motivate the use of multilinear subspace

learning (MSL) methods, e.g., Tucker decomposition, which are often more data-efficient, as they preserve/exploit spatial locality information among tensor entries (Lu et al., 2013).

In this work, we propose a novel nonlinear subspace learning technique for tensorial data based on autoencoders (AEs). Specifically, we avoid flattening (vectorizing) the tensors and instead use a separate sub-model to (de)compress each mode of the tensor. Exploiting the intrinsic structure in this manner greatly reduces the number of AE parameters that must be learned. Using a case study of real-world produce data, we show that our novel AE architecture with linear activations can closely match the compression ability of standard MSL approaches, while enabling memory-efficient training using semi-batch gradient descent. Nonlinear activations can further improve compression ability. Finally, as the model architectures are generic, we expand the AEs to include classification within the compressed, latent space (i.e., multi-task learning). In the context of providing feedback in supply chain management, this approach enables attributes to be predicted in online applications without requiring data to be reconstructed and re-processed.

2. Methodology

The goal of our proposed autoencoder (AE) architecture is to learn a low-dimensional, nonlinear manifold underlying tensorial data, while simultaneously sustaining the spatial locality of the data. For this, a residual tensor architecture consisting of N linear AEs and N non-linear AEs for an Nth-order tensor is trained, as shown in Figure 1. It can be shown that the linear AE architecture performs as well as common algebraic linear subspace learning techniques, such as Tucker decomposition and High-Order Singular Value Decomposition (HOSVD), if trained properly. Therefore, we do not simply train a nonlinear AE end-to-end, but rather a nonlinear AE that learns the residuals of a linear AE. This simplifies learning, enables stable training, and improves generalization performance.

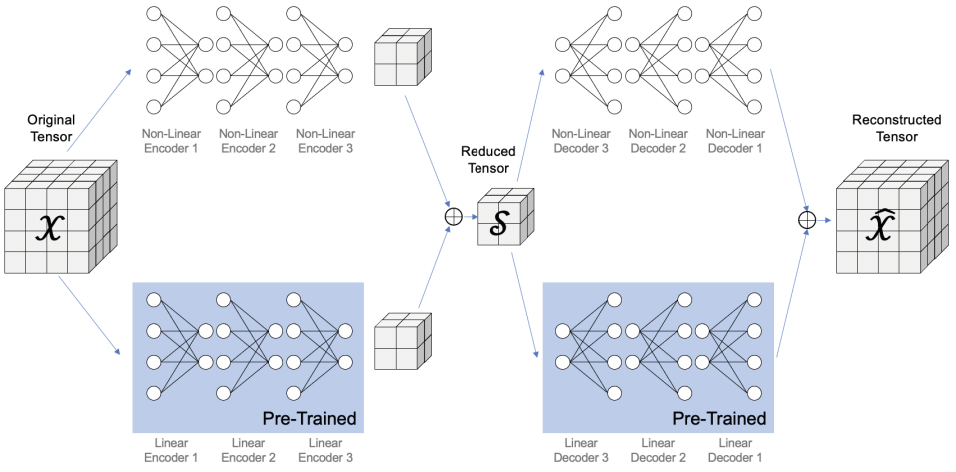


Figure 1: Residual tensor AE architecture for a third-order tensor with (fixed) pre-trained weights of the linear AE and trainable weights of the non-linear AE model.

2.1. Residual Tensor Autoencoder (AE) Architecture

Figure 1 depicts the architecture for a third-order tensor, but the proposed architecture can easily be generalized to Nth-order tensors. Both the linear and nonlinear encoders reduce the input tensor $\mathcal{X}$ to the desired subspace dimension. The reduced tensors are added together, resulting in a combined tensor $\mathcal{S}$, and subsequently fed into both the linear and non-linear decoders. The reconstructed tensor $\hat{\mathcal{X}}$ results from the addition of the outputs of the linear and non-linear decoders. We employ a mean squared error (MSE) loss function that minimizes the l_2-norm of the error between $\hat{\mathcal{X}}$ and $\mathcal{X}$, i.e., $\text{MSE}(\mathcal{X}, \hat{\mathcal{X}}) := ||\mathcal{X} - \hat{\mathcal{X}}||_2^2$.

The weights of the linear AEs are pre-trained and fixed while training the nonlinear components. Our proposed architecture comprises N autoencoders: each reduces one dimension of an Nth-order tensor. The AE pairs encode "fibers" of the tensor independently. Each AE slices the tensor into its different fibers and feeds the fibers (i.e., vector data) into fully connected encoder and decoder models. Figure 2 depicts how an encoder model slices input tensor $\mathcal{X}$ into its component fibers to reduce dimensions sequentially. The decoder model works identically. Note that the number of data samples fed into the fully connected encoders and decoders decreases sequentially, as previous dimensions are already reduced, and fibers are processed independently by the AEs. This results in fewer data samples to be trained on by some AEs. However, we note that slicing the tensors creates more "samples" in the first place and hypothesize that, since these processed samples have fewer correlated dimensions, the effective dataset size remains similar.

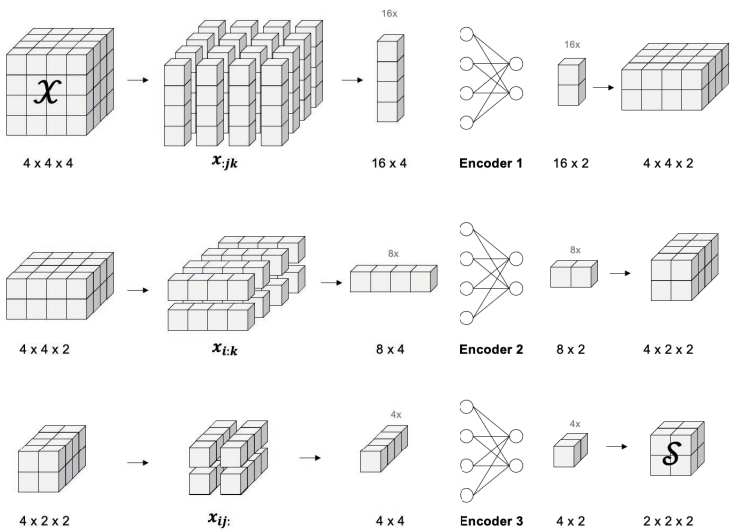


Figure 2: Encoder models for a third-order tensor which slice each tensor into its different fibers along each mode to reduce all three dimensions in sequence.

2.2. Non-Linear Architectures of the Fully Connected Autoencoder

For each of the N fully connected encoders and decoders, a simple nonlinear architecture is applied. In particular, each encoder and decoder comprises two fully connected layers.

After the first layer, a nonlinear activation function is applied, namely the Rectified Linear Unit (ReLU). While multiple extensions to this simple nonlinear architecture are possible, e.g., deeper architectures, batch normalization, or residual skip connections, none of these led to significantly better results in our experiments. This can be explained through the resulting depth—and hence number of trainable parameters—between input and output signals, which makes the non-convex loss function more difficult to train. For a third-order tensor with only two layers for each encoder and decoder, there are already $3 \times 2 \times 2 = 12$ nonlinear layers between input tensor $\mathcal{X}$ and reconstructed tensor $\hat{\mathcal{X}}$.

2.3. Multi-Task Learning: Classification of Reduced Tensors

One advantage of the proposed AE architecture, in addition to its trainability via gradient-descent-based optimization, is the potential to learn multiple tasks simultaneously and in an end-to-end fashion, i.e., multi-task learning. In this case, a classifier is learned in addition to minimizing the reconstruction loss when training the model. The loss function is a linear combination of the MSE loss (as described above) and the cross entropy loss for classification. The goal is to classify each tensor based on its compressed representation and, in particular, to learn a representation of tensors that yields good classification performance. For this, in addition to the existing architecture, a classification model on the compressed tensor is trained, which can be simultaneously learned using backpropagation.

3. Experimental Results

A real-world hyperspectral image dataset with 186 tensor data samples of avocados is used for the experiments. All data samples are cropped to an equal size of $236 \times 187 \times 224$. The avocados are imaged in three-by-four trays, but each is stored independently. Ambient lighting differs among the four tray columns (far left, middle left, middle right, far right). Therefore, given the confidentiality of industrial data, we consider predicting the column an avocado was in during imaging as a simple, yet representative, classification task.

We compare performance of the proposed model against Tucker decomposition as a benchmark. Tucker decomposition can be seen as a multilinear extension of PCA to higher order tensor data (Lu et al., 2013). As we do not operate on a single tensor, i.e., how Tucker decomposition is conventionally denoted, but rather desire to learn a linear manifold for a set of K tensors, we optimize:

$$\min_{U(n)} \frac{1}{K} \sum_{k=1}^{K} \|\mathcal{X}_k - \mathcal{S}_k \times_1 U_1 \times_2 U_2 \cdots \times_N U_N\|_F^2 , \tag{1}$$

where $\mathcal{X}$ is the original tensor, $\mathcal{S}$ the reduced tensor, and U_N the Nth projection matrix. A low-rank approximation can be found via High Order Orthogonal Iteration (HOOI) (Sheehan and Saad, 2007).

Residual Tensor Architecture. We compare the performance of the proposed AE architecture to Tucker decomposition for several different subspace dimensions. For simplicity of comparison, the tensors are always reduced from their original size of $236 \times 187 \times 224$ to cubes with subspace dimensions between $1 \times 1 \times 1$ up to $40 \times 40 \times 40$. As we observe in Table 1, the proposed model outperforms Tucker decomposition, especially for small

Table 1: MSE reconstruction loss for residual tensor learning for various cube subspace dimensions using two-layer nonlinear encoder and decoder models and training for 100 epochs. Training times reported for a single GPU NVIDIA GeForce GTX TITAN X.

Dim.	Train Loss / 10^{-3}	Test Loss / 10^{-3}	%-Change v Tucker Test Loss	Training Time	Reduction in Size
1	7.088	8.386	-44.03%	23 min	99.99%
5	1.034	1.427	-34.00%	24 min	99.99%
10	0.918	1.138	+18.54%	20 min	99.98%
15	0.588	0.602	-2.90%	19 min	99.96%
20	0.400	0.424	-2.08%	19 min	99.91%
25	0.300	0.320	+4.92%	19 min	99.84%
30	0.240	0.236	+10.28%	20 min	99.72%
35	0.175	0.192	+22.29%	22 min	99.56%
40	0.139	0.149	+24.17%	22 min	99.35%

Table 2: MSE Reconstruction Loss & Cross Entropy Classification loss for simultaneous residual tensor learning and classification tasks for various cube subspace dimensions using two-layer nonlinear encoder and decoder models and training for 100 Epochs. Training times reported for a single GPU NVIDIA GeForce GTX TITAN X.

Dim.	Test Loss /10^{-3}	%-Change v Tucker Test Loss	Classifier Test Loss /10^{-3}	# Correctly classified (Test)	Training Time
1	7.871	-47.47%	1333.622	29/36	16 min
5	1.441	-33.33%	50.055	35/36	17 min
10	1.824	+90.00%	5.757	36/36	17 min
20	0.449	+3.68%	0.355	36/36	17 min
30	0.241	+12.76%	0.324	36/36	18 min
40	0.144	+19.73%	0.916	36/36	21 min

subspace dimensions up to $20 \times 20 \times 20$. For subspace sizes below $5 \times 5 \times 5$ the model reduces the reconstruction error on the test dataset by over 30%. The fact that the proposed architecture performs better for smaller subspace dimensions may be due to having fewer trainable parameters to optimize, given the highly non-convex loss function.

Simultaneous Classification. When training a classifier on the compressed tensor simultaneously with learning the lower dimensional manifold, the model is able to classify over 97% data samples of the test dataset correctly for subspace sizes of $5 \times 5 \times 5$ and larger. Even when reducing each tensor to a $1 \times 1 \times 1$ scalar, the model classifies $> 80\%$ of test data correctly. The reconstruction loss remains comparable to the results for models when only learning the lower dimensional manifold. The results are presented in Table 2.

4. Conclusions

In this work, we introduced a novel AE architecture for hyperspectral produce data. We demonstrated that the proposed residual tensor architecture outperforms existing subspace

learning techniques, especially for smaller subspace dimensions, where our model can reduce the reconstruction loss compared to Tucker decomposition by over 30%. Furthermore, we showed that multi-task learning, i.e., including a classification task on the reduced tensors, is possible and promising. Compared to existing subspace learning techniques our model is more scalable, as it can be efficiently trained using stochastic gradient descent and involves fewer parameters compared to other AE-based methods.

Future work can involve more effective uses of spatial locality in tensorial data. Our proposed method assumes the vector input of each AE (the fibers) to be drawn as independent samples of a distribution. However, in the case of hyperspectral images, nearby pixels are highly correlated, and the data exhibits strong spatial structure as a result. The sheer size of hyperspectral images, however, may prohibit directly employing standard approaches, such as convolutional layers. Given the current optimization difficulties for higher dimensional subspaces, future work can also explore alternative gradient-based algorithms for training, as well as gradient-free methods. A final interesting direction of future research could investigate the impact of the order in which the dimensions are reduced, as this impacts the numbers of samples each AE is trained on.

5. Acknowledgements

The authors gratefully acknowledge support from the Engineering & Physical Sciences Research Council (EPSRC) through fellowship grant EP/T001577/1, Apeel Sciences, and an Imperial College Research Fellowship to CT.

References

Feng, Y.Z., and Sun, D.W., Application of hyperspectral imaging in food safety inspection and control: a review. *Crit. Rev. Food Sci. Nutr.*, 52(11), 1039-1058, 2012.

Huang, H., Liu, L., and Ngadi, M.O., Recent developments in hyperspectral imaging for assessment of food quality and safety. *Sensors*, 14(4), 7248-7276, 2014.

Kakarla, S., Gangula, P., Singh, C., and Sarma, T. H., Dimensionality Reduction in Hyperspectral Images Using Auto-encoders. In *ICACCI 2019*, 101-107, 2019

Kramer, M.A., Nonlinear principal component analysis using autoassociative neural networks. *AIChE J.*, 37(2), 233-243, 1991).

Lejarza, F., and Baldea, M., Closed-loop optimal operational planning of supply chains with fast product quality dynamics. *Comput. Chem. Eng.*, 132, 106594, 2020.

Lu, H., Plataniotis, K.N., and Venetsanopoulos, A., *Multilinear subspace learning: dimensionality reduction of multidimensional data*, CRC Press, 2013.

Rong, A., Akkerman, R., and Grunow, M., An optimization approach for managing fresh food quality throughout the supply chain. *Int. J. Prod. Econ.*, 131(1), 421-429, 2011.

Sheehan, B.N., and Saad, Y., Higher order orthogonal iteration of tensors (HOOI) and its relation to PCA and GLRAM. *SIAM Int. Conf. Data Min.*, 355-365, 2007.

Tsay, C., and Baldea, M., 110th anniversary: Using data to bridge the time and length scales of process systems. *Ind. Eng. Chem. Res.*, 58(36), 16696-16708, 2019.

World Food Programme (WFP), Zero Hunger. `https://www.wfp.org/zero-hunger`, 2020.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50265-7

Molecular Representations in Deep-Learning Models for Chemical Property Prediction

Adem R.N. Aouichaoui, Fan Fan, Seyed Soheil Mansouri, Jens Abildskov
Gürkan Sin*

Process and Systems Engineering Center (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Kgs.Lyngby, DK-2800, Denmark
gsi@kt.dtu.dk

Abstract

A molecular property prediction model is dependent on the interplay between the quality of data, and expressive representation (or descriptor), and a suitable algorithm to relate the descriptors to the target property. In this work, a deep neural network (DNN) is used to regress two types of descriptors: fixed descriptors (Group fragments and Morgan fingerprints) and learned descriptors (from a Graph Neural Network, GNN). Bayesian optimization was used for hyperparameter tuning and a set of 5 models were benchmarked and used to predict the enthalpy of formation of organic compounds. GNN based models provided the best overall results compared to descriptor-based models which the attentive fingerprint model that combines RNN and graph attention mechanism (AFP) achieved the best results of 5.9 kJ/mol mean absolute deviation and a coefficient of determination of 0.99 in the training, validation, and test set. Despite not achieving chemical accuracy of 4 kJ/mol, the model has shown great promise in distinguishing between isomers and provides a baseline for future improvements to achieve chemical accuracy.

Keywords: Deep-Learning, Molecular Property Prediction, Enthalpy of Formation

1. Introduction

The molecular properties of chemical compounds must be known *a priori* to execute many chemical engineering applications such as risk assessment, P-V-T calculation using equations of states as well as material selection. These properties are either the direct product or derived quantities from experimental measurements. Conducting such measurements on demand whenever the need arises is not a viable option due to time and expenses. Predictive models capable of describing these properties provide an attractive alternative to quickly evaluate the properties of a compound. Quantitative structure-property relation (QSPR) models are predictive models relating the chemical structure of a compound to a target property. The molecular structure is converted into a numerical representation that is then used as input to a mathematical model to produce the target property. The mathematical model is usually selected by observing the trend of the property with increasing carbon numbers for homologous series. Recently, deep neural networks (DNN) have gained popularity in many engineering applications and have also been used as part of the QSPR model(Aouichaoui *et al.*, 2021). This increasing attention is due to their ability to approximate any non-linear functions (universal approximation theory). Despite being an integral part of any QSPR, the selection of the mathematical model is less challenging than developing the molecular representation or descriptors used as input, which remains an issue and a detriment factor to the success of the model

(good accuracy and ability to distinguish compounds) to a higher degree than the mathematical model. The molecular representation used can either be the product or a combination of domain knowledge, heuristics, or a data-driven approach. Group-contribution models (GC) employ both domain knowledge and heuristics to represent the molecule as an occurrence vector of a set of predefined groups(Hukkerikar *et al.*, 2012). The Morgan fingerprints and it's variation such as the extended connectivity fingerprints (ECFP) are another widely popular descriptor in the fields of cheminformatics and drug discovery (Rogers and Hahn, 2010). The ECFPs are circular fingerprints that represent the presence of a particular substructure that is encoded through a hashing function. The circular fingerprints can be generated for different diameters by combining features from the previous diameter length that are stored in a variable-length bit vector. Various QSPR models have used the ECFP to model a variety of molecular properties such as predicting water solubility (Xiong *et al.*, 2020). The group fragments and the ECFP are fixed descriptors that are proper to the molecule chosen and known before the modeling process. However, the molecular descriptors can also be learned so they become not only property to the molecule but also exclusive for the dataset through graph neural networks (GNNs) (Gilmer *et al.*, 2017). These models take a graph representation of the molecule where the nodes represent the atoms and the edges represent the bonds. Each node and edge is assigned a feature vector with information related to the atoms (type, valency, etc) and the bonds (type of bonds, etc.). These feature vectors are then updated based on the information contained in the feature vector of their neighboring nodes by applying graph convolutions or message passing layers. A readout function is then applied to the graph representation to produce a vector representation that is then supplied to a DNN to produce the target prediction. The feature update and the repression procedures both employ a series of algorithms that integrates learnable parameters that are adjusted using error backpropagation, which produces a representation that fits the compound and the target property at the same time. In this work, we benchmark the above-presented descriptors using the same property data, to compare their performance and highlight some of their advantages and drawbacks.

2. Methods

2.1. Models

We distinguish between two types of models, those that use a fixed representation in the form of the group fragments and ECFP and those that generate their representation from a molecular graph representation.

2.1.1. Fixed representation models

Models with a fixed molecular representation are used in conjunction with a DNN to correlate the descriptor to the target property. The GC-DNN uses the group fragmentation developed by (Hukkerikar *et al.*, 2012), where the molecule is described through 3 levels (orders) with increasing levels of complexity containing: 224 first-order groups, 134-second order groups, and 74 third-order groups. Although third-level groups are based on convenience and a more heuristic approach, they are included in this study to take advantage of the full predictive power of GC-based methods. The ECFP-DNN uses the extended circular fingerprints generated through the Morgan algorithm as described in (Rogers and Hahn, 2010). These descriptors are generated using a Python-based cheminformatics package RDKit (Landrum, 2020). The representation is hashed into a bit vector of length 1024 as used in previous studies (Xiong *et al.*, 2020). The hyper-parameters were optimized: dimensions of hidden neurons in the first layer [256, 1024],

number of layers [2, 4], activation functions ['LeakyReLU', 'Sigmoid', 'Tanh', 'SELU'], L2 regularization [0, 0.05], initial learning rate [1e-5, 1e-1], learning rate reduce factor [0.2, 0.8]. Note that the size of hidden neurons in the following layers is designed to be half of the previous layer.

2.1.2. Adaptive/learned representation models

Three graph neural networks are used to evaluate the performance of models with an adaptive representation of the molecule. In the following, the main features of the models are highlighted. For a more in-depth explanation of the models and their hyperparameters, the reader is encouraged to inspect the references provided.

The Message Passing Neural Network (MPNN) by (Gilmer *et al.*, 2017) is a versatile model used for various property prediction purposes such as predicting water solubility. The model takes an undirected molecular graph with attributed nodes and edges. The operation of the models is described in two phases: a message-passing phase where the node and edge features are transmitted to the neighboring nodes and used to update its representation and a readout phase transforming the graph representation to a vector representation that is supplied to a DNN to regress the target property. More details on mathematics can be found in (Gilmer *et al.*, 2017). The main hyper-parameters were optimized: hidden dimensions [1, 128], number of layers [1, 4], L2 regularization [0, 0.05], learning rate [1e-5, 1e-1], learning rate reduce factor [0.2, 0.8].

Graph Isomorphism Network (GIN) by (Xu *et al.*, 2019) is a simple GNN that is intended to achieve a similar ability to the Weisfeiler-Lehman graph isomorphism test. The model only relies on node features to aggregate and update the node feature through a deep neural network and uses the sum function and concatenating the resulting representation from each iteration as the readout function. The main hyper-parameters for tuning together with their search domains are the same as MPNN.

Attentive Fingerprint (AFP) by (Xiong *et al.*, 2020) is considered the state-of-the-art GNN model that combines a series of deep-learning techniques to enhance its representative capabilities. A recursive neural network is used to agglomerate the messages from nearby and distant nodes in addition to the graph attention mechanism that allows the model to weigh the information and assign importance to it thus only focusing on the relevant structural information. Besides those hyper-parameters mentioned above, Attentive FP has one extra hyper-parameter to be optimized: the number of time steps [1, 4].

All previously described GNN models operate on a graph representation of the molecule. The node is attributed with the atom type (C, N, F, Br, Cl, S, I), the atomic mass, the atom degrees (nr. of covalent bonds), the type of hybridization (sp,sp^2,sp^3,sp^3d,sp^3d^2), whether the atom is part of an aromatic configuration, whether it is part of a ring structure as well as the number of hydrogen attached and whether it is a chiral center. The edges contain information on whether the bond is single, double, triple, or part of an aromatic structure. The feature vector also includes whether the bond is conjugated or part of a ring structure as well as a stereo-configuration it might be part of (E/Z, cis/trans).

2.2. Training & Optimization

The data are split into three folds: 90% for training, 5% for validation, and 5% for testing. Training has been prioritized since the dataset size is small compared to other deep-learning applications such as image or speech recognition. During training, the adaptive learning rate is used to adapt to the optimization surface as well as early stopping and L2 weight regularization to avoid overfitting the objective function used for model training

is the mean-squared error. The hyperparameters of the various models have been tuned using a multi-objective Bayesian optimization (MOBO) toolbox (Galuzio *et al.*, 2020) The root-mean-squared error (RMSE) of the training and validation set are chosen to describe the performance of the model. The hyper-parameters of each model and their range have been previously described when presenting the models. The MOBO is done by constructing a posterior distribution function using a Gaussian process using the Matén covariance function as shown in Eq.(1), where "l" is the length scale, Γ() is the gamma function, "Kv()" is the modified Bassel function and "r" is the distance between two arguments of Kernel and "v" is a positive parameter set to 1.5 in this study.

$$C_v(\vec{x},\vec{x}') = \frac{2^{1-v}}{\Gamma(v)}\left(\frac{r\sqrt{2v}}{l}\right)^{v} K_v\left(\frac{r\sqrt{2v}}{l}\right) \quad (1)$$

3. Case Study: Predicting the enthalpy of formation of organic compounds

3.1. Property Data

The methodology model described previously is applied to predict the standard enthalpy of formation (HFOR) of organic compounds. The HFOR is defined as the change in enthalpy associated with the reaction forming the given chemical in its standard state from the elements in its standard state. The data were collected from the DIPPR database (Wilding *et al.*, 2017) containing a total of 741 compounds. Figure 1 shows the distribution and range of the data. Only experimental values are selected and only organic compounds with either of the following atoms: Oxygen (O), Bromide (Br), fluorine (F), chlorine (Cl), Iodine (I), Nitrogen (N), and Sulfur (S). An important challenge associated with the enthalpy of formation is to achieve " chemical accuracy" which is stated as 4 kJ/mol (Meier, 2021).

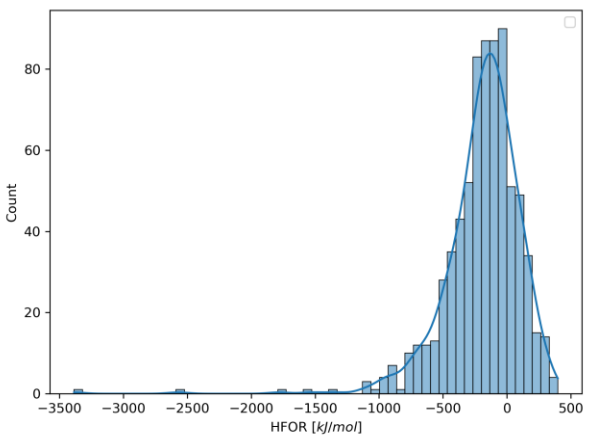


Figure 1: distribution of the heat of formation

3.2. Results

Table 1 provides the number of learned parameters (N), the coefficient of determination (R^2), the mean absolute error (MAE), and the mean absolute percentage error (MAPE) of the best-performing type of model. The best metrics are in bold.

Table 1: Perfoamnce metrics of the tuned models

Model	N	R^2	MAE(kJ/mol)	MAPE (%)
ECFP-DNN	1,494,510	0.977-0.916-0.944	26.0-74.9-60.5	27.4-80.9-56.7
GC-DNN	946,297	0.996-0.991-0.987	11.8-16.2-16.3	16.1-19.7-16.0
MPNN	557,665	0.997-0.973-0.981	6.7-28.7-31.1	10.5-33.6-26.8
GIN	123,481	0.994-0.992-0.991	12.9-34.8-16.2	23.3-29.8-43.5
AFP	97,995	**0.999-0.999-0.997**	**5.1-7.9-12.6**	**6.4-13.4-12.8**
GC+	428	0.999	1.75	-

3.3. Discussion

The AFP model outperforms all other GNN models as well as the models based on GC and ECFP across all metrics with an overall R^2 of 0.999, an MAE of 5.9 kJ/mol, and a MAPE of 8%. Compared to the model developed by (Hukkerikar *et al.*, 2013) (GC^+) the AFP falls short of achieving the desired chemical accuracy of 2 kJ/mol. However, it is important to note that this model uses all data to perform the regression and even defined new additional higher order that is not funded in any chemical or property knowledge to reduce the error and accomplish the chemical accuracy. This raises a concern about the models' ability to extrapolate. The GC-DNN also falls short of the model by (Hukkerikar *et al.*, 2013) across all metrics. The reason is not all data are used for regression and the fact that it did not rely on all groups defined by (Hukkerikar *et al.*, 2013). Compared to classical QSPR models, GNN and in general deep-learning-based models are high parametric models with the number of parameters much higher than the number of data points available. Despite this, AFP is the GNN with the least model parameters and still outperforms the remaining models. This could be related to the attention mechanism and use of RNN to process node features.

Table 2: Prediction and experimental value for Methylpentenes. The absolute relative error compared to the experimental value in % are given in parenthesis

	2-METHYL-1-PENTENE	3-METHYL-1-PENTENE	2-METHYL-2-PENTENE
Experimental	-59.2 kJ/mol	-49.4 kJmol	-66.8 kJ/mol
GC-DNN	-57.0 kJ/mol (03.7)	-40.9 kJ/mol (17.2)	-59.4 kJ/mol (11.1)
ECFP-DNN	-48.6 kJ/mol (17.9)	-20.8 kJ/mol (57.9)	-87.5 kJ/mol (30.9)
MPNN	-57.8 kJ/mol (02.3)	-50.7 kJ/mol (02.6)	-62.6 kJ/mol (06.3)
GIN	-50.7 kJ/mol (14.4)	-46.2 kJ/mol (06.5)	-63.4 kJ/mol (05.1)
AFP	-57.9 kJ/mol (02.2)	-49.0 kJ/mol (00.8)	-65.2 kJ/mol (02.4)
GC+	-54.6 kJ/mol (07.8)	-45.7 kJ/mol (07.5)	-64.7 kJ/mol (03.1)

Table 2 contains the experimental value and corresponding predictions using various models of the methylpeneten isomers. The table clearly shows that the AFP model is superior in distinguishing between isomers compared to the rest of the models with the lowest absolute relative errors across all models. Although a rare occasion, the GC^+ model can distinguish the isomers presented herein due to the presence of second-order groups.

Another interesting aspect is the fact that no prior knowledge is incorporated into the GNN models other than basic chemistry-related information, and despite this, the model achieves very promising results.

4. Conclusions

A successful QSPR model heavily relies on the interplay between data, representation, and model. The focus of this work was on some of the methodologies to represent the molecular structure in a machine-readable way. The data and the general prediction model were identical. GNN based models showed superior performance to descriptor-based models with the AFP model achieving the best results with 5.9 kJ/mol mean absolute error, and although it falls short of the target chemical accuracy (also referred to as the holy grail), the model has shown it is much better at distinguishing between isomers. Furthermore, the results suggest that deep-learning-based models such as GNN do provide a powerful tool to correlate molecular structure to the desired target property without time-consuming descriptor design or extensive domain knowledge.

References

Aouichaoui, A.R.N., Al, R., Abildskov, J. and Sin, G. (2021) 'Comparison of Group-Contribution and Machine Learning-based Property Prediction Models with Uncertainty Quantification', in *Computer Aided Chemical Engineering*. Elsevier Masson SAS, pp. 755–760.

Galuzio, P.P., de Vasconcelos Segundo, E.H., Coelho, L. dos S. and Mariani, V.C. (2020) 'MOBOpt — multi-objective Bayesian optimization', *SoftwareX*, 12.

Gilmer, J., Schoenholz, S.S., Riley, P.F., Vinyals, O. and Dahl, G.E. (2017) 'Neural Message Passing for Quantum Chemistry', *34th International Conference on Machine Learning, ICML 2017*, 3, pp. 2053–2070. Available at: http://arxiv.org/abs/1704.01212.

Hukkerikar, A.S., Meier, R.J., Sin, G. and Gani, R. (2013) 'A method to estimate the enthalpy of formation of organic compounds with chemical accuracy', *Fluid Phase Equilibria*, 348(1), pp. 23–32.

Hukkerikar, A.S., Sarup, B., Ten Kate, A., Abildskov, J., Sin, G. and Gani, R. (2012) 'Group-contribution + (GC +) based estimation of properties of pure components: Improved property estimation and uncertainty analysis', *Fluid Phase Equilibria*, 321, pp. 25–43.

Landrum, G. (no date) 'RDKit: Open-source cheminformatics'.

Meier, R.J. (2021) 'Group Contribution Revisited: The Enthalpy of Formation of Organic Compounds with "Chemical Accuracy"', *ChemEngineering*, 5(2).

Rogers, D. and Hahn, M. (2010) 'Extended-connectivity fingerprints', *Journal of Chemical Information and Modeling*, 50(5), pp. 742–754.

Wilding, W.V., Knotts, T.A., Giles, N.F. and Rowley, R.L. (2017) 'DIPPR® Data Compilation of Pure Chemical Properties', *Design Institute for Physical Properties, AIChE* [Preprint].

Xiong, Z., Wang, D., Liu, X., Zhong, F., Wan, X., Li, X., Li, Z., Luo, X., *et al.* (2020) 'Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism', *Journal of Medicinal Chemistry*, 63(16), pp. 8749–8760.

Xu, K., Hu, W., Leskovec, J. and Jegelka, S. (2019) 'How Powerful are GNN', *Int. Conf. on Learning Representations*, pp. 1–17.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50266-9

Deep Reinforcement Learning for Continuous Process Scheduling with Storage, Day-Ahead Pricing and Demand Uncertainty

Gustavo Campos[a], Simge Yildiz[a], Nael H. El-Farra[a,*], Ahmet Palazoglu[a]

[a]*Department of Chemical Engineering, University of California, Davis, CA 95616, USA*

nhelfarra@ucdavis.edu

Abstract

In this work, we evaluate the application of a Deep Reinforcement Learning (DRL) method for the scheduling of continuous process/energy systems under day-ahead electricity rate and demand forecast uncertainty. We employ the Soft Actor Critic (SAC) method, a stochastic, off-policy, actor-critic method with built-in entropy maximization that balances exploration and exploitation. We choose as a case study the dispatching of energy systems with storage, which can be posed as a continuous scheduling problem. Results from the computational case study demonstrate that the DRL agent is able to surpass a heuristic policy using very little data, and ultimately reaches a performance comparable to a model predictive control (MPC) solution. The effect of demand forecast uncertainty is further analysed and it is shown that, while the MPC performance degrades steadily as the forecast error and recalculation period increase, the DRL method exhibits a more robust performance.

Keywords: Deep Reinforcement Learning; Operation; Optimization; Energy Systems; Demand Response.

1. Introduction

Data-based methods for optimization and control have been gaining traction over recent years due to advances in the field of Deep Reinforcement Learning (DRL), which merges the power of nonlinear approximators with strategies for online exploration, parameter estimation and optimization. Along with this trend, there has also been an increased effort to reduce carbon emissions and increase energy efficiency and renewable penetration. In this context, establishing an efficient operation of energy-intensive processes becomes a critical component for achieving these goals. The operation of these systems, particularly under time-varying electricity rates, has been typically approached using model-based optimization. For instance, in the case of district cooling plants that produce cooling utilities by running electricity-driven industrial chillers, previous works include Economic Model Predictive Control (Ma *et al.*, 2011) and closed-loop scheduling (Risbeck *et al.*, 2017; Campos *et al.*, 2021). While powerful, these methods require considerable modelling effort, which is exacerbated when attempting to model the effects of uncertainty (e.g., in stochastic programming or robust optimization). In this case, DRL emerges as a promising alternative for performing the operation of complex (e.g.,

nonlinear, stochastic, multiscale) systems while avoiding the high associated modelling and computational costs (Badgwell *et al.,* 2018).

In the Process Systems Engineering (PSE) literature, studies using classical data-based techniques such as Reinforcement Learning (e.g., Cassol *et al*., 2018) or Approximate Dynamic Programming (e.g., Lee and Wong, 2010) have been regularly proposed over the past few decades. However, the recent developments in DRL and Deep Learning, more specifically over the last 5 years, have enabled more powerful applications with high dimensional state and action spaces, and policy complexity. DRL techniques have been recently employed in a few process systems applications, including chemical production scheduling using Advantage Actor-Critic (A2C) (Hubbs et al., 2020), control of liquid-liquid extraction columns in biopharmaceutical processes using Deep Q-Networks (DQN) (Hwangbo and Sin, 2020), and control of batch polymerization processes using Deep Deterministic Policy Gradient (DDPG) (Yoo et al., 2021).

While the performance of DRL methods for more traditional problems has been addressed, less attention has been paid to the demand responsive operation of processes. In addition, as opposed to the majority of works that employ deterministic agents (e.g., DDPG, TD3 and DQN), in this paper we employ a stochastic agent using the Soft Actor-Critic (SAC) method, which generates actions following a probability distribution and has been shown to be more robust and have lower brittleness to hyperparameters. Compared to other stochastic methods such as A2C (or A3C), SAC has the advantage of being an off-policy method, an important property for practical applications that allows the use of a replay memory buffer for reducing sample complexity. We provide insights into the applicability of SAC for demand response through a case-study that demonstrates how the agent's performance compares to heuristic and optimal policies with and without forecast uncertainty.

2. Soft Actor-Critic (SAC) Method

The SAC formulation (Haarnoja et al., 2018a, b) aims to learn a policy that maximizes the expected sum of rewards (traditional RL objective), while simultaneously maximizing the policy's entropy or stochasticity. The entropy maximization encourages exploration of the state space and avoids local optima, in addition to other practical advantages (e.g., reducing hyperparameter sensitivity). The entropy augmented objective is posed as $\mathbb{E}_{(s_t,a_t)}\left[r(s_t,a_t)+\alpha\, H\left(\pi(\cdot\,|s_t)\right)\right]$, where α is a temperature parameter that controls the relative importance of the policy's entropy function $H(\pi)$. By applying a soft policy iteration procedure using function approximators (for the Q-function and the policy) and stochastic gradient descent, one can derive equations for the loss functions (minimization objectives) of the actor L_π, critic L_Q and alpha L_α (its logarithm), which are given as follows.

$$L_\pi = -\left(Q_\theta(s_t\,,a_t) - \alpha \log\, \pi_\phi(a_t|s_t)\,\right) \tag{1}$$

$$L_Q = MSE\,\left(Q_\theta(s_t\,,a_t),\; r_t + \gamma\left(Q_{\underline{\theta}}(s_{t+1}\,,a_{t+1}) - \alpha \log \pi_{\underline{\phi}}(a_{t+1}|s_{t+1})\right)\right) \tag{2}$$

$$L_{\log\alpha} = \log\alpha\;\left(\log \pi_\phi(a_t|s_t)\;+\hat{H}\right) \tag{3}$$

The policy update maximizes the Q-function, corrected by a term that steers the policy away from actions with high probability. The critic loss is a mean square error function

between the current critic prediction and a target value, the latter calculated using target networks $Q_{\underline{\theta}}$ and $\pi_{\underline{\phi}}$, which are updated using a Polyak rule. Double estimation of the Q-function (i.e., two critic networks) is employed to mitigate positive bias. A condensed version of the algorithm is presented in Table 1, for a complete version the reader is referred to Haarnoja *et al.* (2018b).

Table 1 – Soft Actor Critic Algorithm (condensed).

Initialize network parameters, hyperparameters, replay buffer and environment.
for *each environment time step,* ***do****:*
    Collect experience (choose a*, observe* r *and* s'*) and store in memory buffer.*
    for *each training step,* ***do****:*
        Sample minibatch of experiences.
        Calculate losses: L_Q *(Eq. (1)),* L_π *(Eq. (2)),* L_α *(Eq. (3)).*
        Calculate gradients (automated with PyTorch).
        Update networks using the Adam stochastic optimizer.
        Perform a Polyak update of the target actor and critics.
    Update state $s_t \leftarrow s_{t+1}$

The information flow through the actor and critic networks is shown in Figure 1. The yellow area represents the action calculation phase, while the blue area represents the training phase, in which the gradient of the parameterized Q-function with respect to the networks parameters is calculated. Reparameterization of the action sampling step explicitly with respect to the distribution parameters using a Gaussian white noise ϵ is employed to allow direct backpropagation of the gradient.

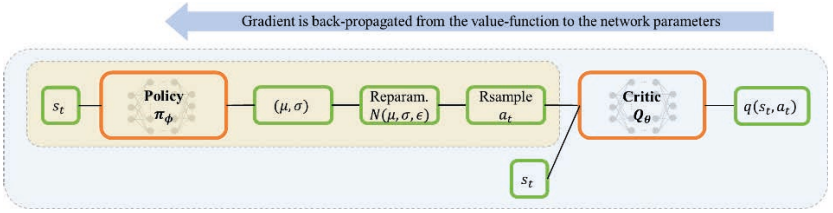


Figure 1. Information flow through networks for action generation and training phases.

3. System Formulation and Description

Markov Decision Process Formulation. We formulate the problem of equipment and storage dispatching under day-ahead electricity prices as a Markov Decision Process (MDP). Our goal is to achieve a minimalistic description that is suitable for the application of DRL methods. We consider the classical MDP framework defined by the tuple (S, A, p, R), i.e., the state and action (continuous) spaces, the unknown state transition probability, and the bounded reward function. The state space includes the current storage value, future demands and future prices. The action space is defined as the current production level (i.e., equipment loads). The reward function is the negative of cost, including both operational cost (i.e., electricity consumption) and constraint violation penalties.

Scheduling of Energy Storage Systems. A typical on-line scheduling formulation is used to obtain the model-based solution. This problem type includes many practical applications such as the dispatching of distributed energy resources (e.g., solar PV and battery systems) and utility production facilities (e.g., district cooling and heating plants). A diagram of the system is presented in Figure 2. For the specific case-study, we consider

the scheduling of district cooling plants, in which the production level represents the cooling load of the chillers, the storage represents a Thermal Energy Storage (TES) tank, the power consumption is a linear function of the production level, the demand corresponds to cooling demand from buildings, and the electricity price varies hourly according to a wholesale day-ahead program.

Figure 2. Energy/process system with production plant, storage and demand.

$$\min \sum_{t \in T} Power_t \, Price_t \tag{4}$$

$$\text{s.t.} \quad Storage_{t+1} = Storage_t + Production_t - Demand_t \tag{5}$$

$$Power_t = f(Production_t) \tag{6}$$

$$\text{hard bounds for } Production_t, \text{ soft bounds for } Storage_t \tag{7}$$

4. Case Study Results

The algorithm was implemented in Python 3.7 using the PyTorch v1.6.0 package. The following settings were used: Adam optimizer with learning rates 1e-4 (actor) and 5e-4 (critic/alpha); networks with two layers, 256 neurons each, and ReLU activation functions; target network Polyak coefficient τ = 0.005; discount factor γ = 0.99; buffer size = 1e6; samples per minibatch = 256; training steps/environment step = 3; weights/biases initialized uniformly $\in (-\mu, \mu)$, μ = (#weights)$^{-0.5}$; log_std output of the policy network $\in$ (-10, 2); single continuous action; 24 forecast steps (state vector with 49 elements); episodic environment using one week of real hourly data (non-episodic formulations were tested and work the same way); optimal solution obtained from a closed-loop MPC with long horizon and perfect information; heuristic policy operates equipment when electricity is cheaper on a weekly average; minimum-level (10%) constraint violation penalty = 30 $/MWh (of storage load), proportional to the violation magnitude (penalty = 90 $/MWh in the case of tank depletion).

4.1. *Case Study 1: Perfect Information Scenario*

We first evaluate how the SAC method performs under a perfect information scenario and compare it to an optimal (upper bound), a heuristic (baseline), and a random policy (soft lower bound). Results are shown in Figure 3 (log-scale is used to amplify the data near the optimum), in which the plotted curve is an average of three runs (shaded area represents a single standard deviation). Within the first week of training, the SAC agent overcomes the performance of the random policy, meaning that the agent learns to avoid constraint violations (largest sources of penalties that dominate the random policy's performance). In the next 2-3 weeks, the agent overcomes the performance of the heuristic policy, which indicates a great potential for practical application. In the remaining weeks, the agent closes the gap towards the optimal solution. Figure 4 presents the scheduling variables for a representative week (action can be seen avoiding peak prices) and the network losses and alpha throughout the training period.

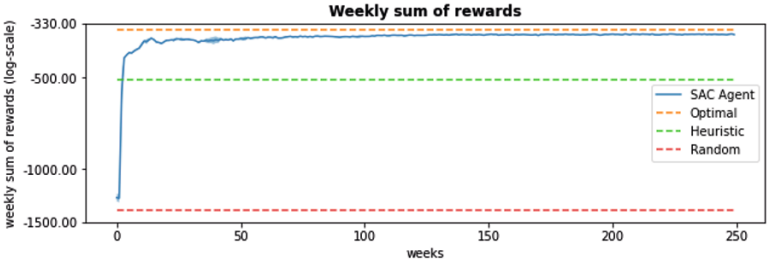


Figure 3. SAC training performance vs. optimal, heuristic, and random policy performances.

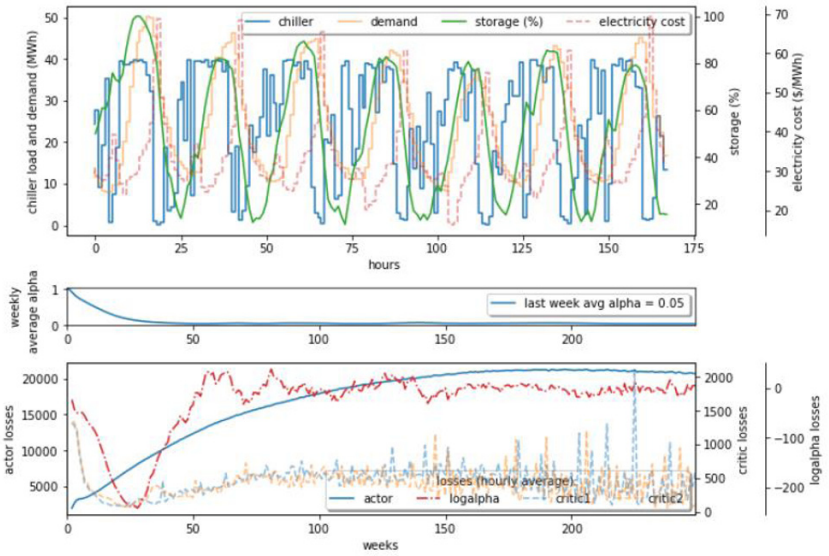


Figure 4. Weekly scheduling results, alpha and losses for the training phase.

4.2. *Case Study 2: Demand Forecast Uncertainty*

We assume a forecast with additive Gaussian noise, which simulates the error increasing further into the future. Five error magnitudes are considered between 0 to 4% of an average demand (24.70 MWh) as the standard deviation of the additive Gaussian noise at each forecast step. Figure 5 presents the MPC solutions with varying recalculation periods (1, 6, 12, 24 h) averaged over 100 weeks to reduce sampling effects from the forecast distribution, and the SAC agent's performance at various training stages (50, 150, 250 weeks). The SAC solution shows good robustness against uncertainty, especially after 250 weeks of training, comparable to employing a fast recalculation frequency for the MPC. The SAC agent shows particular advantage when there is a combination of long recalculation period (or high computational cost) and considerable uncertainty.

5. Conclusions

In this work, we evaluated the application of a DRL method for performing demand response of energy/process systems. The SAC method was employed, a stochastic, off-policy, actor-critic method with built-in entropy maximization that balances exploration and exploitation. Results demonstrate that the SAC agent quickly learns to avoid

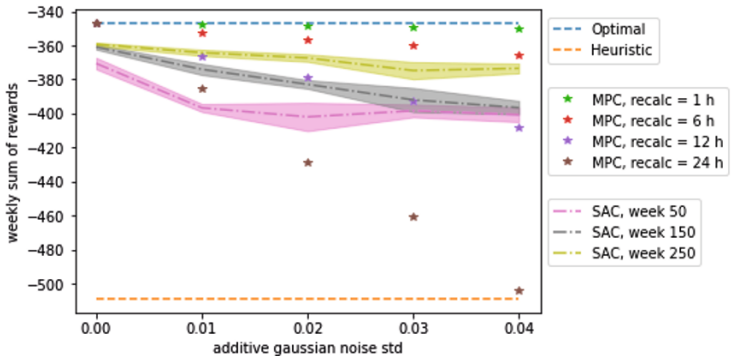


Figure 5. SAC vs. MPC performance under varying forecast uncertainty and recalculation period

constraint violations and continuously closes the gap towards the optimal solution, indicating a good potential for practical application. An analysis of the effects of forecast uncertainty indicated scenarios in which the DRL approach would present advantages over a model-based approach, i.e., when there is a combination of high recalculation period and high uncertainty magnitude. This result is of particular significance when dealing with complex (e.g., nonlinear, multi-scale, mixed-integer, stochastic) systems.

References

Badgwell, T. A., Lee, J. H., Liu, K.-H., Reinforcement Learning – Overview of Recent Progress and Implications for Process Control, *Comput. Aided Chem. Eng.*, 44, 71-85 (2018)

Campos, G., Liu, Y., Schmidt, D., Yonkoski, J., Colvin, D., Trombly, D. M., El-Farra, N. H., Palazoglu, A., Optimal Real-time Dispatching of Chillers and Thermal Storage Tank in a University Campus Central Plant, *Applied Energy*, 300, 117389, (2021)

Cassol, G. O., Campos, G., Thomaz, D. M., Capron, B. D. O., Secchi, A. R. Reinforcement Learning Applied to Process Control: A Van de Vusse Reactor Case Study. *Comput. Aided Chem. Eng.*, 44, 553–558 (2018)

Haarnoja, T., Zhou, A., Abbeel, P., and Levine, S., Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor, *Proc. of PMLR 80* (2018)

Haarnoja, T., Zhou, A., Hartikainen, K., Tucker, G., Ha, S., Tan, J., Kumar, V., Zhu, H., Gupta, A., Abbeel, P., and Levine, S., Soft Actor-Critic Algorithms and Applications. *CoRR* abs/1812.05905. URL http://arxiv.org/abs/1812.05905

Hubbs, C. D., Li, C., Sahinidis, N. V., Grossmann, I. E., and Wassick, J. M., A deep reinforcement learning approach for chemical production scheduling, *Comput. Chem. Eng.* 141, 106982 (2020)

Hwangbo S., Sin, G., Design of control framework based on deep reinforcement learning and monte-carlo sampling in downstream separation, *Comput. Chem. Eng.* 140, 106910 (2020)

Lee, J., Wong, W., Approximate dynamic programming approach for process control, *J. Process Control*, 20, 1038-1048, (2010)

Ma, Y., Borrelli, F., Hencey, B., Coffey, B., Bengea, S., Haves, P., Model predictive control for the operation of building cooling systems, *IEEE Transactions on control systems technology*, 20 (3), 796–803 (2011)

Risbeck, M. J., Maravelias, C. T., Rawlings, J. B., Turney, R. D., A mixed-integer linear programming model for real-time cost optimization of building heating, ventilation, and airconditioning equipment, *Energy Build*, 142, 220–235 (2017)

Yoo, H., Kim, B., Kim, J. W., Lee, J. H., "Reinforcement learning based optimal control of batch processes using monte-carlo deep deterministic policy gradient with phase segmentation, *Comput. Chem. Eng.*, 144, 107133 (2021)

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50267-0

Convolutional Neural Network based Detection and Measurement for Microfluidic Droplets

Shuyuan Zhang[a,b], Xinye Huang[a,b], Kai Wang[a,c], Tong Qiu[a,b*]

[a]*Department of Chemical Engineering, Tsinghua University, Beijing, 100084, CHINA*
[b]*Beijing Key Laboratory of Industrial Big Data System and Application, Tsinghua University, Beijing, 100084, CHINA*
[c]*State Key Laboratory of Chemical Engineering, Department of Chemical Engineering, Tsinghua University, Beijing 100084, CHINA*

qiutong@tsinghua.edu.cn

Abstract

Modern microfluidic systems realize the envisioned idea to perform continuous process operations on a small scale using miniaturized devices and present superiorities in terms of plant modularization, reaction intensification and waste reduction. In microfluidic engineering, droplet size is central to desired function. Therefore, an effective droplet detection and size measurement method is highly-demand to quantitatively reveal the relationship between operation parameters and outcome droplet size. Herein, with recent impressive developments of computer vision, we propose a novel two-step convolutional neural network method to detect and measure droplets in microscopic images. The proposed model first locates droplets with bounding boxes and then calculate the droplet size with detailed coordinates. This convolutional neural network model not only exhibits outstanding performance for droplet size measurement, but also reveals the convenience of deep learning for digital, comprehensive and intelligent microfluidic researches.

Keywords: Microfluidics; Droplet detection; Size measurement; Deep learning; Convolutional neural network.

1. Introduction

Modern continuous-flow process engineering has been explored toward the small-scale device for the prominence in yield, selectivity, scale-up and controllability (Rossetti and Compagnoni, 2016). Among all small-scale reaction technologies, microfluidic engineering is appealing wide attentions and substantially growing in terms of pharmaceuticals, fine chemicals, green chemistry, catalytic reactions and material synthesis that is tough for traditional batch operations (Yan et al., 2021). Microfluidic engineering technology reveals the envisioned idea to perform continuous process operation on a small scale with miniaturized (lab-on-a-chip) devices. For reactions in microfluidic equipment, the process intensification is strengthened with increased mass/heat transfer rates, due to the relatively much larger mass/heat transfer area in the confined volume. Therefore, the size and size distribution of dispersed phase is central to the properties and functions of microfluids (Duraiswamy and Khan, 2009).

Currently, the microfluidic droplet size measurement methods are approximately divided into experimental and imaging methods. Experimental methods are based on probes,

which can be intrusive or non-intrusive, focusing on relating some measurable parameters with the microdroplet size. In intrusive experiments, direct contact between probes and microfluids are inevitable, thus the flow state of microfluids would be disturbed and the measurement result uncertainty would be increased (Chen et al., 2004). Non-intrusive techniques avoid this drawback by adopting non-contacted experimental techniques. For instance, Lucas *et al.* used 3D printing technique to fabricate microfluidic devices with integrated electrodes based on contactless conductivity detection (Duarte et al., 2017). However, non-contacted techniques are still confronted with problems on expensive sensors, complicated installation and poor portability.

With the application of high-speed camera, optical imaging method is attracting wide attentions from microfluidic research communities. However, the post-processing of photographed microscopic images to measure microdroplet size is still rudimentary, mainly by manual measurement on the images at the present stage (Basu, 2013). Commonly, the mechanical image measurement would cost several weeks or months to obtain sufficient data to measure the droplet size. As a consequence, the low efficiency of imaging post-processing will severely harm the scale-up of microfluidic processes, as well as microfluidic device design.

Motivated by recent advancements of computer vision, especially deep learning, it is promising to realize an intelligent method to detect droplets and measure their sizes precisely (Cerqueira and Paladino, 2021; Zhang et al., 2022). Herein, we propose a two-step method based on convolutional neural network (CNN) for microdroplet size measurement. The proposed method first detects droplets in the image by locating them with bounding boxes (Bboxes), then calculates droplets' equivalent diameters with detailed coordinates of their Bboxes. In this way, the droplet size can be rapidly calculated and the size distribution curves can be easily acquired, with high droplet detection precision and low size measurement error. This method not only exhibits outstanding performance of CNN for droplet size measurement, but also reveals the convenience of deep learning for digital, comprehensive and intelligent microfluidic researches.

2. Experiment

We utilized a popular capillary-assembled microchannel as the research microfluidic system in this work, which was fabricated with polymethylmethacrylate (PMMA). The experimental setup is shown in Figure 1. The outer and inner diameters of the inserted capillary were 710 μm and 410 μm, respectively. A stepped T-junction was used as the micromixer to disperse one phase into another phase for producing microdroplets, due to its enhanced shearing effect (Wang et al., 2015). The narrow slit of the stepped T-junction is 200 μm high. After that, an observing chamber was attached with 2 mm wide window on both sides for microphotography. The channel size is ~3 times larger in the observing chamber, where the generated droplets are slowed down to acquire clear microscopic images. An optical microscope (XSP-63B, Shanghai Optical Instrument, China) is equipped with a CMOS camera (B742-F, PixeLink, Canada) to snap images of produced microfluidic droplets. The optical information was sent to computers and converted to digital images with a scale bar of 1.65 pix/μm for post-processing. A collector was set at the end to collect produced microdroplets.

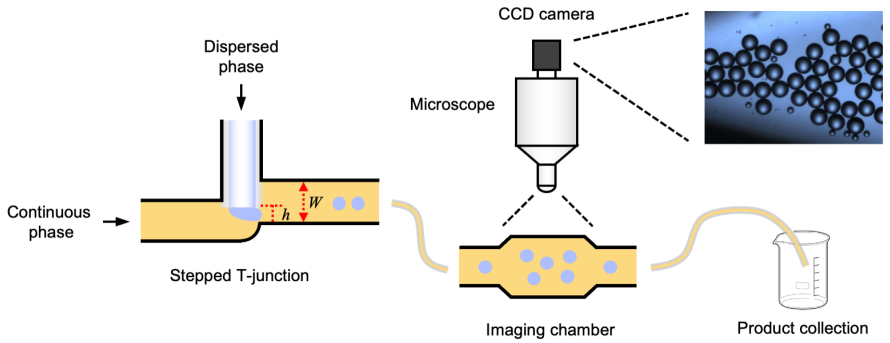


Figure 1. Microfluidic experimental setup.

In the experiment, styrene was selected as the dispersed phase, an aqueous solution of 3.0 wt% sodium dodecyl sulfonate (SDS) and 2.0 wt% polyvinylpyrrolidone K30 (PVP-K30) was selected as the continuous phase. SDS is a surfactant to decrease the interfacial tension, and PVP-K30 is the thickener of the solution for producing droplet swarm. The continuous and dispersed phases were continuously pumped into the T-junction with 50 mL and 10 mL gastight syringes (SEG, Australia), respectively, using commercial syringe pumps (Fusion 6000 for the continuous phase and Fusion 4000 for the dispersed phase). The width and height of photographed images are 1024 and 768 pixels, respectively. To realize a stable jetting flow regime, the flow rate of continuous phase is set between 3~8 mL/min and the flow rate of dispersed phase is set between 1~20 μL/min. This extreme phase ratio is as set to obtain tiny microdroplets. Besides, in some high flow rate ratio cases, the retracted neck of the dispersed phase may by stretched by the continuous phase, and break up into a smaller droplet, of which the size was smaller than dominant droplets.

After experiment, microfluidic images are collected to construct training and testing datasets, which are summarized in Table 1. For training neural networks, it is required to annotate droplets in these images (LeCun et al., 2015). Those droplets truncated by the image borders are also annotated with Bboxes but excluded out in size measurement, because the detailed ordinates of some corner points of the Bbox are missing. There are 2276 and 862 droplets in training and testing datasets, respectively. The mean, maximum and minimum diameters of the training dataset are close to those of the testing dataset, demonstrating the constructed datasets are qualified to evaluate the proposed CNN-based size measurement method.

Table 1. Annotated training and testing datasets of microdroplet image.

	The number of droplets	Mean diameter/μm	Min diameter/μm	Max diameter/μm
Train	2276	35.3	7.3	70.0
Test	862	38.2	8.8	69.4

3. Method

The proposed two-step CNN-based droplet size measurement method is exhibited in Figure 2. The first step is to detect droplets and regress their coordinates with Bboxes. The second step is to filter droplets not truncated by image borders and calculated the equivalent diameters.

In the first step, microfluidic images are input into a CNN backbone to extract abstract features. Then, potential regions that contain droplets are proposed. After that, features in each proposed regions are transformed into a same size with the regions of interest (RoI) pooling operation. Finally, a sharing fully-connected (FC) layer and two separate FC layers are used to distinguish whether the proposed region contains a real droplet and regress the droplet locations. In the second step, droplets whose bounding boxes reach any image border would be excluded out in size measurement and the size is determined as the mean value of height and width of the Bbox.

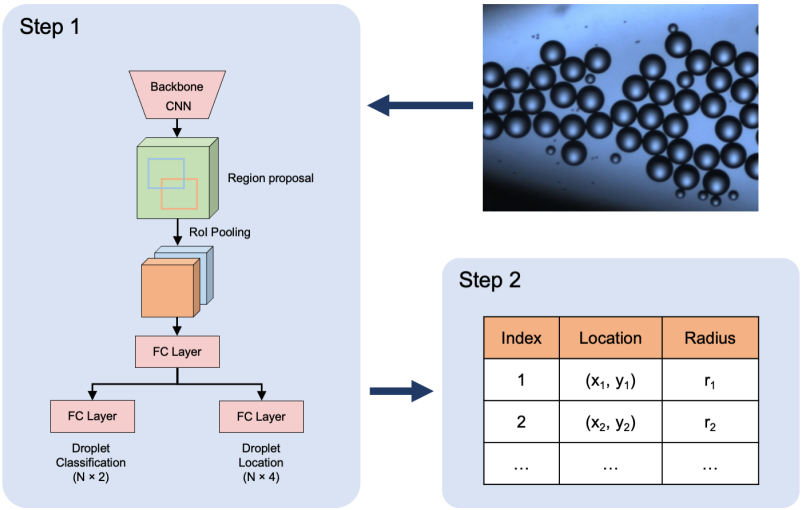


Figure 2. Schematic of the proposed CNN-based microdroplet size measurement method.

For model evaluation, the average precision (AP) under certain intersection-of-union (IoU) condition is utilized (Ren et al., 2016). For a pair of predicted and actual Bboxes, if their IoU is larger than a predetermined threshold, the prediction can be regarded as a matched one. AP is the averaged ratio of matched predictions to all predictions over all images, which can reflect the microdroplet detection accuracy. The IoU threshold is set as 0.75 in this work and the corresponding AP is denoted as AP75. Meanwhile, mean absolute error (MAE) of predicted diameter is also considered as the most straightforward criterion.

4. Results and discussions

4.1. Model tuning

During training, many reported CNN architectures can be implemented as the CNN backbone for effective feature extraction. In this work, an advanced CNN model series, ResNet, is selected as the backbone for feature extraction. ResNet is characterized by the skip identity mapping from shallower layers to deeper layers (Figure 3(a)), which realizes an incremental learning pattern in deep learning (He et al., 2016). Normally, ResNet that contains more CNN layers is inclined to perform better. Three ResNets with different number of layers are investigated, which are ResNet18, ResNet34 and ResNet50, respectively.

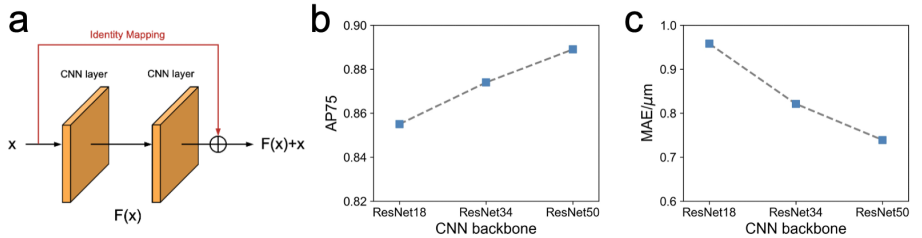


Figure 3. (a) Residual block of ResNet. (b) AP75 and (c) MAE results using ResNet backbones.

The AP75 and MAE results of ResNet50, ResNet75 and ResNet50 are exhibited in Figure 3(b) and (c), respectively. It can be observed that, as the ResNet backbone becomes deeper, AP75 is increasing and MAE is decreasing. AP75 reflects the droplet detection accuracy and MAE reflects the droplet measurement preciseness. Therefore, a deeper ResNet backbone can contribute a precise droplet detection and measurement, although the calculation cost is also increased because more CNN layers are included in the model. In application, ResNet with certain number of layers can be selected as the CNN backbone to meet the requirement of measurement preciseness or calculation speed (or both). Specifically, a high AP75 of 0.889 and a low MAE of 0.739 μm are achieved by ResNet50.

4.2. Measurement results

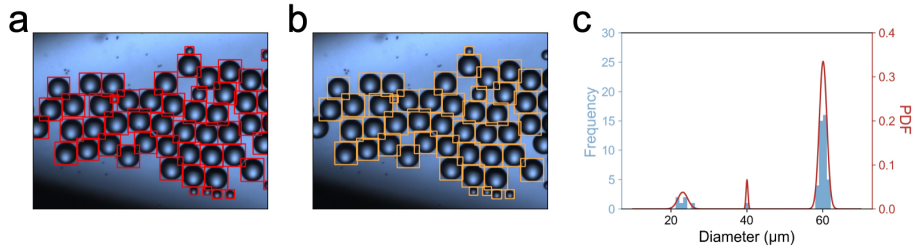


Figure 4. (a) All detected Bboxes, (b) non-truncated Bboxes, and (c) diameter frequency and probability density function (PDF) curve of a typical droplet image.

Figure 4 exhibits a typical photographed droplet image and corresponding inference result of the proposed CNN model. With the powerful feature extraction capability of CNN, almost all droplets can be detected and precisely located with a Bbox to depict their 2-dimensional existence limits (Figure 4(a)). After that, Bboxes that cover the image border are excluded out, as shown in Figure 4(b). The measured droplet diameter is determined as the average of the height and width of the Bbox. Finally, frequency can be easily calculated and PDF curve can be fitted with Gaussian mixture model. In Figure 4(c), the major peak can be ascribed to dominant droplets and the other two peaks can be ascribed to aforementioned satellite droplets.

5. Conclusions

In conclusion, this novel work proposes an intelligent and precise CNN-based model for detecting and measuring droplets in microfluids. This CNN model can comprehensively and precisely locate microdroplets with Bboxes. Based on that, microdroplet diameters can be easily measured and diameter distribution can be obtained for analysis. It is

noteworthy that this CNN model can be implemented to analyse microdroplets with different diameters, even those satellite droplets with diameter as small as ~20 μm. With advanced deep learning technique, this CNN model not only reaches a human-level preciseness in droplet size measurement, but also sheds light on fast and precise microfluid analysis for microfluidic device design and in-depth microfluidic flow research in the future.

References

Basu, A.S., 2013. Droplet morphometry and velocimetry (DMV): a video processing software for time-resolved, label-free tracking of droplet parameters. Lab. Chip 13, 1892–1901. https://doi.org/10.1039/C3LC50074H

Cerqueira, R.F.L., Paladino, E.E., 2021. Development of a deep learning-based image processing technique for bubble pattern recognition and shape reconstruction in dense bubbly flows. Chem. Eng. Sci. 230, 116163. https://doi.org/10.1016/j.ces.2020.116163

Chen, J.Z., Darhuber, A.A., Troian, S.M., Wagner, S., 2004. Capacitive sensing of droplets for microfluidic devices based on thermocapillary actuation. Lab. Chip 4, 473. https://doi.org/10.1039/b315815b

Duarte, L.C., Chagas, C.L.S., Ribeiro, L.E.B., Coltro, W.K.T., 2017. 3D printing of microfluidic devices with embedded sensing electrodes for generating and measuring the size of microdroplets based on contactless conductivity detection. Sens. Actuators B Chem. 251, 427–432. https://doi.org/10.1016/j.snb.2017.05.011

Duraiswamy, S., Khan, S.A., 2009. Droplet-Based Microfluidic Synthesis of Anisotropic Metal Nanocrystals. Small 5, 2828–2834. https://doi.org/10.1002/smll.200901453

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539

Ren, S., He, K., Girshick, R., Sun, J., 2016. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. arXiv:1506.01497.

Rossetti, I., Compagnoni, M., 2016. Chemical reaction engineering, process design and scale-up issues at the frontier of synthesis: Flow chemistry. Chem. Eng. J. 296, 56–70. https://doi.org/10.1016/j.cej.2016.02.119

Wang, K., Qin, K., Lu, Y., Luo, G., Wang, T., 2015. Gas/liquid/liquid three-phase flow patterns and bubble/droplet size laws in a double T-junction microchannel. AIChE J. 61, 1722–1734. https://doi.org/10.1002/aic.14758

Yan, Z., Tian, J., Wang, K., Nigam, K.D.P., Luo, G., 2021. Microreaction processes for synthesis and utilization of epoxides: A review. Chem. Eng. Sci. 229, 116071. https://doi.org/10.1016/j.ces.2020.116071

Zhang, S., Liang, X., Huang, X., Wang, K., Qiu, T., 2022. Precise and fast microdroplet size distribution measurement using deep learning. Chem. Eng. Sci. 247, 116926. https://doi.org/10.1016/j.ces.2021.116926

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50268-2

Deep Reinforcement Learning Based Controller for Modified Claus Process

Jialin Liu[a*], Bing-Yen Tsai[a], Ding-Sou Chen[b]

[a]*Research Center for Smart Sustainable Circular Economy, Tunghai University, No. 1727, Sec.4, Taiwan Boulevard, Taichung, Taiwan*
[b]*Green Energy and System Integration Research and Development Department, China Steel Corporation, No. 1, Chung Kang Road, Hsiao Kang, Kaohsiung, Taiwan*
jialin@thu.edu.tw

Abstract

The modified Claus process is used to recover sulfur from acid gases containing high concentrations of H_2S. The downstream process, a tail gas treatment system, is required to reduce atmospheric emission of sulfur compounds to the level required by air pollution control regulations. In this study, a deep reinforcement learning (DRL) based controller was developed to minimize the concentration variations of H_2S and SO_2 in the tail gas from the modified Claus process. In addition, the sequence to sequence (Seq2Seq) networks were trained by the plant data to capture the dynamic information between the manipulated and controlled variables. Thereafter, the optimal operating policy can be found through that the advantage actor-critic (A2C) algorithm was implemented to the DRL agent by interacting with the environment constructed by Seq2Seq. The results show that the variations of H_2S and SO_2 can be reduced 40 % and 36 %, respectively, compared with that of applying the traditional control strategy.

Keywords: Deep Reinforcement Learning; Sequence to Sequence Networks; Modified Claus Process.

1. Introduction

The reinforcement learning (RL) framework consists of a learning agent interacting with a stochastic environment. The agent selects an action (a_t) at time t according to the probabilities that are generated by a learning policy $\pi(a|s)$ with the current state (s_t) observed from the environment. The selected action interacts with the environment to obtain the reward (r_t) and the next state (s_{t+1}). The deep reinforcement learning (DRL) is referred to approximate the learning policy and the cumulative reward, which is also known as the state-value function, by two deep neural networks (DNNs). Williams (1992) proposed a policy-gradient learning algorithm that generates an episode of actions, rewards and states following the policy. Thereafter, the DNN weightings of the learning policy is updated to maximize the cumulative rewards. Mnih et al. (2015) proposed a deep Q-network (DQN) to maximize the action-value function (also known as Q) and reported that the DQN can achieve a high level of performance on any of a collection of different problems by the same architecture of DNN. The actor-critic algorithm combines the features of the policy-gradient and value-based approaches (Sutton and Barto, 2018). The DNN weightings of the state-value function are updated by minimizing the errors of approximating the observed cumulative reward, whereas the probability function of the policy is updated to maximize the advantage between the reward and the estimated one by the state-value function.

Recently, the DRL approaches have been introduced into the model predictive control (MPC) community; for example, Spielberg et al. (2019) based on the actor-critic (AC) algorithm to develop a DRL-based controller that learned the control policy in real time by interacting with the simulation examples, in which setpoint tracking problem on single-input single-output, multiple-input multiple-output, and nonlinear systems were demonstrated. Petsagkourakis et al. (2020) applied the policy gradient (PG) method from batch-to-batch data to update a control policy to maximize the product concentration of a bioprocess. In their approach, a preliminary optimal control policy was obtained by interacting with the environment that simulated the real bioprocess. Subsequently, the policy was refined by implementing into the true system. Therefore, the number of evaluations with the true system was reduced, which may be costly and time consuming. Ma et al. (2019) designed a DRL controller based on the deep deterministic policy gradient (DDPG) method interacting with a simulated semi-batch polymerization system. More recently, Kang et al. (2021) proposed a two-stage training deep deterministic policy gradient (2S-DDPG) algorithm to control the boiler drum level, which was simulated by a set of transfer functions. The above-mentioned DRL approaches were interacted with the over-simplified mathematical models. The potential of applying DRL-based controllers to the real processes was not demonstrated.

Ma et al. (2020) applied a four-layer feedforward neural network to build a step-ahead prediction model using the experimental data from a bioreactor. In their approach, the DRL-based controller was developed by the asynchronous advantage actor-critic (A3C) algorithm. The experimental results showed that the A3C controller significantly improves the yield of the desired product compared to that of using a traditional control method. Adams et al. (2021) proposed a deep reinforcement learning optimization framework in which the environment was built by a 5-layer DNN from over 1.5 years of plant data with a 1 min sampling time, which interacted with an A2C agent. The objective of the framework was to maximize the power generation of a coal-fired plant while reducing the NO_x emission. In their approach, a static DNN was applied as the surrogate model to predict the power generation and the NO_x emission by the process variables. The actions of the manipulated variables were determined by the A2C agent with the predictions of the static DNN. Therefore, the process dynamic information was not incorporated into their proposed framework. In this study, the process model, which interacts with the DRL agent, is constructed by the sequence to sequence (Seq2Seq) networks (Sutskever et al., 2014). Chou et al. (2020) developed a physically consistent soft sensor by the Seq2Seq networks. They reported that the process dynamics can be fully extracted by the Seq2Seq model from the plant data, because the estimation of the process gains is consistent with the domain knowledge.

2. Seq2Seq and A2C Networks

In this study, the Seq2Seq networks are constructed by the gated recurrent units (GRUs, Cho et al., 2014), which was proposed to modify the drawback of RNNs. The original RNNs suffer from exploding or vanishing gradient problems through backpropagation on multiple time steps. The reset and update gates are added on the structure of RNNs to solve this issue. Figure 1 shows that the encoder extracts the dynamic information from the operating data with past window length w and capsules into a hidden state vector ($\mathbf{z}_0$). In the encoder, the input layer of each node contains the previous hidden state vector ($\mathbf{h}_{i-1}$) and the current data of input and output variables, which are the disturbance ($\mathbf{d}_i$), manipulated ($\mathbf{m}_i$) and controlled ($\mathbf{c}_i$) variables, respectively. The decoder predicts

the future f samples of the controlled variables by the corresponding data of the manipulated variables and the hidden state vector from the previous GRU, as shown in Figure 1.

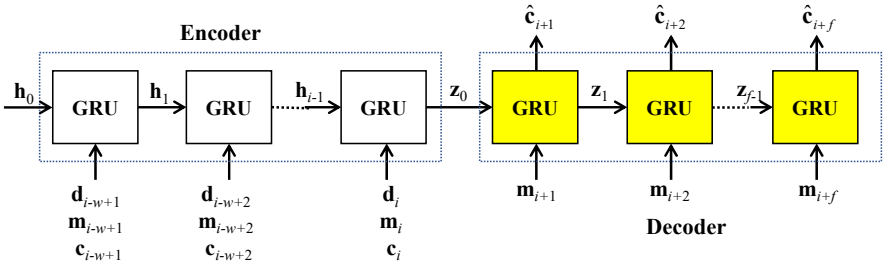


Figure 1. Seq2Seq networks for process modelling

The A2C algorithm uses two deep neural networks to approximate the learning policy and the state-value function, respectively. The probability function of learning policy, which is also called the actor network denoted by $\pi\left(A_t \middle| S_t, \boldsymbol{\theta}\right)$ with weightings $\boldsymbol{\theta}$, maps the current state (S_t) into an action (A_t), interacting with the environment to observe the reward (R_t) and the next state (S_{t+1}). The cumulative reward at time t can be expressed by $G_t = R_t + \gamma R_{t+1} + \gamma^2 R_{t+2} + ... = R_t + \gamma G_{t+1}$ with a discount rate γ. The state-value function, which is the critic network denoted by $\hat{v}\left(S_t, \mathbf{w}\right) \approx G_t$ with weightings $\mathbf{w}$, estimates the cumulative reward at the current state. The correlation of the cumulative reward can be applied into the state-value function, i.e., $\hat{v}\left(S_t, \mathbf{w}\right) = R_t + \gamma \hat{v}\left(S_{t+1}, \mathbf{w}\right)$. Therefore, the temporal difference (TD) error is defined as: $\delta \equiv R_t + \gamma \hat{v}\left(S_{t+1}, \mathbf{w}\right) - \hat{v}\left(S_t, \mathbf{w}\right)$. To improve the accuracy of the cumulative reward approximated by the critic network, the weightings of state-value function ($\mathbf{w}$) are updated by minimizing the square of TD error, i.e., $\mathbf{w} \leftarrow \mathbf{w} + \alpha^w \delta \nabla \delta$ where α^w is a learning rate. In addition, the TD error can be expressed as an advantage function, i.e., $\delta = R_t - \left[\hat{v}\left(S_t, \mathbf{w}\right) - \gamma \hat{v}\left(S_{t+1}, \mathbf{w}\right)\right]$ in which the square bracket term is the current reward estimated by the critic network. A proper policy function should be designed to maximize the advantage; therefore, the weightings of policy network ($\boldsymbol{\theta}$) are updated by $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \alpha^\theta \delta \nabla \ln \pi\left(A_t \middle| S_t, \boldsymbol{\theta}\right)$ where α^θ is a learning rate. The details of A2C algorithm can be found in Sutton and Barto (2018).

In this study, the A2C agent interacting with the Seq2Seq model is proposed. The pseudocode of the proposed approach is listed in Table 1. Two GRU networks, which were initialized with parameters $\boldsymbol{\theta}$ and $\mathbf{w}$, are used as the actor and critic networks, respectively. Each network is deployed with two layers and 30 hidden nodes to map the current state into actions and state values. As listed in Table 1, the initial state at time t ($S_{t,0}$) is defined by the disturbance data ($\mathbf{d}_t$), the measurements of manipulated variables ($\mathbf{m}_t$) and the corresponding predictions ($\hat{\mathbf{c}}_t$) by the decoder of the Seq2Seq networks at Line 3. Starting from Line 4, the weighting $\mathbf{w}$ and $\boldsymbol{\theta}$ are updated 20 times (T = 20) for each sample. The future actions ($\tilde{\mathbf{m}}_{t,i}$) with f prediction horizon and the corresponding state value ($v_{t,i}$) are generated by the actor-critic networks using the current state at Line 5. Thereafter, the predictions are made by the decoder incorporating with the future actions to form the next state with the same disturbance data at Line 6. Line 7

describes that the next state value ($v_{t,i+1}$) is estimated by the critic network with the new state and the previous state is replaced by the new one. The variance of predictions is calculated and compared with the benchmark (var_B) at Line 8. If the current variance is less than the benchmark, the reward is set to 10 and the benchmark is replaced. Otherwise, the reward is given by the negative variance. Consequently, the TD error can be calculated and used to update the network parameters at Line 9.

Table 1. Pseudocode for integrating A2C agent with Seq2Seq networks

1.	Input: a policy $\pi(a\|s,\boldsymbol{\theta})$ and a state-value function $\hat{v}(s,\mathbf{w})$ with parameters $\boldsymbol{\theta}$ and $\mathbf{w}$
2.	For each sample in training dataset:
3.	Initialize the first state: $S_{t,0}=[\mathbf{d}_t \quad \mathbf{m}_t \quad \hat{\mathbf{c}}_t]$
4.	For $i=0,1,\ldots,T-1$:
5.	$\tilde{\mathbf{m}}_{t,i}=\pi(a\|S_{t,i},\boldsymbol{\theta}),\quad v_{t,i}=\hat{v}(S_{t,i},\mathbf{w})$
6.	$\hat{\mathbf{c}}_{t,i}=Decoder(\tilde{\mathbf{m}}_{t,i}),\quad S_{t,i+1}=[\mathbf{d}_t \quad \tilde{\mathbf{m}}_{t,i} \quad \hat{\mathbf{c}}_{t,i}]$
7.	$v_{t,i+1}=\hat{v}(S_{t,i+1},\mathbf{w}),\quad S_{t,i}=S_{t,i+1}$
8.	$\text{var}_i=Var(\hat{\mathbf{c}}_{t,i})$; If $\text{var}_i\le\text{var}_B$: $R_t=10;\quad \text{var}_B=\text{var}_i$ else: $R_t=-\text{var}_i$
9.	$\delta\leftarrow R_t+\gamma v_{t,i+1}-v_{t,i}$; $\mathbf{w}\leftarrow\mathbf{w}+\alpha_w\delta\nabla\delta$; $\boldsymbol{\theta}\leftarrow\boldsymbol{\theta}-\alpha_\theta\delta\nabla\ln\pi(A_t\|S_t,\boldsymbol{\theta})$

3. Industrial Example

In this study, the DRL-based controller is applied to minimize the concentration variations of SO_2 and H_2S in the tail gas from the modified Claus process. The process flow diagram is shown in Figure 2. The sour gas is fed into the burner reactor, in which H_2S is burned with air to form SO_2 and H_2O, i.e., $2H_2S+3O_2\rightarrow 2H_2O+2SO_2$. Thereafter, the effluent gas from the burner reactor is cooled and fed into the converter reactors, which are labeled as R1 and R2 in Figure 2, for catalytic conversion of H_2S and SO_2 to elemental sulfur and water ($2H_2S+SO_2\rightarrow 2H_2O+3S$). In addition to maintain the reactor temperatures, two control loops are used to stabilize the process operations by adjusting the air flowrates. According to the operational guidelines, the primary air flowrate is determined by the sour gas flowrate with a ratio controller whose setpoint of the air to sour gas ratio is recommended as 1.1 by volume. Furthermore, the secondary air flowrate is manipulated to maintain the molar ratio of H_2S to SO_2 whose setpoint ought to be 2. However, the historical data show that the variation of the air to sour gas ratio ranges between 1.2 and 1.6. On the other hand, the molar ratio of H_2S to SO_2 spreads from 2 to 16. That indicates the air flowrate controllers might not work properly; thereafter, the downstream process, the tail gas treating unit, suffers from the large variations of the H_2S and SO_2 concentrations in the tail gas.

The Seq2Seq model was built by five-month operating data, which were collected once per minute around 180,000 samples. The encoder contained all variables listed in Figure 2. On the other hand, the inputs for the decoder were the primary and secondary air flowrates, and the outputs were the H_2S and S_2O concentrations in the tail gas. One layer of GRU with 30 hidden nodes was applied to the encoder and decoder, respectively. The optimal window length (w) of the encoder was determined as 40 samples by the mean absolute percentage errors (MAPEs) of the test dataset in which

5600 samples after modelling data were applied. The Seq2Seq networks predicted the future sixty samples of the controlled variables once per minute using the corresponding manipulated variables. The MAPEs of the test data, for which the predictions were made by the different time periods in 10, 30, and 60 minutes, are listed in Table 2. The MAPEs of predicted H_2S and SO_2 are around 4% and 8%, respectively. The accuracy of predictions show that the Seq2Seq networks capture the process dynamic behavior, properly. Therefore, the Seq2Seq model can be used as the environment interacting with the DRL agent.

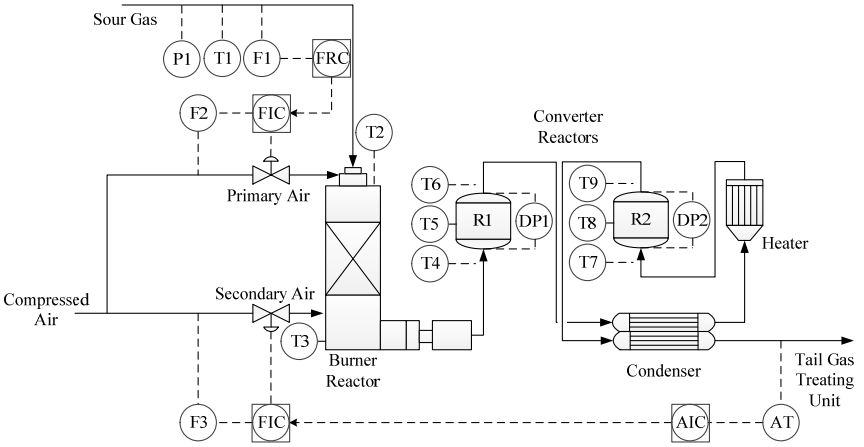


Figure 2. The process flow diagram of the modified Claus process

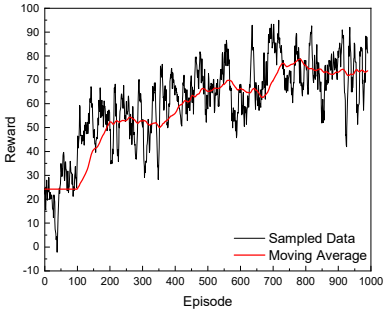


Table 2. Result summary

		10 min	30 min	60 min
		MAPE		
	H_2S	3.7%	4.0%	4.1%
	SO_2	8.1%	8.0%	8.1%
Standard Deviation (H_2S: 0.045, SO_2: 0.013)				
S2S	H_2S	0.040	0.039	0.039
	SO_2	0.011	0.011	0.011
A2C	H_2S	0.024	0.026	0.028
	SO_2	0.007	0.008	0.009

Figure 3. Reward evolution for training the A2C agent

The A2C agent was trained during 1000 episodes where the weightings of actor-critic networks were updated 20 times in each episode. For each update, the future sixty actions were drawn from the actor network; thereafter, the corresponding H_2S and SO_2 concentrations were predicted by the Seq2Seq model incorporating with the future actions. Consequently, the reward was evaluated by the variances of the predictions. As shown in Figure 3, the reward was initially around 20 that could be improved to 70 at the end of training. The standard deviations of the H_2S and SO_2 concentrations were 0.045 and 0.013 for the raw data, as listed in Table 2. The standard deviations of the predictions by the Seq2Seq model are comparable with those of the raw data in the different time periods, implementing with the original measurements of the primary and

secondary air flowrates. On the other hand, the A2C agent was implemented to determine the future actions of primary and secondary air flowrates every minute in the periods of 10, 30, and 60 minutes, respectively. Compared with the results of the standard deviations by the Seq2Seq and A2C networks listed in Table 2, the variations were reduced around 28%–40% for the H_2S and 18%–36% for the SO_2, respectively.

4. Conclusions

The DRL-based controller was developed by the A2C agent interacting with the environment constructed by the Seq2Seq model. The process dynamic feature could be captured by the encoder; meanwhile, the correlation between manipulated and controlled variables was extracted by the decoder. Thereafter, the reward of future actions generated by the A2C networks was evaluated by the multistep-ahead predictions. The results showed that the proposed approach can reduce the variations of H_2S and SO_2 concentrations, effectively.

References

D. Adams, D-H Oh, D-W Kim, C-H Lee, M. Oh, 2021, Deep reinforcement learning optimization framework for a power generation plant considering performance and environmental issues, J. Clean. Prod., 291, 125915.

K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, 2014, Learning phrase representations using RNN encoder-decoder for statistical machine translation, in EMNLP, https://arxiv.org/abs/1406.1078.

C.H. Chou, H. Wu, J.L. Kang, D.S.H. Wong, Y. Yao, Y.C. Chuang, S.S. Jang, J. D.Y. Ou, 2020, Physically consistent soft-sensor development using sequence-to-sequence neural networks, IEEE T. Ind. Inform., 16, 2829–2838.

J-L Kang, S. Mirzaei, J-A Zhou, Robust control and training risk reduction for boiler level control using two-stage training deep deterministic policy gradient, J. Taiwan Inst. Chem. Eng., (2021), https://doi.org/10.1016/j.jtice.2021.06.050.

Y. Ma, W. Zhu, M.G. Benton, J. Romagnoli, 2019, Continuous control of a polymerization system with deep reinforcement learning, J. Process Control, 75, 40–47.

Y. Ma, D.A. Noreña-Caro, A.J. Adams, T.B. Brentzel, J.A. Romagnoli, M.G. Benton, 2020, Machine-learning-based simulation and fed-batch control of cyanobacterial-phycocyanin production in Plectonema by artificial neural network and deep reinforcement learning, Comp. Chem. Eng., 142, 107016.

V. Mnih, K. Kavukcuoglu, D. Silver, A.A. Rusu, J. Veness, M.G. Bellemare, ..., Ostrovski, G., 2015, Human-level control through deep reinforcement learning, Nature, 518, 529–533.

P. Petsagkourakis, I.O. Sandoval, E. Bradford, D. Zhang, E.A. del Rio-Chanona, 2020, Reinforcement learning for batch bioprocess optimization, Comp. Chem. Eng., 133, 106649.

S. Spielberg, A. Tulsyan, N.P. Lawrence, P.D. Loewen, R.B. Gopaluni, 2019, Toward self-driving processes: a deep reinforcement learning approach to control, AIChE, 65, e16689.

I. Sutskever, V. Oriol, Q.V. Le, Sequence to sequence learning with neural networks, 2014, Adv. Neural Inf. Process. Syst., 3104–3112.

R.S. Sutton, A.G. Barto, 2018, Reinforcement learning: An introduction (2nd ed.) MIT Press.

R.J. Williams, 1992, Simple statistical gradient-following algorithms for connectionist reinforcement learning. Machine Learning, 8, 229–256.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50269-4

Process performance prediction based on spatial and temporal feature extraction through bidirectional LSTM

Changrui Xie[a], Runjie Yao[b], Zhengbang Liu[b], Lingyu Zhu[b], Xi Chen[a*]

[a]: *State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou, Zhejiang 310027 P.R. China*
[b]: College of Chemical Engineering, Zhejiang University of Technology, Hangzhou, Zhejiang, 310014, China
*Corresponding author：*xi_chen@zju.edu.cn*

Abstract

With the development of deep learning, it has been a trend to build data driven soft sensors in process industries with neural networks. There are a number of networks proposed to deal with time series prediction, such as Recurrent Neural Network (RNN) and Long Short-Term Memory Network (LSTM). However, it is a critical part to extract nonlinear and dynamic characteristics hiding in process data collected from industrial production. This paper proposes a novel approach for performance prediction based on the spatial and temporal feature extraction through bidirectional LSTM networks (BiLSTM) for a reactor network. Due to the superiority of processing sequences from both directions, BiLSTM are utilized to simulate the physical structure of the reactor network. With both spatial and temporal feature extraction, the deep learning model through BiLSTM achieves nice prediction performance.

Keywords: Bidirectional Long Short-Term Memory Network, Soft sensing, Feature extraction, Deep learning.

1. Introduction

In comparison to the traditional offline analysis in laboratory, soft sensing provides a more fast and economical way to predict critical quality variables, which has been widely used in plenty of industrial plants. With the rapid development of machine learning and statistics, great progress has been made in the field of data-driven model based soft sensing. Different from first-principle models, data-driven models are developed with available data collected during industrial productions even without exact mechanism. However, data-driven models have to attach more emphasis on how to extract as many relevant nonlinear and dynamic features as possible to capture the valid characteristics of the complex chemical processes, because the validity of features will determine the performance of soft sensor directly (Ma et al. 2018). Kaneko et al. (2009) developed a new soft sensor combining independent component analysis (ICA) and partial least squares (PLS) together, where independent components can be seen as features sensitive to the outliers, then a PLS model can be updated with normal samples. Corrigan et al. (2021) proposed a soft sensor model based on dynamic kernel slow feature analysis, which was utilized to extract slowly varying features. Sun et al. employed multi-layer perceptron (MLP) to model the complex desulfurization process, based on which a soft

sensor was built for SO_2 emission. Although many approaches have been proposed for feature extraction for the purpose of soft sensing, such as principal component analysis (PCA), ICA and deep learning methods, there is still a lot of room for improvement to extract the hidden nonlinear and dynamic features from vast process data especially in certain industrial scenarios. In this paper, a deep learning model with spatial and temporal feature extraction is proposed for a reactor network (Dorgo et al. 2019). The BiLSTM networks are utilized to simulate the physical structure of the reactor network to extract spatial features, then a unidirectional LSTM network is followed to process the feature maps output from bidirectional LSTM networks to extract the temporal features at multiple time points. Thus, the nonlinear and dynamic characteristics of the reactor network can be well captured through with both spatial and temporal feature extraction.

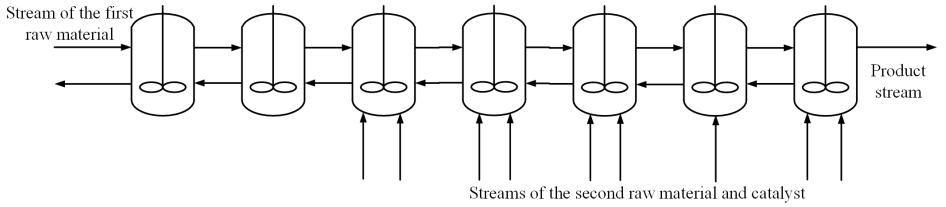


Figure 1. Structure diagram of the reactor network

2. Reactor network process

The reactor network considered in this paper is a system consisting of seven successively connected reactors, as is shown in Figure 1. There are mainly two raw materials entering the system, one from the first reactor directly and the other from the third to the seventh reactors. Overall, the reactor network keeps a counter-flow structure. In each reactor, two liquid phases exist and two outflow streams are kept via a separator inside. On account of the complex structure of the reactor network, it is natural that the dynamic and nonlinear characteristics of each reactor will propagate among the reactor network. During the industrial production, the outlet concentration of components from the fourth reactor are of the most significance, because they are seemed as a flag to reveal the real-time state of reactions in the whole reactor network. Therefore, they are usually selected as the critical quality variables that determines the process performance and operators usually adjust the feed flow of raw materials to maintain the production stability based on this observation. In this task, we select the outlet concentration of three components from the fourth reactors as the predicting target.

3. Feature extraction by bidirectional Long short-term memory networks

Sequence problems is considered as one of the hardest problems in many industrial cases. RNN is widely used to deal with this kind of problem. But RNN works well only towards short sequences, it may suffer from carrying information from earlier time steps to later ones when sequences are long enough due to gradient vanishing. LSTM, as a variant of RNN, improves the architecture by introducing a mechanism named cell state, by which it can preserve the relevant information to the later units even when the sequence is very long. As shown in Figure 2, LSTM propagates the information with cell state and hidden state produced by three gates inside, named forget gate, input gate and output gate. These three gates determine the information needed to be remembered or forgot.

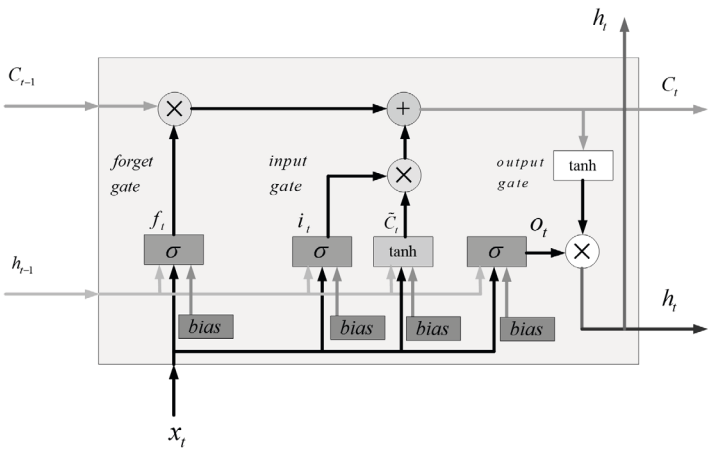


Figure 2. Data flow in the cell in LSTM network

With the help of mechanism of LSTM, information in the past can be preserved and passed to the future then an excellent prediction result can be obtained. However, the propagation of information in sequences may be in both directions in some cases, such as natural language processing. In terms of this issue, bidirectional long short-term memory network is proposed to construct two independent LSTMs at the same time, one of which processes the input sequences from past to future as traditional LSTM, and the other one processes from future to past inversely. By combining the outputs of both independent LSTMs together in some ways, the output of BiLSTM at each time point has the ability to preserve the information in both directions.

Considering the structural similarity between the reactor network and the BiLSTM, we propose using BiLSTM to build the soft sensor for the reactor network. Instead of extracting the temporal features from past to future and future to past in natural language processing, BiLSTM here is used to extract the spatial features at one time point. The input of BiLSTM layer is a feature matrix in shape with the complete information acquired from the whole reactor network. In this way, each column of feature maps output from BiLSTM layer denotes the spatial features for an individual reactor; thus, the fourth of them is selected to predict the outlet concentrations of the fourth reactor. In order to extract as many spatial characteristics as possible, two BiLSTM layers are set in order at the top of networks. Afterwards, these feature vectors are concatenated together and transferred into a unidirectional LSTM layer to extract the temporal features. Then some fully connected layers are followed and an output layer is added at last. Figure 3 illustrates the basic framework of the proposed system for the spatial and temporal feature extraction. Corresponding neuron number and activation function for each layer are listed in the Table 1. Besides, L2 regularization terms are added to the layers to avoid overfitting. The Root Mean Squared Propagation (RMSProp) is selected as the optimization algorithm to conduct gradient descent during network training with the learning rate set at 0.001.

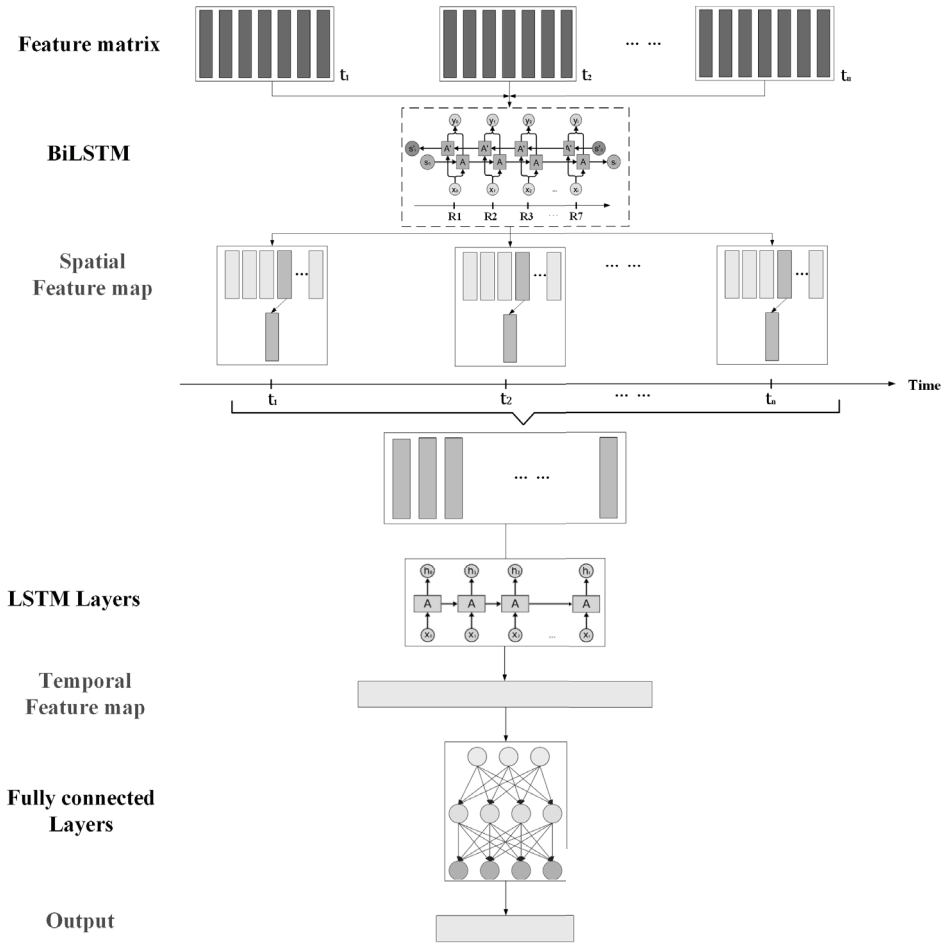


Figure 3. Spatial and temporal feature extraction based soft sensing network

Table 1. Number of neurons and activation function for each layer

Layers	Activation Function	Model with spatial and Temporal feature extraction
Input layer	-	(24,7,7)
BiLSTM1	sigmoid	64
BiLSTM2	sigmoid	32
LSTM	sigmoid	48
Fully connected layer1	tanh	48
Fully connected layer2	tanh	32
Output layer	sigmoid	3

4. Results

In the real industrial operation, the concentration of the target stream is collected and offline analysed for every 4 hours. Although what we aim to predict is the outlet concentration from the fourth reactor, the influence from other reactors is not negligible due to the non-decoupling property of the whole system. Therefore, it is necessary to take variables such as the feed flow of raw materials of other reactors into consideration. Eventually, 49 online measurable process variables are collected from the reactor network system with a sampling rate at 5 minutes. Considering the fact that the real process residence time of the whole system is about 2 hours, the input includes all the information from the past 2 hours. In the end, 1272 data pairs are prepared and are randomly divided as training set and testing set for the neural network model. To ensure consistent distribution between the training set and testing set, 70% of dataset are randomly split as the training set, and the remaining samples as the testing set.

To quantitatively evaluate the discrepancy, the root-mean-square error (RMSE) is used as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{1}$$

where i is the sample index and N is the number of samples in the testing set.

The RMSE for all samples in the testing set are presented in Table 2. The comparison for the first component of the target stream is presented in Figure 4. The soft sensing model with both spatial and temporal feature extraction achieves a good prediction performance. Instead of just inputting all relevant information as a flattened feature vector at each time point, the proposed soft sensing model has the ability to eliminate the useless features or the existence of information redundancy included in the feature matrices.

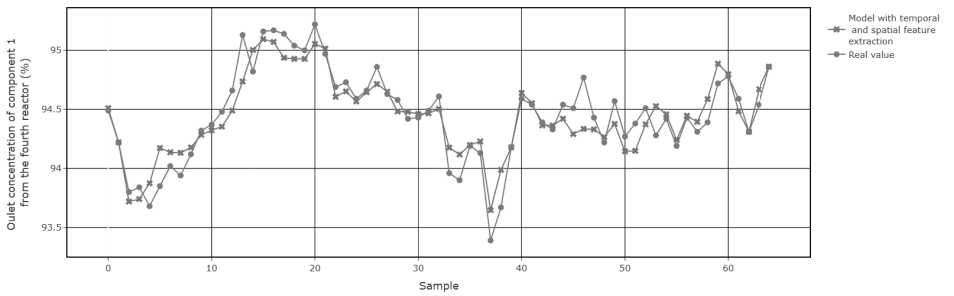


Figure 4. Prediction results of the testing samples

Table 2. RMSE of prediction results by soft sensing model

RMSE	Component 1	Component 2	Component 3
Model with spatial and Temporal feature extraction	0.189	0.152	0.147

5. Conclusions

Based on BiLSTM networks, a novel spatial and temporal features extraction method is proposed for reactor network system in this paper. The BiLSTM networks are utilized to simulate the physical structure of the equipment deployment, with the purpose of extracting spatial information of the whole reactor network at one time point as much as possible; then a unidirectional LSTM layer is followed to extract the temporal features from multiple timesteps. The prediction results indicate that feature extraction by deep learning is beneficial to capture nonlinear and dynamic characteristics of complex reactor systems, which contributes to more accurate predictions for soft sensing. This study provides a new point of view to take advantage of neural networks, that is simulating the real deployment of reactor network with similar structures designed by neural networks, especially for those non-decoupling complex systems.

Acknowledgement
Financial support of the Zhejiang Provincial Natural Science Foundation of China (No. LZ21B060001) is gratefully acknowledged.

References

Y. Ma, B. Huang, 2018, Extracting dynamic features with switching models for process data analytics and application in soft sensing, AIChE Journal, 64, 2037–2051.

H. Kaneko, M. Arakawa, K. Funatsu, 2009, Development of a new soft sensor method using independent component analysis and partial least squares, AIChE Journal, 55, 87–98.

J. Corrigan, J. Zhang, 2021, Developing accurate data-driven soft-sensors through integrating dynamic kernel slow feature analysis with neural networks. Journal of Process Control, 106, 208–220.

K. Sun, X. Wu, J. Xue, F. Ma, 2019, Development of a new multi-layer perceptron based soft sensor for SO2 emissions in power plant, Journal of Process Control, 84, 182-91.

G. Dorgo, J. Abonyi, 2019, Learning and predicting operation strategies by sequence mining and deep learning, Computers & Chemical Engineering, 128, 174-187.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50270-0

Exploring the Potential of Fully Convolutional Neural Networks for FDD of a Chemical Process

Ana Cláudia O. e Souza[a*], Maurício B. de Souza Jr.[b], Flávio V. da Silva[a]

[a] *Chemical Systems Engineering Department, School of Chemical Engineering, University of Campinas, 500 Albert Einstein Avenue, Campinas 13083-872, Brazil*
[b] *School of Chemistry, Federal University of Rio de Janeiro, 149 E-207 Athos da Silveira Ramos Avenue, Ilha do Fundão, Rio de Janeiro 20000-000, Brazil*
a192507@dac.unicamp.br

Abstract

This study investigated three distinct variations of convolutional neural network (CNN) topologies to model a fault detection and diagnosis system. The primary goal was to determine if fully convolutional networks, which do not present fully connected layers or apply the pooling operation, could outperform the well-known traditional convolutional topology. To explore this issue in a chemical process context, the Tennessee Eastman Process was the study case used. Data corresponding to four years of operation was simulated to mimic the big data scenario faced by many industries nowadays. The fully convolutional model provided better average precision and recall results. On top of it, there was a reduction of 80% of the time (elapsed real time) demanded in the training stage when compared with the traditional CNN model evaluated.

Keywords: Fault detection and diagnosis; Deep Learning; Big Data; Fully convolutional neural networks.

1. Introduction

Chemical process safety is one of the biggest concerns of engineers and process operators. It is crucial to ensure the protection of employees and facilities during the large-scale manufacturing of chemical products. Besides that, it is desirable to extend the useful life of equipment as much as possible to reduce operating costs and avoid compromising equipment availability. Therefore, the development of fault detection and diagnosis systems (FDD) is essential for any industrial process. In the last few years, the application of deep learning techniques to modeling FDD frameworks has achieved outstanding results revealing how promising artificial intelligence is to solve complex problems.

Convolutional neural networks (CNNs) are among the most known deep learning neural architectures. Its application for modeling FDD frameworks has recently been addressed in the literature (Wu and Zhao, 2018; Ge et al., 2021). Still, there is much more to explore regarding fully convolutional neural networks (FCNs). Fully convolutional neural networks consist of an end-to-end convolutional network. They do not have fully connected layers (FC) because the classification stage is also performed by convolutional layers (Conv). One of the advantages of FCNs is to demand simpler structures since only Conv layers are necessary. FCNs were successfully used to develop FDD models for navigation systems (Xu and Lian, 2018), insulators of power lines (Chen et al., 2019), and continental sandstone reservoirs (Wu et al., 2021). An

excellent benchmark for evaluating FDD techniques in the process systems engineering field is the Tennessee Eastman Process (TEP) case study (Downs and Vogel, 1993). Therefore, TEP will be used for investigating if FCNs are a good choice for modeling FDD systems in chemical processes. Since the TEP data generation is done by using its available simulation models, it is possible to create a very realistic data distribution domain, similar to the big data scenario that is a reality for many processes and chemical industries nowadays.

This work aimed to model an FDD system for the TEP benchmark comparing the performance of a fully convolutional neural network and a traditional convolutional network with FC layers. In the following Section (2), the study case is described as well as the methodology to collect and preprocess the data. Section 3 is dedicated to presenting and discussing the FDD framework modeled. Finally, Section 5 summarizes the discussion, and some possible future developments are pointed out.

2. Methods

2.1. Tennessee Eastman Process Benchmark

Figure 1 presents the Tennessee Eastman Process (TEP) diagram. This benchmark was proposed by Downs and Vogel (1993) as an attempt to provide a complex study case well-suited for addressing topics like plant-wide control strategies, optimization, multivariable and predictive control, process monitoring and diagnostics, among others. The TEP represents an industrial chemical process composed of a reactor, a product condenser, a vapor-liquid separator, a recycle compressor, and a product stripper.

There are 11 manipulated variables, 22 continuous process measurements, and 19 sampled measurements, totalizing 52 variables that can be used as the input data for the model's training stage. There are also 20 process disturbances implemented, which will be used as process faults for the present work. Heat and material balance data, the list of process operating constraints, as well as the detailed description of the process operating modes can be found in Downs and Vogel (1993).

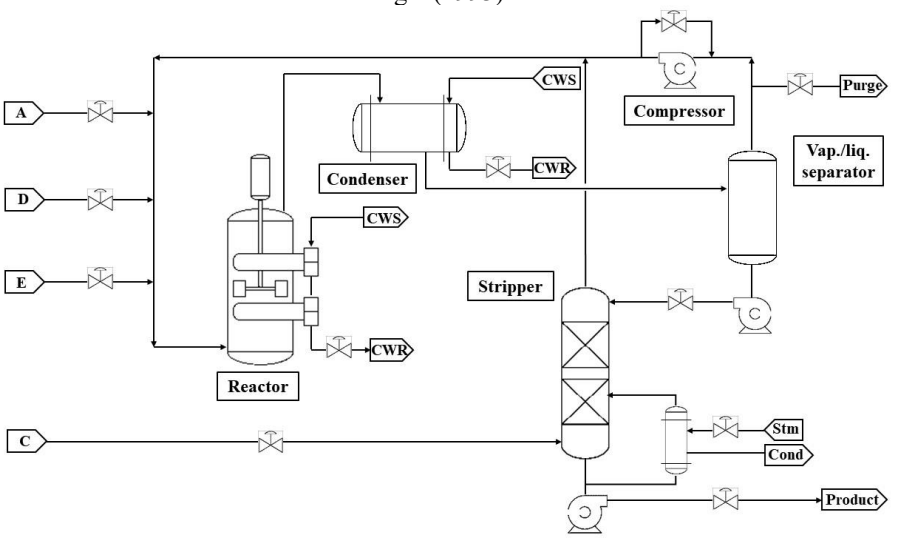


Figure 1. Tennessee Eastman Process diagram.

2.2. Data collection and preprocessing

Two different datasets were generated, one for training with data corresponding to three years of operation (525.600 samples) and the other for testing with one year of data (175.200 samples). Continuous simulations were performed with a sampling time equal to 3 minutes with the occurrence of only one fault at a time (which was randomly chosen among the 20 process faults implemented in the TEP simulator). The duration of each fault was also randomly selected between the established range of 24 and 48 hours.

Regarding data pretreatment, no feature selection technique was applied since the convolutional layers themselves can identify and isolate the most relevant attributes present in the input data. The only preprocessing consisted of normalizing each of the 52 input variables and transforming the data into 4-dimensional tensors. Therefore, *m x n* matrices were generated from the simulated data frames, where *m* represents the time span of each matrix and *n* is the number of input variables. To work with square matrices, which enhance the efficiency of the CNN training (Aggarwal, 2018), matrices with a shape of 52x52 were obtained. Thus, each matrix corresponds to a period of about two and a half hours of data points.

2.3. CNN based FDD framework

Figure 2 shows the proposed framework. After data acquisition, a validation dataset was separated from the training set for the application of early stopping to avoid overfitting. Then, the training, validation, and test data frames were normalized and transformed into 52x52 matrices. The best hyperparameters values for the traditional convolutional neural network (denoted by TCN) were determined by trial and error. The hyperparameters investigated were the number of convolutional layers, number of filters, optimizer method, the number of fully connected layers and their neurons, learning rate, and batch size. The size of the convolutional and pooling kernels was kept constant and equal to (3,3) and (2,2), respectively. Also, for every trial tested, max pooling was the pooling method, categorical cross-entropy was the loss function, ReLu was the activation for the intermediate layers, a SoftMax function was applied in the output layer, and the strides of convolution and pooling were 1 and 2, respectively.

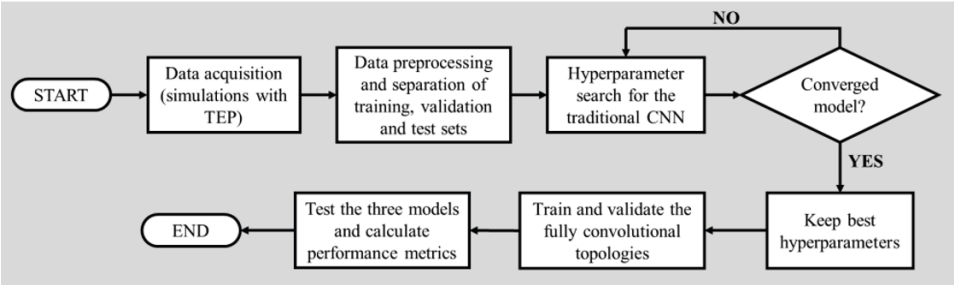


Figure 2. CNN-based fault detection and diagnosis system developing framework.

Once the best hyperparameters for TCN were found, the topology of the network was altered to generate two new networks, denoted by FCN1 and FCN2. The first fully convolutional network, FCN1, does not have fully connected layers (FC) to perform the final classification of the process state. Instead, the target classification is performed by convolutional layers (Conv) with adequate kernel sizes. In the second fully convolutional network, FCN2, not only the FC layers are replaced but also the pooling operation is omitted. Therefore, the downsampling operation is achieved by increasing

the stride of the preceding convolutional layer. The idea of FCN2 corresponds to the "Strided-CNN" presented by Springenberg et al. (2015). No hyperparameter search was conducted for FCN1 and FCN2 to allow the straight comparison of the training and testing stages performances of the traditional CNN topology and the altered ones.

Precision and recall were the performance criteria applied. The simulations for data acquisition, and the training, validation, and testing stages of the models were conducted on a computer with Intel i7-9700 CPU (9th gen) 3.00 GHZ 12MB, 32 GB RAM, and Ubuntu 20.04.1 LTS. The programming language used was Python 3.7.4 with the open-source libraries Keras and Tensorflow.

3. Results and Discussion

For the traditional convolutional network, TCN, with pooling and FC layers, from all the topologies tested during the hyperparameter search, the best one was Conv(20)-Conv(30)-Pool-Conv(40)-Pool-FC(21) with a learning rate of 0.001, batch size of 500 and Adam as the optimizer. Once defined the TCN topology, FCN1 and FCN2 were created. For FCN1, the last FC layer was replaced by a Conv layer with stride 1 and kernel size equal to (13,13). On the other hand, for FCN2 the max pooling layers were removed, the stride of the two preceding Conv layers [Conv(30)-Conv(40)] was increased from 1 to 2, and the last FC layer was also replaced by a Conv layer with stride 1 but kernel size equal to (13,13). Finally, FCN1 and FCN2 were trained using the same hyperparameters of TCN. Table 1 summarizes the described topologies and presents the total number of trainable parameters as well as the time spent to update and optimize the weights and biases of each model (elapsed real time).

Despite the topology differences in the three models, the number of total trainable parameters remains the same, allowing a straightforward comparison between them. The TCN model took one hour and 39 minutes to complete the training stage. The FCN1, which does not use a fully connected layer of neurons for the final classification, was 43 minutes faster to complete the training step. In this case, the number of total parameters does not change because the only thing happening is the conversion of one kind of layer to another. The use of filters with kernels of (13,13) for the last layer of FCN1 guarantees that the total number of learnable parameters will be maintained. On the other hand, FCN2 completes the weights and biases optimization in only 20 minutes. The significant decrease in training time results from the reduction in the overlapping presented by the strided convolution. This is a consequence of increasing the stride by which the filters move across the output of the previous layer since in FCN2 the convolutions themselves are downsampling the intermediate feature maps.

Table 1. Models' topology description and training duration.

Model	**Topology**	**Total training time (min)***	**Total trainable parameters**
TCN	Conv(20)-Conv(30)-Pool-Conv(40)-Pool-FC(21)	99	196,651
FCN1	Conv(20)-Conv(30)-Pool-Conv(40)-Pool-Conv(21)	56	196,651
FCN2	Conv(20)-Conv(30)-Conv(40)-Conv(21)	20	196,651

* Elapsed real time.

Since the reduction in the training time is not the primary goal here (despite being a useful outcome), the models were tested on a separate dataset never seen before during the training and validation stages to evaluate their performances. Precision and recall for the three models were calculated and are presented for each process state (normal operation and 20 different faults) in Table 2. The KPIs of FCN1 and FCN2 with equal or better performances than the traditional topology (TCN) are highlighted.

Among the three trained models, FCN2 is the one that better detects and diagnosis the fault occurrences of the test set with 80.8 % and 80.1 % of precision and recall, respectively. Compared with TCN, FCN2 shows an expressive improvement in detecting faults considered insipient and difficult to learn in the literature (Zhang and Zhao, 2017; Wu and Zhao), like Faults 3, 16, and 18. Also, FCN2 was able to provide a better separation between the faulty instances and the periods of normal operation; this is represented by the observed increase in the precision and recall of the normal status. Besides that, the maintenance of detection of faults with particular dynamic signatures, like Faults 1, 2, 5, 6, 7, and 14, proves that the conversion of FC layers into Conv layers and the removal of max pooling operations do not harm the performance of the FDD system in general.

Table 2. Detailed results of the three models on the test set.

Process status	**TCN**		**FCN1**		**FCN2**	
	Precision (%)	**Recall (%)**	**Precision (%)**	**Recall (%)**	**Precision (%)**	**Recall (%)**
Normal	84.5	82.0	84.5	70.4	86.8	88.2
Fault 1	99.8	99.9	99.9	100.0	99.8	99.9
Fault 2	100.0	100.0	100.0	100.0	99.2	100.0
Fault 3	11.6	13.1	8.4	6.3	29.3	44.9
Fault 4	99.7	98.7	99.2	98.8	99.1	99.9
Fault 5	100.0	100.0	100.0	100.0	100.0	100.0
Fault 6	100.0	92.0	100.0	100.0	100.0	98.0
Fault 7	100.0	100.0	100.0	100.0	100.0	100.0
Fault 8	91.7	94.9	92.5	91.5	93.8	91.6
Fault 9	6.4	2.3	7.3	21.5	8.4	4.2
Fault 10	96.6	85.1	79.0	93.4	97.0	88.6
Fault 11	99.8	99.5	100.0	98.6	99.8	98.5
Fault 12	100.0	100.0	89.7	100.0	100.0	96.2
Fault 13	93.4	89.2	89.7	93.4	92.3	92.7
Fault 14	100.0	100.0	100.0	100.0	100.0	100.0
Fault 15	4.2	13.7	2.9	4.6	4.8	5.8
Fault 16	36.7	48.2	34.8	37.7	58.0	53.9
Fault 17	97.8	97.6	95.8	99.1	97.5	97.9
Fault 18	44.4	30.8	48.8	51.3	62.1	46.2
Fault 19	99.8	96.9	99.8	96.5	91.5	98.8
Fault 20	97.4	98.0	98.2	97.7	98.5	98.6
Average	79.2	78.2	77.6	79.1	80.8	80.1

Therefore, the superior performance of the FCN2 model is clear. The fully convolutional network demanded a simpler topology with only convolution layers and

outperformed the TCN model in all the metrics evaluated. Besides that, it is essential to emphasize that the better results were observed for the FCN2 with a reduction of 80% of the time demanded by the TCN to complete the training stage. The performance of FCN1 was not uniform. Despite the increase in the recall, due to improvements in the detection of some faults (like 9, 10, and 18), the model was not accurate regarding the normal operation, which led to the observation of some false alarms. The ideal scenario is the one where the rates of false alarms and false negatives are both low. So, FCN2 remains the best option between the models investigated in this work.

4. Conclusions

In this work, the potential of fully convolutional neural networks to model a fault detection and diagnosis system for a chemical process was explored. The outstanding performance of a convolutional topology – that does not possess fully connected layers nor max pooling operations – was proved using the Tennessee Eastman Process benchmark. The conversion of FC layers into Conv layers, and the increase of the convolutions stride to perform the downsampling of the internal feature maps, allowed the development of a model cheaper to train and with an improved ability of generalization when facing new data. The parsimonious nature of the fully convolutional neural networks appears to be a promising paradigm for designing adaptive FDD systems applied to processes subject to novel faults or new operational conditions. Given these promising outcomes, some other techniques can be tested to improve even more the achieved results. In future work, an automated hyperparameter tuning will be used to enlarge and optimize the search for the best FCN model settings. Also, the application of transfer learning to improve the detection of incipient faults by the FCN2 will be further investigated.

References

C. C. Aggarwal, 2018, Neural networks and deep learning – A Textbook. 1 ed., Switzerland: Springer Nature.

J. Chen, X. Xu, H. Dang, 2019, Fault detection of insulators using second-order fully convolutional network model, Mathematical Problems in Engineering, 1–10.

J. J. Downs, E. F. Vogel, 1993, A plant-wide industrial process control problem, Computers & Chemical Engineering, 17, 245–255.

X. Ge, B. Wang, X. Yang, Y. Pan, B. Liu, B. Liu, 2021, Fault detection for reactive distillation based on convolutional neural network, Computers & Chemical Engineering, 145, 1–14.

J. T. Springenberg, A. Dosovitskiy, T. Brox, M. Riedmiller, 2015, Striving for simplicity: the all convolutional net, International Conference on Learning Representations (ICLR), 1-14.

J. Wu, B. Liu, H. Zhang, S. He, Q. Yang, 2021, Fault detection based on fully convolutional networks (FCN), Journal of Marine Science and Engineering, 9, 1–13.

H. Wu, J. Zhao, 2018, Deep convolutional neural network model based chemical process fault diagnosis, Computers & Chemical Engineering, 115, 185–197.

H. Xu, B. Lian, 2018, Fault detection for multi-source integrated navigation system using fully convolutional neural network, IET Radar, Sonar & Navigation, 1, 774–782.

Z. Zhang, J. Zhao, 2017, A deep belief network based fault diagnosis model for complex chemical processes, Computers & Chemical Engineering, 107, 395-407.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50271-2

Data-driven online scenario selection for multistage NMPC

Zawadi Mdoe, Mandar Thombre, Johannes Jäschke*

Department of Chemical Engineering, Norwegian University of Science & Technology, Sem Sælands vei 4, 7491 Trondheim, Norway.

johannes.jaschke@ntnu.no

Abstract

This paper aims at reducing the conservativeness of the robust and computationally efficient sensitivity assisted multistage nonlinear model predictive controller. The approach uses a hyperbox over-approximation for the parametric uncertainty set that often results into conservativeness. We propose the use of principal component analysis (PCA) on available process data to extract more information to tighten the approximation of the parametric uncertainty set. It is approximated by a polytope whose vertices lie on the principal components. Then we define the multistage nonlinear problem with a linear transformation of the uncertain parameters. This transformation ensures consistency with the required conditions for sensitivity assisted multistage MPC algorithm used for scenario tree pruning. Finally, the method was implemented on a case study of a system of four tanks and the controller exhibited reduced conservativeness and fast computational performance.

Keywords: Robust MPC, Dynamic optimization, Parametric uncertainty, Data-driven

1. Introduction

Model predictive control (MPC) is a model based control strategy that reoptimizes a nonlinear process system with respect to a control objective subject to constraints at each sampling time. MPC includes constraints for online decision making, and has good control performance even when the system is disturbed away from the desired reference trajectory (Rawlings & Mayne, 2009). Although MPC has inherent robustness against uncertainty, the property may break when there are significant disturbances, causing infeasibilities. As a result, robust MPC approaches have been developed. One of them was proposed by Lucia et al. (2013) and is known as the multistage MPC.

1.1. Multistage MPC

Multistage MPC explicitly considers a selection of possible future scenarios along a prediction horizon to formulate its optimization problem. The scenarios are determined by propagating from the current state to the end of the prediction horizon, a finite number of uncertain parameter realizations using a scenario tree. When the prediction horizon is long the number of scenarios in the scenario tree increases exponentially resulting into an intractable problem. Lucia et al. (2013) proposed a robust horizon where the scenario tree branching is stopped before the end of the horizon, and the uncertain parameters are kept constant until the end of the prediction horizon. The robust horizon makes the problem

practically feasible to solve but can still be expensive, especially for nonlinear problems, leading to a significant computational delay. In order to reduce the computational cost and computational delay of the multistage MPC, Thombre et al. (2020) proposed the sensitivity assisted multistage MPC. It has an algorithm to prune irrelevant scenarios from the scenario tree using NLP sensitivities in order to speed up computations. The sensitivity assisted multistage MPC is discussed further in Section 2.

1.2. Motivation

Even though multistage MPC is robust against constraint violations, it is rather conservative, resulting into performance loss. The conservativeness is highly dependent on how uncertainty set is represented. So far, its implementation has mainly been done using a hyperbox over-approximation of the uncertainty set. The over-approximation is often very poor if the true uncertainty set is ellipsoidal. Although the computational delay of multistage MPC can be reduced by the sensitivity assisted algorithms (Thombre et al., 2020), it has been implemented with an over-approximation of the uncertainty set leading to conservative control performance. However, in combination with statistical data analysis methods used for uncertainty identification, one can significantly reduce the conservativeness. Krishnamoorthy et al. (2018) suggested that detailed information on process uncertainty could be extracted via statistical data analysis to obtain more representative scenarios. Moreover, Shang & You (2019) rigourously present on calibration of approximate uncertainty sets for a scenario-based stochastic MPC in linear systems using support vector clustering with stability guarantees based on some mild assumptions. The contribution of this paper is to demonstrate how principal component analysis can be specifically applied to the sensitivity assisted multistage MPC framework in order to reduce conservativeness and retain its computational efficiency.

1.3. Notation

We assume a nonlinear system model $z_{i+1} = f(z_i, \nu_i, d_i)$ that predicts the evolution of the states z_i from time t_{k+i} with control actions ν_i and uncertain parameters d_i. Let us define the notation used in this manuscript. The time index $k \geq 0$ corresponds to sampling time t_k. A perfect state measurement is always assumed, and the state at time t_k is denoted by x_k. The time index of a model prediction is denoted by $i \in \mathbb{Z}_+$ which corresponds to sample time t_{k+i}. The nominal parameters are denoted as d_i^0 such that the nominal model becomes $z_{i+1}^0 = f(z_i^0, \nu_i^0, d_i^0)$. For a nonlinear system we obtain a nonlinear optimization problem (NLP) resuling inot a class of MPC known as nonlinear MPC (NMPC).

2. Sensitivity assisted multistage NMPC

The algorithm for the sensitivity assisted multistage NMPC (samNMPC) that performs online critical scenario selection based on NLP sensitivities was first developed by Holtorf et al. (2019). This selection is done by solving the NMPC problem for the nominal scenario together with a lower level optimization problem (LLP) that gives the parametric realizations that maximize the inequality constraints. This gives the constraints that are most likely violated. However, when the inequalities are interval bounds there exists a trivial solution to the LLP that lies on the vertices of the uncertainty hyperbox. Assume that the constraints are monotonically increasing or decreasing in the uncertain parame-

ter space. The multistage MPC problem is parametric in the disturbances thus the online critical scenarios selection is based on parametric NLP sensitivities from the nominal scenario. This algorithm determines the realization most likely to violate a constraint using the sign of the parametric sensitivity. It formulates a pruned scenario tree with only the critical scenarios and the nominal, leading to a smaller NMPC problem that is cheaper to solve. The stability and recursive feasibility properties of the samNMPC were established by Thombre et al. (2020). A sensitivity assisted multistage NMPC problem at time t_k is written as follows:

$$V_N^{sam}(x_k) = \min_{\substack{z_i^c,\nu_i^c \\ c\in\widehat{\mathbb{C}}\cup\{0\}}} \sum_{c\in\widehat{\mathbb{C}}\cup\{0\}} \omega_c\Big(\psi(z_N^c, d_{N-1}^c) + \sum_{i=0}^{N-1} \ell(z_i^c, \nu_i^c, d_i^c)\Big) + \sum_{c\in\bar{\mathbb{C}}} \omega_c\Big(\psi(z_N^0 + \Delta z_N^c, d_{N-1}^c) + \sum_{i=0}^{N-1} \ell(z_i^0 + \Delta z_i^c, \nu_i^0 + \Delta\nu_i^c, d_i^c)\Big) \tag{1a}$$

$$\text{s.t. } z_{i+1}^c = f(z_i^c, \nu_i^c, d_i^c), \quad i = 0, \ldots, N-1 \tag{1b}$$

$$z_0^c = x_k,\ z_N^c \in \mathbb{X}_f, \tag{1c}$$

$$\nu_i^c = \nu_i^{c'}, \quad \{(c, c') \mid z_i^c = z_i^{c'}\} \tag{1d}$$

$$d_{i-1}^c = d_i^c, \quad i = N_R, \ldots, N-1 \tag{1e}$$

$$z_i^c \in \mathbb{X},\ \nu_i^c \in \mathbb{U},\ d_i^c \in \mathbb{D}, \forall\, c, c' \in \widehat{\mathbb{C}} \cup \{0\} \tag{1f}$$

where the sets $\widehat{\mathbb{C}}$ and $\bar{\mathbb{C}}$ are the critical and noncritical scenario index sets, respectively and $\{0\}$ repesents the nominal scenario. $\mathbb{D} \in \mathbb{R}^{n_d}$ is the uncertain parameter set containing a finite number of realizations, $\mathbb{X} \in \mathbb{R}^{n_x}$, $\mathbb{U} \in \mathbb{R}^{n_u}$ are the feasible sets for states and inputs, respectively and $\mathbb{X}_f$ represents the terminal set. N is the prediction horizon length and N_R is the robust horizon. z_i^c and ν_i^c are the predicted state and control variable vectors for scenario c at time t_{k+i}, respectively. The stage cost function is given by ℓ, terminal cost is denoted by ψ, and ω_c represents the weights on scenario c to the objective function. The variables and constraints in problem (1) are only those associated with critical constraints, thus making the problem smaller than that of the ideal multistage NMPC with a robust horizon.

3. Data driven sensitivity assisted multistage NMPC

This section presents the main idea which is to integrate principal component analysis (PCA) and samNMPC in order to reduce its conservativeness, hence enhancing its performance. The goal is to achieve that while retaining the computational speed of samNMPC.

3.1. Principal component analysis

Principal component analysis (PCA) is a multivariate data analysis tool that reveals hidden information from data. This method evaluates the variability in the data set and identifies principal components (PC) which are the unit directions that explain the total variation in the data. As a result, PCA fits a hyperellipsoid to the data with the principal components corresponding to the ellipsoids axes. The principal components are listed in order of decreasing component variance.

Assume we have a data set with n_s samples for each uncertain parameter and the data set is a represented by a matrix $\mathbf{D} \in \mathbb{R}^{n_s \times n_d}$. Before decomposition, the data set must be mean centered and scaled because PCA is sensitive to scale differences. Let the scaled and mean centered data corresponding to $\mathbf{D}$ be denoted as $\mathbf{D_0} \in \mathbb{R}^{n_s \times n_d}$. PCA on $\mathbf{D_0}$ results in the linear model $\mathbf{D_0} = \mathbf{\Lambda C}^\top$ where $\mathbf{\Lambda} \in \mathbb{R}^{n_s \times n_p}$ is a matrix with the scores corresponding to each data sample. The scores are a projection of the data points onto the principal components directions. The matrix $\mathbf{C} \in \mathbb{R}^{n_p \times n_p}$ is made up of the weights on the original samples required to obtain the component score.

3.2. Algorithm for scenario selection using both data and NLP sensitivities

This algorithm combines PCA that determines the maximum and minimum scores in the principal component directions with the samNMPC algorithm presented by Thombre et al. (2020). In order to use the samNMPC algorithm with data, we make a linear transformation of the uncertain parameters in the optimization problem using the PCA matrix. The algorithm has the following steps

(a) Scale or normalize and mean-center the data set $\mathbf{D}$ to obtain $\mathbf{D_0}$.
(b) Perform PCA on $\mathbf{D_0}$ to determine the principal component scores $\mathbf{\Lambda}$ and the corresponding principal component matrix $\mathbf{C}$.
(c) Transform the uncertain parameter vectors d_i^c into the new orthogonal space using the matrix $\mathbf{C}$, such that, $d_i^c = \mathbf{C}\overline{d}_i^c + d_i^0$ where $\overline{d}_i^c$ are the transformed parameters.
(d) Substitute the transformation from step (c) above in problem (1) to obtain an NLP in terms of the transformed parameters.
(e) At the current time t_k, determine critical scenarios $\hat{\mathbb{C}}$ and non-critical scenarios $\bar{\mathbb{C}}$ with respect to the transformed parameters using the samNMPC algorithm.
(f) Generate a pruned scenario tree with only the critical scenarios and the nominal scenario and then solve the transformed problem (1).

4. Case study

Consider the quadtank problem with a four tank configuration from Raff et al. (2006). The levels of water in the four tanks are described by the following set of differential equations:

$$\dot{x_1} = -\frac{a_1}{A_1}\sqrt{2gx_1} + \frac{a_3}{A_1}\sqrt{2gx_3} + \frac{\gamma_1}{A_1}u_1 \qquad \dot{x_3} = -\frac{a_3}{A_3}\sqrt{2gx_3} + \frac{1-\gamma_2}{A_3}u_2$$
$$\dot{x_2} = -\frac{a_2}{A_2}\sqrt{2gx_2} + \frac{a_4}{A_2}\sqrt{2gx_4} + \frac{\gamma_2}{A_2}u_2 \qquad \dot{x_4} = -\frac{a_4}{A_4}\sqrt{2gx_4} + \frac{1-\gamma_1}{A_4}u_1$$

where the states x_i are the tank levels, the inputs u_i are pump flow rates, and the uncertain parameters are the valve coefficients γ_1 and γ_2. The controller tracks setpoint levels x_1 and x_2 with minimum input usage such that the objective is $\ell = (x_1 - x_1^*)^2 + (x_2 - x_2^*)^2 + r_1u_1^2 + r_2u_2^2$. There are constraints on x_3 and x_4 and the system experiences predefined pulses in x_1 as described by Thombre et al. (2020).

4.1. Data analysis

The uncertain parameters have a process data cloud shown in the left plot of Figure 1. PCA on the data gives $\mathbf{C} = [0.6571,\ -0.7538;\ 0.7538,\ 0.6571]$. The red circled points

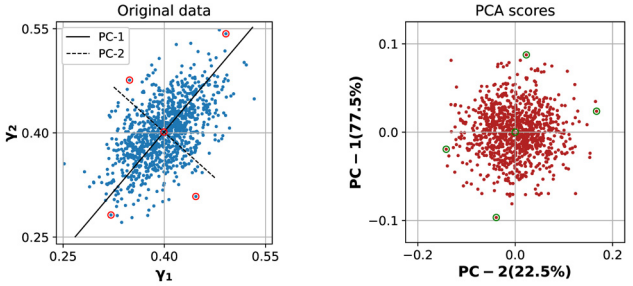


Figure 1: PCA on process data. Left shows original data, right shows the PCA scores.

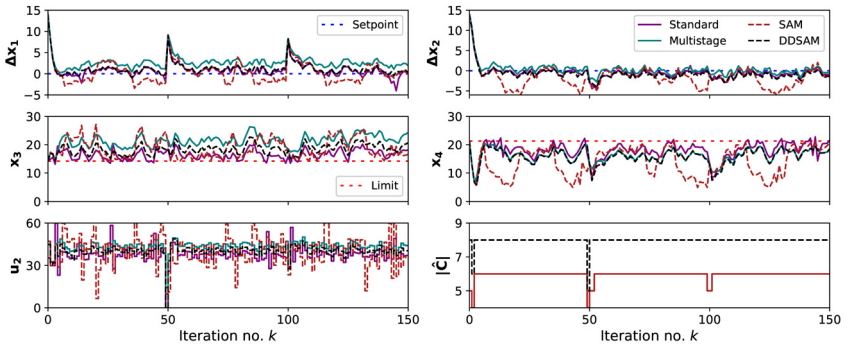


Figure 2: Comparing the control performance of the data-driven samNMPC with standard (nominal) NMPC, multistage, sensitivity-assisted multistage NMPC in the quadtank problem.

are the data points corresponding to the extreme scores on each principal component. The scores in the principal components are shown in plot on the right of Figure 1.

4.2. Results

The uncertain parameters γ_1 and γ_2 are random values generated from the multivariate distribution of the process. Then simulations were performed for both standard NMPC, multistage NMPC, samNMPC and the data-driven samNMPC. It was done for 150 iterations and the results for robust horizon $N_R = 2$ are shown in Figure 2. The tracking performance of the samNMPC is improved by the data transformation. Data-driven samNMPC tracks closer to the set point hence it is less conservative than original samNMPC and multistage NMPC. It is also robust against constraint violations for x_3 and x_4. To show the improvement of the tracking performance, we computed the accumulated cost in the simulation as shown in the bar chart on the right of Figure 3. For robust horizons lengths 1 to 3, data-driven samNMPC shows a slightly better setpoint tracking performance than the standard NMPC. It also shows a significant improvement from the original samNMPC tracking performance. In terms of computational efficiency, Figure 3 shows that the data-driven samNMPC is as fast as the original samNMPC.

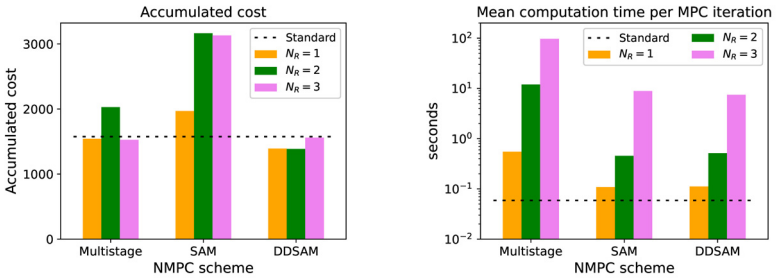


Figure 3: Comparing tracking costs (left - absolute scale) and the average computation time (right - logarithmic scale) for different robust horizons (N_R).

5. Conclusions

We have demonstrated how analysis on process data can extract more information on the uncertainty set used to formulate the sensitivity assisted multistage MPC problem. The integration of data with samNMPC requires transforming the uncertain parameters into new variables corresponding to the principal components. As a result, the samNMPC becomes less conservative while still being computationally efficient. A caveat to the method is that the uncertainty set representation is an under-approximation using a polytope whose vertices are the maximum and minimum PCA scores. There is still a chance that the process may be outside the polytope especially when a dominant principal component does not exist. However, we expect such cases to be rare and we include soft constraints on the state bounds with penalties to avoid the infeasibilities. Future work would be investigating how scaling up to a higher-order system with more uncertain parameters affects the method's performance.

References

F. Holtorf, A. Mitsos, & L. T. Biegler (2019). Multistage NMPC with on-line generated scenario trees: Application to a semi-batch polymerization process. Journ. Proc. Control.

D. Krishnamoorthy, M. Thombre, S. Skogestad, & J. Jäschke (2018). Data-driven scenario selection for multistage robust model predictive control. IFACPapersOnLine.

S. Lucia, T. Finkler, S. Engell (2013). Multi-stage nonlinear model predictive control applied to a semi-batch polymerization reactor under uncertainty, Journ. Proc. Control.

T. Raff, S. Huber, Z. K. Nagy, & F. Allgower (2006). Nonlinear model predictive control of a four tank system: An experimental stability study. Int. Symp. on Intelligent Control. IEEE.

J. B. Rawlings, & D. Q. Mayne (2009). Model Predictive Control: Theory and Design.

C. Shang, & F. You (2019). A data-driven robust optimization approach to scenario-based stochastic model predictive control, Journ. Proc. Control.

M. Thombre, Y. Zhou, J. Jäschke, & L. T. Biegler (2020). Sensitivity-assisted multistage nonlinear model predictive control: Robustness, stability and computational efficiency, Comp. & Chem. Eng.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50272-4

Data-driven Robust Model Predictive Control with Disjunctive Uncertainty for Building Control

Guoqing Hu[a,*], Fengqi You[a]

[a]*Cornell University, Ithaca, New York 14853, USA*
gh429@cornell.edu

Abstract

Model Predictive Control (MPC) has gained popularity in recent years and is widely adopted in building control. This study proposes a novel data-driven robust MPC to make the optimal heating plan, specifically for the multi-zone single-floor building. In this study, the room temperature and relative humidity (RH) will be highly valued in the optimization decision. To better incorporate RH into the state-space model (SSM), the linear relations between RH and other room temperature parameters in the thermal zones are formulated, ensuring the better linear fitting of SSM to the original nonlinear model. Afterward, k-means clustered, principal component analysis (PCA), and kernel density estimation (KDE) based data-driven uncertainty set is constructed and applied to MPC. The other three kinds of MPC's are compared to our proposed data-driven robust MPC (RMPC), including conventional RMPC, k-means clustered, data-driven RMPC, PCA and KDE based data-driven RMPC. The results demonstrate that the optimality of our proposed k-means clustered, PCA and KDE based data-driven RMPC, which consumes 9.8 % to 17.9 % less energy in controlling both temperature and RH, compared to other data-driven robust MPC's, and essentially follow the constraints which certainty equivalent MPC and conventional RMPC cannot conform.
Keywords: model predictive control, disjunctive uncertainty, multi-zone building control

1. Introduction

According to the EIA report in 2019, heating and humidity control dominate energy usage, contributing 30 % of total power consumption. Controlling temperature is essential to the building control since overheating is another problem that consumes significant energy and deteriorates the living condition.

Among all possible control methods, model predictive control (MPC) provides the new scope for controlling the building temperature, saving a tremendous amount of energy usage compared to the rule-based control strategies (Prívara et al., 2011). However, the conventional MPC does not possess the capability of hedging against the uncertainty (Shang et al., 2019), i.e. being applied under stochastic conditions (Ning and You, 2019). In building control, weather information can never be perfectly predicted, and thus can be treated as the sources for uncertainties (Shang et al., 2020). Consequently, it remains a knowledge gap needs to be filled with the new designed MPC which is not only robust to the disturbances from uncertainty, but also can avoid the "over-conservative" problem proposed by Chen et al. (2021). Therefore, we focus on developing the better control strategy to multi-zone building's room temperature and RH under realistic condition, k-

mean clustered, principal component analysis (PCA) and kernel density estimation (KDE) based data-driven RMPC (KM-PKDDRMPC). We apply this model to the multi-zone building's SSM, which incorporates both room temperature and RH. In this work, the SSM of the building is generated from based on both building element construction and the study of the dynamic airflow within the building. Afterward, the uncertainty set is constructed based on the historical forecast error to the weather information, i.e., the differences between forecast and real-measured values. This uncertainty set can be further clustered by the k-means algorithm, and PCA combined with KDE can return the polyhedral-shaped applied to the RMPC. The optimization problem at each control horizon is solved using affine disturbance policy (ADF). The contributions of this paper are summarized below:

- A novel data-driven robust model predictive control framework with disjunctive uncertainty to control the multi-zone building's room temperature and RH;
- A simulation of multi-zone building's temperature and RH control based on actual weather data demonstrates better control performance of KM-PKDDRMPC comparing to other MPC's

2. Model formulation

2.1. Complete state-space model

The BRCM MATLAB toolbox is adopted for finding the state space matrix (SSM). BRCM can generate the linear resistance-capacitance models from self-designed building geometry construction. The dynamic multi-input multi-output system is given by:

$$x_{t+1} = Ax_t + B_u u_t + B_v v_t + B_w w_t \tag{1}$$

Where A is the state matrix that correlates state variables x_t to SSM. The state variables returned from BRCM are room temperature, wall temperature, floor temperature and ceil temperature. B_u, B_v, B_w are control input matrix, disturbance matrix, and uncertainty matrix, respectively, corresponding to u_t, v_t, w_t, which are control input, disturbances, and uncertainty. The control inputs include heater, radiator, humidifiers and dehumidifiers; the disturbances are from ambient temperature and ambient RH condition. Uncertainties are the forecasted temperature and RH errors. Meanwhile, RH within each room is calculated based on the air dynamic within the building (Cengel, 1997). The mass of airflow is initially found as:

$$m_{air,t-1} = \frac{Q_{t-1}}{c_p \Delta T} \tag{2}$$

ΔT is calculated as follows:

$$\Delta T = \max\left(\left(T_{room,t} + \delta T - T_{air,t}\right), 0\right) \tag{3}$$

where δT is temperature difference of room and air heating unit (AHU). Unlike in previous research, $m_{air,t-1}$ is not regarded as a constant because the simulation process is conducted in the winter season. The constant intake airflow rate indicates that the room is constantly exchanging the air with a colder ambient environment. The heater, most of the time, is active to maintain room temperature within the thermal comfort standard. Alternatively, we assume the difference between the room temperature and heated air from the AHU is constant. Subsequently, the heating airflow can be turned off when

heating is not necessary. When the mass of airflow is calculated, the mass of water vapor brought by airflow can be found by the following equation:

$$m_{AC,in,t-1} = \rho_{AC,in,t-1} \cdot RH_{out,t-1} \cdot \frac{m_{air,t-1}}{\rho_{air}} \tag{4}$$

And so can be found the mass of water vapor taken away by airflow:

$$m_{AC,out,t-1} = \rho_{water,sat,t-1} \cdot RH_{t-1} \cdot \frac{m_{air,t-1}}{\rho_{air}} \tag{5}$$

Where SVD values are found through equation *f*, which is a linear equation of SVD values over temperature (T) expressed as follows:

$$\rho_{water} = f(T) = 1.0272T - 1.8959 \tag{6}$$

Afterward, the mass of water vapor stored in each room can be found as:

$$m_{water,t} = \rho_{water,sat,t-1} \cdot RH_{t-1} \cdot V_{room} + m_{hum} + m_{AC,in,t-1} - m_{dehum} - m_{AC,out,t-1} \tag{7}$$

Eventually, RH values within each room at *t* can be found, which is the ratio of absolute and SVD:

$$\rho_{abs,t} = \frac{m_{water,t}}{V_{room}},\ \rho_{sat,t} = f\left(T_{room,t}\right),\ RH_t = \frac{\rho_{abs,t}}{\rho_{sat,t}} \tag{8}$$

At this point, the RH values within each room can be found based on the room temperature, control input and room volume. The next step is to add system identification toolbox found in MATLAB to obtain the SSM required for the MPC. The testing data, instead of training data, is used to ensure the feasibility of SSM to be applied in simulation within the real condition. The average value of mean absolute percentage error (MAPE) for RH in all rooms is 4.65 % and the average MAPE for temperature in all rooms is 0.95 %, indicating this model is acceptable for the MPC problem.

2.2. PCA and KDE based data-driven uncertainty sets clustered by K-means algorithm

Disjunctive uncertainty sets are constructed to better learn the trend of the uncertainty data (Ning and You, 2017). Therefore, the K-means clustering method is adopted in this work to cluster the uncertainty into multiple groups. The groups are identified by minimizing the sum of intracluster variances, i.e., squared Euclidean distance:

$$D^* = \arg\min\left(\sum_{i=1}^{k}\sum_{w\in D_i}\left\|w-\mu_i\right\|^2\right) \tag{9}$$

Despite multiple groups of uncertainty data, the traditional norm-based uncertainty set cannot be applied directly to deal with the uncertainty data with varied structure and complexity (Ning and You, 2021). Therefore, PCA and KDE are adopted for coping with the data with polyhedral shapes. PCA can then maximise the variance of the uncertainty under the same scale. The covariance matrix can be approximated as:

$$S_i = \frac{1}{N-1} w_i^T w_i \tag{10}$$

As the covariance matrix S_i can be further decomposed as $S_i = Q_i \Lambda_i Q_i^T$, where Q_i's column contains all the eigenvectors, corresponding to the eigenvalues stored in the diagonal matrix Λ_i. The individual eigenvalue will represent the variance of this axis if data is projected on this eigenvector.

Finally, it can be further studied the distributional information of the uncertainty dataset within each component *j* within the cluster *k* via the KDE approach:

$$f_{j,k} = \frac{1}{N}\sum_{n=1}^{N} K(\xi_{j,k}, p_{j,k}^{(n)}) \tag{11}$$

With probability density function, the cumulative density function will be written as follows:

$$F_{j,k}^{-1}(\alpha) = \min\{\xi_{j,k} \mid F_{j,k}(\xi_{j,k} \geq \alpha)\} \tag{12}$$

where α is the pre-specified small quantile parameter, ranging from 0 to 0.5, and ξ is the inferred latent variable. The uncertainty set W_k within cluster *k* can be formulated by introducing forward and backward deviation variable z^+ and z^- (Ning and You, 2018):

$$\mathbb{W}^k = \left\{ \mathbf{w}^k \in R^H \;\middle|\; \begin{array}{l} \mathbf{w}^k = \hat{\boldsymbol{\mu}}_k + Q_k\boldsymbol{\xi}_k,\ \boldsymbol{\xi}_k = \underline{\boldsymbol{\xi}}_k z^- + \overline{\boldsymbol{\xi}}_k z^+ \\ \mathbf{0} \leq z^+, z^- \leq \mathbf{1},\ \ z^+ + z^- \leq \mathbf{1},\ \ \mathbf{1}^T\left(z^+ + z^-\right) \leq \Gamma \\ \underline{\boldsymbol{\xi}} = \left[\hat{F}_{1,k}^{-1}(\alpha),...,\hat{F}_{1,k}^{-1}(\alpha)\right]^T \\ \overline{\boldsymbol{\xi}} = \left[\hat{F}_{1,k}^{-1}(1-\alpha),...,\hat{F}_{1,k}^{-1}(1-\alpha)\right]^T \end{array} \right\} \tag{13}$$

3. Control strategy

The next step is to develop the optimization problem to get the control strategy to the multi-zone building. To ensure the tractability of the RMPC optimization problem, ADF is adopted to get control input u_t based on past disturbances. The equation is expressed as following (Goulart et al., 2006):

$$u_i = h_i + \sum_{j=0}^{i-1} M_{i,j} w_j,\ \forall w \in \mathbb{W}^k \tag{14}$$

where M is regulated as follows:

$$M = \begin{bmatrix} 0 & \cdots & \cdots & 0 \\ M_{1,0} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ M_{H,0} & M_{H,1} & \cdots & 0 \end{bmatrix} \tag{15}$$

Only the first u_0 will be applied for the control to the model and the rest will be discarded. The optimization problem with ADF can be formulated as follows (Shang et al., 2017):

$$\begin{aligned} &\min \sum_{i \in B_u} c_i u_i + \lambda^T L \lambda \\ &\text{s.t. } F_u[Mw + h] \leq f_u, \qquad \forall w \in \mathbb{W}^k \\ &\quad F_x[Ax_0 + B_u h + B_v v + (B_w + B_u M)w] \leq f_x + \lambda, \quad \forall w \in \mathbb{W}^k \end{aligned} \tag{16}$$

where F_x, F_u, f_x, f_u represent the state variable constraints matrix, control input constraints matrix, constraints for state variables, and constraints for the input. *L* is the weighted cost matrix that penalizes the violation to the constraints. Λ is the slack variable that allows some extent of violation to the hard constraints (Jia et al., 2020).

4. Case study

In this study, the single-floor multi-zone building located in Ithaca, New York, USA is selected for the simulation of close-loop data-driven RMPC to control the temperature and RH in each individual room. The constraints for the control conditions are: For the room temperature should be within 15 ℃ to 25 ℃, and RH should sit in between 30 % to 60 %, according to ASHRAE Standard 62-2001.

The model was simulated from 0:00 AM, November 1st, 2016 to 0:00 AM, on November 8th, 2016, ranging from precisely one week. The initial conditions for temperature values in all rooms are 21 ℃ and RH values are 40 %. One of rooms' results are selected for demonstration, as shown in Fig 1. Based on the result, both certainty equivalence MPC (CEMPC) and RMPC violate the constraints more severely. CEMPC which only considers the deterministic conditions, fails to compose the strategy against the prediction error from ambient temperature and RH. Meanwhile, the RMPC fails to obey the RH constraints, indicating an irregular shape of the uncertainty data of RH. On the other hand, the rest three control strategies can be more conservative in maintaining both temperature and RH within the constraints. KM-DDRMPC will be the most conservative one since there is nearly no violation at all, but, it will have the highest power consumption across all control methods. On the other hand, though there are slightly more violation cases and more computation time, KM-PKDDRMPC will draw significantly less power in controlling the temperature and RH compared to KM-PKDDRMPC and PCA coupled with KDE based data-driven RMPC (PKDDRMPC).

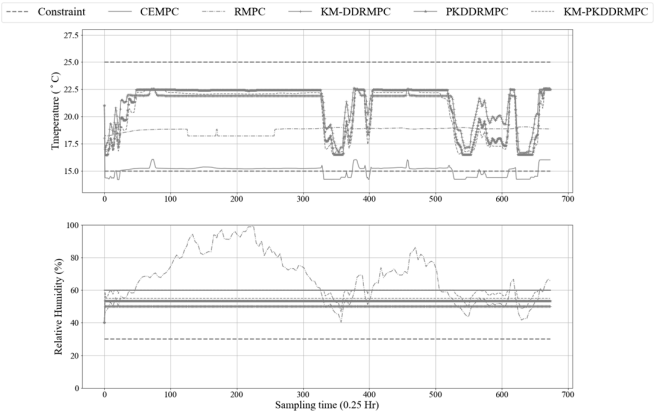


Figure 1. Control profile in Ithaca, New York, in the first week of November 2016

5. Conclusions

In this work, we develop a KM-PKDDRMPC framework for the multi-zone building SSM, which includes indoor temperature and RH control. In order to maintain temperature and RH within the comfortable range, KM-PKDDRMPC is capable of handling the uncertainty sets from temperature and RH forecast. The steady-state system with RH is constructed with the help of system identification. Then the optimization problem can be further developed with the SSM and disjunctive uncertainty sets. The proposed KM-PKDDRMPC was compared with the CEMPC and other MPC strategies, including RMPC, KM-DDRMPC, PKDDRMPC. The result demonstrated that the

proposed KM-PKDDRMPC has outperformed the rest from the overall perspective, using 17.9 % less power consumption than KMDDRMPC and 9.8 % fewer compared to PKDDRMPC. Though CEMPC and RMPC have used less power, the high violation rate will exclude them from the final consideration to the practical application.

6. Nomenclature

SVD – saturated vapor density
VD –vapor density
$m_{air,t-1}$ – mass of airflow at t-1, kg
Q_{t-1} – heat input at t-1, J
c_p – specific heat of air, kJ/(kg-K)
ΔT – temperature change, ℃
$T_{room,t}$ – room temperature at t, ℃
$T_{air,t}$ – ambient temperature at t, ℃
ρ_{air} – air density, kg/m^3
RH_{t-1} – relative humidity in room at t-1
$RH_{out,t-1}$ – ambient relative humidity at t-1
$\rho_{water,sat,t-1}$ – SVD of T_{room} at t-1, g/m^3
$\rho_{AC,sat,t-1}$ – SVD of T_{air} at t-1, g/m^3
V_{room} – room volume, m^3
$m_{water,t}$ – mass of VD at t, kg
m_{hum} –mass of VD from humidifier, kg
m_{dehum} – mass of VD taken by dehumidifier, kg
$m_{AC,in,t-1}$ – mass of VD from air circulation at t-1, kg
$m_{AC,out,t-1}$ – mass of VD taken by air circulation at t-1, kg
$\rho_{abs,t}$ – absolute VD density at t, kg/m^3
$\rho_{sat,t}$ – SVD at t, kg/m^3

References

ISO, 2001, Standard 62-2001, Ventilation for acceptable indoor air quality. American Society of Heating, Refrigerating and Air Conditioning Engineers.
Y. A. Cengel, 1997, Introduction to thermodynamics and heat transfer, McGraw-Hill New York.
W.-H. Chen, C. Shang, S. Zhu, et al., 2021, Data-driven robust model predictive control framework for stem water potential regulation and irrigation in water management. Control Engineering Practice, 113, 104841.
EIA, 2019, Use of energy explained[Online]. U.S. Energy Information Administration.
P. J. Goulart, E. C. Kkerrigan, J. M. Maciejowski, 2006, Optimization over state feedback policies for robust control with constraints. Automatica, 42, 523-533
R. Jia, F. You, 2020, Multi-stage economic model predictive control for a gold cyanidation leaching process under uncertainty. AIChE Journal, 67, e17043
C. Ning, F. You, 2017, Data-Driven Adaptive Nested Robust Optimization: General Modeling Framework and Efficient Computational Algorithm for Decision Making Under Uncertainty. AIChE Journal, 63, 3790-3817.
C. Ning, F. You, 2018, Data-driven decision making under uncertainty integrating robust optimization with principal component analysis and kernel smoothing methods, Computers & Chemical Engineering, 112, 190 – 210.
C. Ning, F. You, 2019, Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming. Computers & Chemical Engineering, 125, 434-448.
C. Ning, F. You, 2021, Online learning based risk-averse stochastic MPC of constrained linear uncertain systems. Automatica, 125, 109402.
C. Shang, W.-H. Chen, A.D. Stroock, et al., 2020, Robust Model Predictive Control of Irrigation Systems With Active Uncertainty Learning and Data Analytics. IEEE Transactions on Control Systems Technology, 28, 1493-1504.
C. Shang, X. Huang, F. You, 2017, Data-driven robust optimization based on kernel learning. Computers & Chemical Engineering, 106, 464-479.
C. Shang, F. You, 2019, A data-driven robust optimization approach to scenario-based stochastic model predictive control. Journal of Process Control, 75, 24-39.
C. Shang, F. You, 2019, Data Analytics and Machine Learning for Smart Process Manufacturing: Recent Advances and Perspectives in the Big Data Era. Engineering, 5, 1010-1016.
S. Privara, J. Siroky, L. Ferkl, J. Cigler, 2011, Model predictive control of a building heating system: The first experience. Energy and Buildings, 43, 564-572

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50273-6

Low-Dimensional Input and High-Dimensional Output Modelling Using Gaussian Process

Jiawei Tang, Xiaowen Lin, Fei Zhao, Xi Chen*

State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China
xi_chen@zju.edu.cn

Abstract

In this paper, a unified low-dimensional input and high-dimensional output modelling method is proposed to deal with complex molecular simulation and design problems. First, a convex optimization framework is constructed to decompose vertically stacked molecular weight distribution (MWD) matrix into low-rank and sparse parts, while the intrinsic structure can be explored, and abnormal points can be eliminated. Then, considering the correlations between independent output channels, an effective coregionalization kernel is adopted in Gaussian Process (GP) to implement the low-dimensional multi-output tasks. The whole procedure consists of data filtering, feature compressing and multi-output GP, which is named by DF-MGP. Case study of an ethylene homo-polymerization with the Ziegler-Natta catalyst system shows the effectiveness of the proposed DF-MGP strategy.

Keywords: Multi-output regression; Gaussian Process; coregionalization kernel; molecular weight distribution

1. Introduction

Machine learning (ML) has been widely applied in almost all areas of science. They are great at problems when inputs lie in high-dimension space and outputs lie in low-dimension space. However, this situation is inverse in molecular simulation, which macro manipulation space is much smaller than micro molecular space. Because of the complex and time-consuming features of molecular simulation, some scholars have utilized ML to improve such phenomenon. Elton et al. (2018) proved that ML techniques can be used to predict CNOHF energetic molecules from their molecular structures. Afzal et al. (2019) applied ML to develop a data-driven prediction model in the study of 1.5 million organic molecules. Moreover, main challenges of the practical applications are missing data, especially when some feature values cannot be observed, presence of noise, and coupling interactions between multiple target variables. To tackle these issues, multi-output regression methods are presented, which are capable to yielding better predictive performance than single-output methods. The multi-output regression methods aim to simultaneously predict multiple real-valued outputs. Kocev et al. (2009) applied ML methods to predict multiple targets describing conditions or quality of vegetation. Tuia et al. (2011) estimated different biophysical parameters from remote sensing images simultaneously.

Furthermore, problems with low-dimensional input and high-dimensional output feature make multi-output regression extremely difficult. For example, to represent a molecular weight distribution of a polymer usually needs a chain length as large as 10^5. Even after discretization, normally 100 grids are required to represent such a curve as measured by Gel Permeation Chromatograp (GPC). This kind of problems requires an elaborate and

elegant technique to process data. In this paper, we propose to apply compressed sensing (CS) method, which is well known in machine vision, to decompose the MWD matrices into low-rank and sparse parts, while a novel strategy is developed to address data missing and noise. An ethylene homo-polymerization with the Ziegler-Natta catalyst system (Lin et al. 2021) is studied to show the effectiveness of the proposed method.

2. Related works

2.1. Decomposing Sparse and Low-Rank Matrices

Compressed sensing is widely studied in computer science, which has been proved to be a complete technique for signal treatment and analysis. It can recover the signal from few of samples. In addition, the compressed sensing technique can also be applied to matrices when signals are stacked vertically in sequence. The general formula is shown as follows.

$$\text{Minimize } \|L\|_* + \lambda\|S\|_1 \qquad \text{s.t. } \mathrm{P}_Q[L+S] = \mathrm{P}_Q[Y] \tag{1}$$

where $\lambda > 0$ is a positive weight parameter; the subscripts, * and 1, denote nuclear norm and L1 norm, respectively; Y is the unprocessed data, P_Q is the projection operator. L and S represent the optimized low-rank and sparse matrices. Because the linear constraint and objective function are convex, the convex problem shown in Eq. (1) can be solved by alternating directions method of multipliers (ADMM).

2.2. Gaussian Process

Gaussian process (GP) is a kind of non-parametric Bayesian approaches (Schulz, E et al. 2009). Theoretically, it can capture a variety of relations between inputs and outputs by using an infinite number of parameters, while determining the level of complexity by means of Bayesian inference. Generally, GP can be formulated in traditional parametric weight space or non-parametric Bayesian function space. A univariate linear Bayesian regression formula is presented as follows:

$$p(f_*|x_*, X, \boldsymbol{y}) = N\left(\frac{1}{\sigma_\epsilon^2} x_*^T A^{-1} X \boldsymbol{y}, x_*^T A^{-1} x_*\right) \tag{2}$$

where *f* is an unknown function, which maps inputs x to outputs y: *f*: $X \to Y$. For the sake of simplicity, the dimension of Y is set to one. x_* is a test case; f_* is latent variable output; σ_ϵ^2 is variance of Gaussian noise. $A = \sigma_\epsilon^2 X X^T + \Sigma_p^{-1}$; X and $\boldsymbol{y}$ represent the training data.

GP is a probabilistic ML, which can predict uncertainties of f_*. We can specify different probability density and construct kinds of likelihood probability density. Moreover, many approximation inference methods can be used for modelling and Bayesian optimization.

3. High-dimensional-output modelling method

In this section, we use an ethylene homo-polymerization with Ziegler-Natta catalyst system to introduce the framework of proposed DF–MGP strategy. As shown in Fig. 1, it mainly contains four steps, i.e., data-collection, feature extraction, kernel design and inference chosen.

Step 1: Data collection

The manipulated variables (MV) involved in this case are hydrogen feed(H2), monomer feed(M), hexane feed (C6H14), catalyst feed (Cp), cocatalyst feed (A) and temperature (T). Latin hypercube sampling (LHS) is applied to sample points in six-dimensional space.

Step 2: Decomposing low-rank and sparse parts

Due to the numerical calculation error, some MVs can lead to a non-convergent MWD in the simulation process. A few outliers can degrade the performance of ML. To tackle this issue, the convex optimization model described in section 2.1 is constructed to minimize the L1 and nuclear norm of MWD matrix to explore the intrinsic structure and cut off the abnormal points. Considering the high dimensionality, the singular value decomposition (SVD) method is performed to recover the best low-rank approximation matrix of MWD and complete feature compressing.

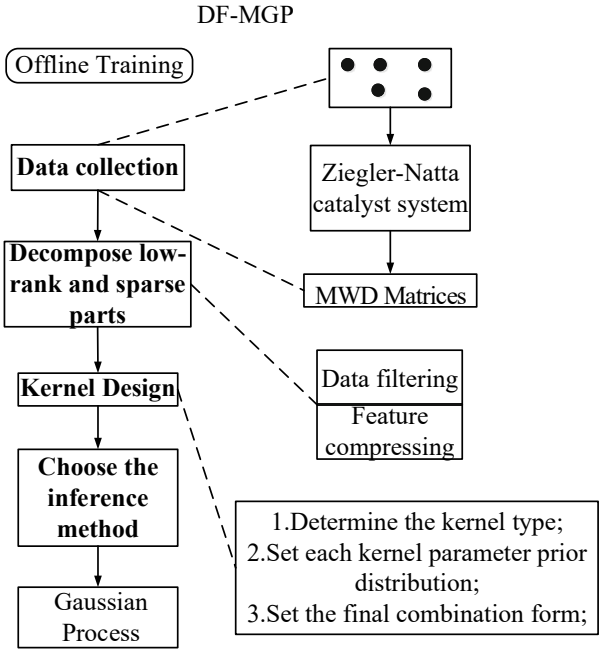


Fig. 1. Framework of the proposed DF–MGP strategy

Step 3: Kernel design

As a nonlinear regression method, the GP models involve many kernel functions. The kernel function defines a covariance function, which can describe the similarity of random variables. There are many kinds of kernel functions, such as linear, Matern32, periodic, polynomial, exponent and radial basis functions. Considering the requirements of multi-output analysis, the coregionalization kernel function is utilized to describe correlation of the outputs. Once the type of kernel is specified, the hyper-parameters involved can be found by Bayesian inference. Then, the corresponding prior of hyper-parameters can be determined. In addition, different kernel functions can be combined, such as sum, product, vertical scaling, warping.

Step 4: Inference chosen

Based on the Gaussian prior and Gaussian likelihood functions, the conditional posterior of f_* can be evaluated by Eq. (2). However, if the kernel function and likelihood function are changed, the conditional posterior will be rather intractable. Thus, different inference

methods are proposed, such as Laplace approximation, expectation propagation, Markov Chain Monte Carlo (MCMC). Different methods can lead to different generalization performance; thus, the step of inference chosen is crucial. Note that the continuous and categorical hyper-parameters involved in above steps are chosen by trial and error, and there also exist some empirical ways to assist in hyper-parameters tuning.

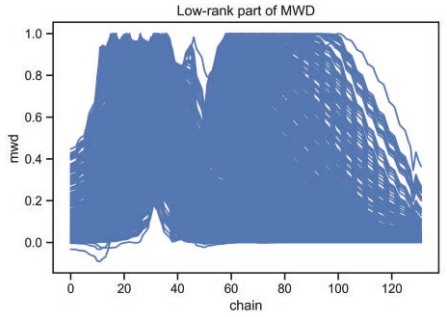
4. Case studies

4.1. Homo-polymerization Ziegler-Natta Catalyst System

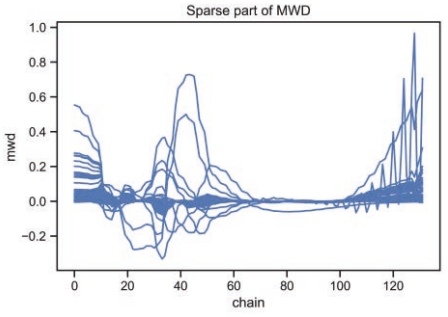
Molecular weight distribution (MWD) is a critical index of the optimization and control for industrial polymerization processes, which indicates the processability and properties of polymers. MWD is a probability distribution function of chain length, which can be predicted by polymerization mechanisms. In this work, an ethylene homo-polymerization with the Ziegler-Natta catalyst system is demonstrated. 1549 sample points are generated in the MV space, and the corresponding MWDs are generated through a kinetic modeling and simulation method (Lin et al. 2021).

4.2. Projection of MWD Matrices on Intrinsic-Low-dimensional Matrices

Since the MWD data is originally defined in a 132-dimensional space, we need to project it onto the low-dimensional subspace. As shown in Fig. 4(a), the raw data have abnormal points because of numerical calculation error. In Fig. 2(a), the low-rank part has extracted most features of the curve. Non-zero parts shown in Fig. 2(b) represent the noise and abnormal points of MWD. Therefore, in Fig. 3, we set an upper threshold 51 to eliminate those parts. Fig. 4(b) demonstrates effectiveness of our proposed data processing method through which the valid data can be identified.



(a)



(b)

Fig. 2. Low-rank part and sparse part visualization.

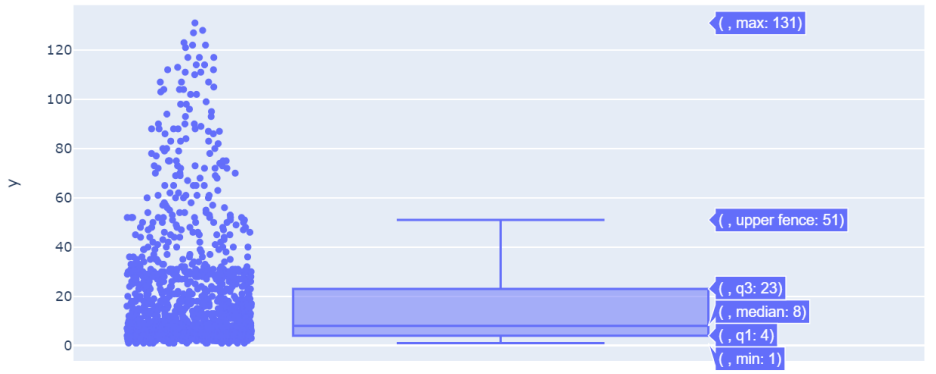
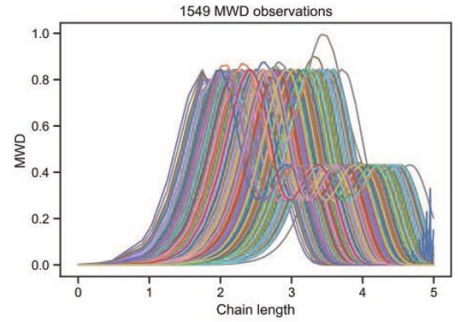
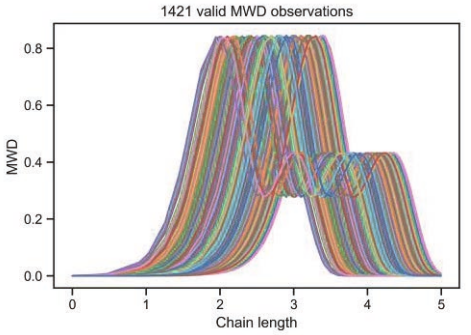


Fig. 3. Box plot of non-zero positions in sparse part



(a)



(b)

Fig. 4. Visualization results of MWD. (a) Raw data (b) Valid data

4.3. LS-MGP Prediction

MCMC inference is implemented to optimize the parameters of GP. 80% of the valid data are used to train the offline model, and the rest are used for validation. Fig. 5 shows one of the visualization results. The shaded part is the uncertainty evaluation of prediction bounded by 0.05 and 0.95 quantiles. The average root mean square error is 0.01. The first solid line in the legends represents the mean value of GP, and the second marker line with left triangle indicates the result of validation point. The rest two dash dot lines show the region of 0.05 and 0.95 quantiles

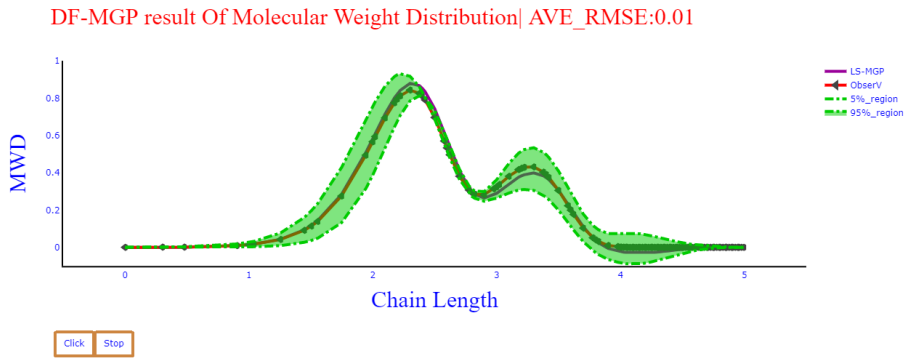


Fig. 5. Prediction of a validation sample

5. Conclusions

Through data processing and feature compressing, the proposed DF-MGP strategy can accurately predict the micro-structure of polymers with different manipulating conditions. Due to the flexibility of GP and powerful performance of MCMC inference, the proposed method is also applicable to other high-dimensional multi-output problems. The current work is a beginning of the future research for molecular design and optimization.

References

Elton, D. C., Boukouvalas, Z., Butrico, M. S., Fuge, M. D., & Chung, P. W. (2018). Applying machine learning techniques to predict the properties of energetic materials. Scientific reports, 8(1), 1-12.

Afzal, M. A. F., Sonpal, A., Haghighatlari, M., Schultz, A. J., & Hachmann, J. (2019). A deep neural network model for packing density predictions and its application in the study of 1.5 million organic molecules. Chemical science, 10(36), 8374-8383.

Kocev, D., Džeroski, S., White, M. D., Newell, G. R., & Griffioen, P. (2009). Using single-and multi-target regression trees and ensembles to model a compound index of vegetation condition. Ecological Modelling, 220(8), 1159-1168.

Tuia, D., Verrelst, J., Alonso, L., Pérez-Cruz, F., & Camps-Valls, G. (2011). Multioutput support vector regression for remote sensing biophysical parameter estimation. IEEE Geoscience and Remote Sensing Letters, 8(4), 804-808.

Lin, X., Chen, X., & Biegler, L. T. (2021). Generalized initialization for the dynamic simulation and optimization of grade transition processes using two‐dimensional collocation. AIChE Journal, 67(1), e17053.

Lin, X., Chen, X., Biegler, L. T., & Feng, L. F. (2021). A modified collocation modeling framework for dynamic evolution of molecular weight distributions in general polymer kinetic systems. Chemical Engineering Science, 237, 116519.

Schulz, E., Speekenbrink, M., & Krause, A. (2018). A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. Journal of Mathematical Psychology, 85, 1-16.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50274-8

Piecewise Smooth Hybrid System Identification for Model Predictive Control

Ilya Stolyarov[c], Ilya Orson Sandoval[a], Panagiotis Petsagkourakis[b], Ehecatl Antonio del Rio-Chanona[a*]

[a]*Department of Chemical Engineering, Imperial College London, London - U.K.*
[b]*Centre for Process Systems Engineering, University College London, London - U.K.*
[c]*Department of Computing, Imperial College London, London - U.K.*
[*]*a.del-rio-chanona@imperial.ac.uk*

Abstract

Complex systems which exhibit different dynamics based on their operating region pose challenges for data driven control because a single global model may not capture the varying dynamics of the system. One solution is to use hybrid system identification to learn the location of operating regions and dynamics within each region from data, yielding a more accurate multi-model of the system. This article proposes a novel method of hybrid system identification through spectral clustering with a custom similarity function. A case study of a chemical process illustrates benefits of this approach for Model Predictive Control.

Keywords: Hybrid System Identification, Model Predictive Control, Data Driven Models

1. Introduction

Systems which exhibit different dynamics based on their operating region, termed hybrid systems, are prevalent within all areas of engineering, ranging from a four-stroke cycle of a combustion engine to chemical processes controlled by a thermostat (Lauer 2019). These systems pose challenges for data driven control, since a single model is often inadequate in capturing the varying dynamics of the system. However, the location or even a number of operating regions may be unknown *a priori*, and hence hybrid system identification is concerned with learning the location and local model of each operating region. This can lead to a more accurate piecewise system model and improve control of hybrid systems.

Due to the presence of discrete operating regions and continuous dynamics, hybrid system identification is an NP-hard mixed integer optimization problem (Lauer 2019), and as such, local methods or relaxation approaches must be employed. Most of the currently available methods further constrain the problem by assuming that the underlying system is linear, leading to piecewise affine (PWA) system identification (Lauer 2019, Ohlsson 2013). On the other hand, only a few algorithms tackle identification of general piecewise smooth (PWS) systems (Lauer 2019, Lauer 2014, Lee 2017) that can be used within the emergent field of nonlinear control. This article proposes a novel PWS system identification approach based on spectral clustering, with ramifications for control illustrated by a case study of a Continuously Stirred Tank Reactor (CSRT).

2. Hybrid System Identification

2.1. Problem Definition

For a hybrid system with states $\boldsymbol{x} \in \mathbb{R}^{d_x}$, controls $\boldsymbol{u} \in \mathbb{R}^{d_u}$ and p operating regions, a latent function $q(\cdot,\cdot) : \mathbb{R}^{d_x} \times \mathbb{R}^{d_u} \to \{1, \dots, p\}$ defines the operating region for a given set of states and controls. Assuming a full state feedback, the discrete state space model of the system then exhibits piecewise behaviour as described in Eq. (1), where $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ is some Gaussian noise.

$$\boldsymbol{x}_{k+1} = f_{q(\boldsymbol{x}_k, \boldsymbol{u}_k)}(\boldsymbol{x}_k, \boldsymbol{u}_k) + \boldsymbol{\varepsilon} \equiv f_{q(\boldsymbol{\zeta}_k)}(\boldsymbol{\zeta}_k) + \boldsymbol{\varepsilon} \tag{1}$$

Given a training dataset $\mathcal{D} = \left\{\boldsymbol{x}_k^{(i)}, \boldsymbol{u}_k^{(i)}, \boldsymbol{x}_{k+1}^{(i)}\right\}_{i=1,\dots,N-1}$, the goal of hybrid system identification is to learn the latent function $q(\cdot,\cdot)$ as well as each local model $\{f_j(\cdot,\cdot)\}_{j=1,\dots,p}$. Notably, the number of regions p may be unknown and hence must also be learned. To simplify the notation, let's further aggregate the state and control vectors into a single vector $\boldsymbol{\zeta}_i = \left[\boldsymbol{x}_k^{(i)}, \boldsymbol{u}_k^{(i)}\right]^T$ and rename $\boldsymbol{x}_{k+1}^{(i)}$ as $\boldsymbol{y}_i$, which allows us to reformulate Eq. (1) and rewrite the training dataset in a more familiar input-output form $\mathcal{D} = \{\boldsymbol{\zeta}_i, \boldsymbol{y}_i\}_{i=1,\dots,N}$

2.2. Challenges of Hybrid System Identification

The main difficulty of hybrid system identification comes from the dual nature of the problem: simultaneously assigning datapoints to regions and learning the model within each region (Lauer 2014). Moreover, the problem is naturally ill defined if the number of regions is unknown, which can be observed by a trivial solution of assigning each datapoint to a different region. The problem can be simplified by assuming that each local model $f_j(\cdot,\cdot)$ is linear, hence deriving a PWA system model. This is attractive because nonlinear systems can be approximated with sufficiently many local affine models and classical linear control can then be employed (Lauer 2019). However, this limits the use of hybrid system identification for nonlinear system control strategies such as nonlinear MPC.

In PWS system identification, on the other hand, the only assumption made about the nature of the underlying local models is that they are smooth. Hence, if the complexity of these models is not limited, a single flexible model can overfit the training dataset. Based on this intuition, (Ohlsson 2013) derives a regularization approach to PWA identification, which is then extended to PWS systems by considering functions in a Reproducing Kernel Hilbert Space (RKHS) in (Lauer 2014). These approaches work by estimating a regularized parametric local model for each datapoint, hence projecting it into some parameter space, where *k*-means clustering can be used to identify operating regions. Although, this works well for linear models where the dimensionality of the parameter space is small, extension to PWS systems faces the difficulty of the large dimensionality of the parameter space associated with functions in RKHS (i.e. dimensionality is equal to the number of training samples).

2.3. Spectral Clustering for Hybrid System Identification

An alternative to clustering points based on compactness in some parameter space, is to employ connectedness based clustering algorithms such as spectral clustering (von Luxburg 2007), which aims to identify clusters such that the similarity of points within a

cluster is maximized and the similarity of points between clusters is minimized. This is done by using a positive similarity function $s(\cdot,\cdot) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^+$ to measure the similarity between any two points in the dataset. Notably, custom similarity function can be used to suit the application needs.

This approach is especially attractive for hybrid system identification because the similarity function can be viewed as a surrogate for the latent region function $q(\cdot)$, such that $s(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j)$ is large if $q(\boldsymbol{\zeta}_i) = q(\boldsymbol{\zeta}_j)$, and small otherwise. Hence, the problem is reformulated as finding such positive similarity function. A thorough background on spectral clustering can be found in (von Luxburg 2007).

3. Local Predictive Clustering

3.1. Similarity Function

If a group of points is generated by the same local model, then they will contain information about each other. Hence, if a set of points is used to construct a local model, which can then accurately predict the value at another point, it is likely that the prediction and at least some training points belong to the same region. This intuition can be encoded into a similarity function by first assuming that the global training dataset $\mathcal{D} = \{\boldsymbol{\zeta}_i, \boldsymbol{y}_i\}$ is sampled uniformly (i.e. at regular intervals) along its input dimensions and defining a standardized distance metric $d(\cdot,\cdot) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^+$ shown in Eq. (2), where $\boldsymbol{\zeta}_i[n]$ refers to the n^{th} component of the vector $\boldsymbol{\zeta}_i$.

$$d(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j) = \sqrt{\sum_{n=1}^{d} \frac{(\zeta_i[n] - \zeta_j[n])^2}{V_\zeta[n]}} \tag{2}$$

where $V_\zeta[n] = Variance(\{\boldsymbol{\zeta}_i[n]\}_{i=1,\dots,N})$

Then for each point $\boldsymbol{\zeta}_i$, a local dataset $\mathcal{D}_i^{(l)}$ containing its k_l neighbours can be used to learn a local model $h_i(\cdot) : \mathbb{R}^d \to \mathbb{R}^{d_x}$. The neighbourhood of the model is then defined as complimentary dataset $\mathcal{D}_i^{(p)}$ such that it includes any point among k_p neighbours of any point in $\mathcal{D}_i^{(l)}$ (but excluding any point in $\mathcal{D}_i^{(l)}$). The local model $h_i(\cdot)$ is then used to make predictions for all points in $\mathcal{D}_i^{(p)}$ giving rise to some prediction RMSE for each point in $\mathcal{D}_i^{(p)}$. Intuitively, if that error is small, then it is likely that point in $\mathcal{D}_i^{(p)}$ is from the same region as some points in $\mathcal{D}_i^{(l)}$.

However, since local models are constructed around each dataset in $\mathcal{D}$, each datapoint will be used for prediction multiple times. Given two close points $\boldsymbol{\zeta}_i$ and $\boldsymbol{\zeta}_j$, let's assemble all predictive errors where $\boldsymbol{\zeta}_i$ was used for model construction and $\boldsymbol{\zeta}_j$ for prediction, and vice versa, into a single array of error $E_{i,j}$. Then, the similarity value between $\boldsymbol{\zeta}_i$ and $\boldsymbol{\zeta}_j$ is taken to be the inverse of the minimum error in $E_{i,j}$, as defined by Eq. (3).

$$s(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j) = \frac{1}{\min E_{i,j}} \tag{3}$$

3.2. Graph Construction and Clustering

The similarity function defined in Eq. (3) can then be used to derive a square similarity matrix $S \in \mathbb{R}^{N \times N}$, such that $S_{i,j} = s(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_j)$. Following the methodology of spectral clustering, an adjacency matrix $A \in \mathbb{R}^{N \times N}$ is constructed according to Eq. (4).

$$A_{i,j} = \begin{cases} S_{i,j} \text{ if } \boldsymbol{\zeta}_i \in k_n \text{ neighbours of } \boldsymbol{\zeta}_j \text{ or } \boldsymbol{\zeta}_j \in k_n \text{ neighbours of } \boldsymbol{\zeta}_i \\ 0 \quad \text{otherwise} \end{cases} \tag{4}$$

Let's now assume that the number of regions is known to be k_c. Then spectral clustering can be performed in a straight forward manner on the graph described by A to yield k_c clusters containing training points within each region. This approach for hybrid system identification, which we term Local Predictive Clustering (LPC), is described in Algorithm 1.

Algorithm 1: Calculating the Similarity Between Points for Local Predictive Clustering

Input: dataset $\mathcal{D} = (\boldsymbol{\zeta}_i, y_i)_{i=1,\dots,N}$ where $\boldsymbol{\zeta}_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$, $k_l \in \mathbb{Z}^+$, $k_p \in \mathbb{Z}^+$, trainable model $h(\boldsymbol{\zeta}_i | \{\boldsymbol{\zeta}_j, y_j\}) \mapsto \mathbb{R}$

Do:

1. Initiate an empty list E of size $N \times N$, where each element contains an empty array
2. For each $\boldsymbol{\zeta}_i$ in $\mathcal{D}$:
   - 2.1. Find a set $\mathcal{D}_l$ of k_l nearest neighbours of $\boldsymbol{\zeta}_i$ using Eq. (3).
   - 2.2. Fit a local model h_i on the dataset $\mathcal{D}_l$
   - 2.3. Find a set $\mathcal{D}_p$ of k_p nearest neighbours of all points in $\mathcal{D}_l$
   - 2.4. Eliminate any point from $\mathcal{D}_p$ which belongs to $\mathcal{D}_l$ (i.e. $\mathcal{D}_p \cap \mathcal{D}_l = \emptyset$)
   - 2.5. Use the local model h_i to make prediction $\bar{y}_j$ for all points $\boldsymbol{\zeta}_j$ in $\mathcal{D}_p$
   - 2.6. For each point $\boldsymbol{\zeta}_j$ in $\mathcal{D}_p$, calculate Root Mean Square Error e_j between y_j and $\bar{y}_j$
   - 2.7. For each point $\boldsymbol{x}_k$ in $\mathcal{D}_l$, append e_j to $E[k, j]$ and $E[j, k]$ where k and j are original indices of $\boldsymbol{x}_k$ and $\boldsymbol{x}_j$ respectively.
3. Initiate an array $S \in \mathbb{R}^{N \times N}$ with zeros
4. For $i = 1, \dots, N$:
   - 4.1. For $j = 1, \dots, N$:
     - 4.1.1. $S[i, j] = \min(E[i, j])$

Return: S

3.3. Local Models

One of the advantages of LPC, is that any regression model can be used to approximate local models $h_i(\cdot)$. One preferred model is a Gaussian Process (GP) (Rasmussen 2006) due to its performance in low data settings and ability to encode additional prior assumptions about the underlying function through the kernel function. As such, LPC can be used for PWA system identification by simply using GPs with a linear kernel. On the other hand, Square Exponential (SE) kernel can be used for PWS system identification.

3.4. Parameter Tuning

The prediction neighborhood size k_p can be fixed to $N - k_l - 1$, i.e. making predictions at all points not used for local model construction, although smaller values can slightly reduce the computational cost without affecting performance. On the other hand, optimal

k_l and k_n are extremely sensitive to the dataset and even similar values can produce widely different clusters. These can be optimized by Bayesian Optimization within the range of 5-50% of the dataset size N.

Hierarchical clustering (Wang 2007) can be used to identify the number of clusters k_c, by iteratively splitting the dataset into two clusters while the resulting cross validation error decreases. An alternative method, enabled through the spectral clustering methodology, is to use the eigengap heuristic (von Luxburg 2007). One drawback of this method is that it is highly sensitive to the derived graph and hence k_l and k_n values, although this can be mitigated by summing the spectra of graphs produced by multiple candidate parameters.

3.5. From Local Predictive Clustering to Model Predictive Control

Once LPC divides the training dataset $\mathcal{D}$ into k_c clusters, the latent function $q(\cdot)$ can be approximated by any supervised classification model, such as an SVM. One caveat of hybrid system identification, is that the target value (i.e. $\boldsymbol{y}_i \equiv \boldsymbol{x}_{k+1}^{(i)} \in \mathbb{R}^{d_x}$) is a multidimensional vector, while LPC is aimed at scalar valued functions. A simple solution is then to treat each dimension of the target value separately, effectively resulting in d_x datasets and identified clusterings, where the optimal clustering is chosen through cross validation.

Once the latent function $q(\cdot)$ is learned, local functions $f_i(\cdot)$ for $i = 1, \dots, k_c$ can be readily approximated by a regression model of choice, for example a GP, hence learning a full piecewise multi-model of the hybrid system. This can then be leveraged in MPC by identifying the currently active local model at each iteration using the learned $q(\cdot)$ and only optimizing the controls over the finite horizon using the identified local model. At the next iteration, the active local model is re-identified.

4. Case Study

To illustrate LPC and advantages of proper hybrid system identification for MPC, let's consider an ideal continuous-stirred tank reactor (CSRT) described in (Kazantis 2000) where concentration and temperature of the reactant, $\boldsymbol{x}_k = [C_A, T] \in \mathbb{R}^2$, is controlled by a dilution rate and inlet temperature, $\boldsymbol{u}_k = [u_1, u_2] \in \mathbb{R}^2$. Hybrid behavior is introduced by modifying the dynamics described in [12], with the piecewise dynamics of the system described in Eq. (5). The system is then controlled from $\boldsymbol{x}_0 = [0.116, 368.5]$ to $\boldsymbol{x}_{target} = [0.666, 308.5]$.

$$\begin{aligned}
\frac{dC_A}{dt} &= \begin{cases} \left(\frac{F}{V} + u_1\right)\left(C_{A,in} - C_A\right) - 2k(T)C_A^2 & \text{if } C_A < 0.6 \\ \left(\frac{F}{V} + u_1\right)\left(C_{A,in} - C_A\right) - 2k(T)C_A & \text{otherwise} \end{cases} \\
\frac{dT}{dt} &= \begin{cases} \left(\frac{F}{V} + u_1\right)(T_{in} + u_2 - T) + 2\frac{(-\Delta H)_R}{\rho c_p}k(T)C_A^2 - \frac{UA}{V\rho c_p}\left(T - T_j\right) & \text{if } C_A < 0.6 \\ \left(\frac{F}{V} + u_1\right)(T_{in} + u_2 - T) + 2\frac{(-\Delta H)_R}{\rho c_p}k(T)C_A - \frac{UA}{V\rho c_p}\left(T - T_j\right) & \text{otherwise} \end{cases}
\end{aligned} \tag{5}$$

The discretized system model is uniformly sampled to yield 1120 training points used to learn the location of operating regions through LPC. Another training dataset of 160 points is sampled randomly and used to train a single, global GP. MPC is then run using the global GP as a system model, failing to control the plant to the target states. Then, the

training dataset is split into regions identified through LPC and local GPs are built for each region. A multi-model MPC is then run using local GPs by identifying an active local GP and using its mean prediction at each iteration, successfully controlling the plant.

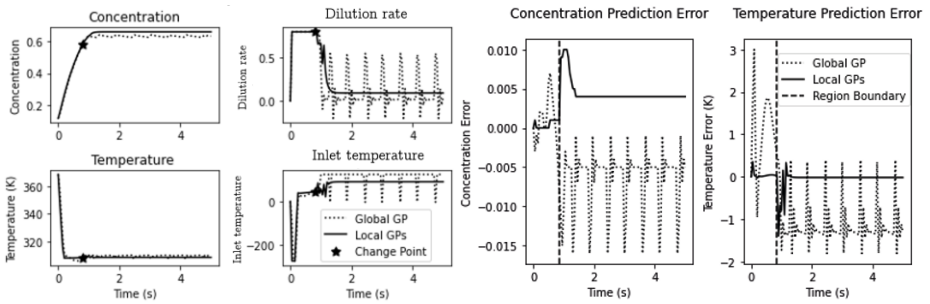


Figure 1: System behavior and model mismatch for MPC with a global and local models.

Figure 1 shows the system behavior as well as the model mismatch for both MPC with a single global GP and two local GPs. While local GPs exhibit large model mismatch close to the boundary ($C_A \approx 0.6$), it quickly decays. On the other hand, a single global GP results in significant model mismatch throughout, illustrating benefits of LPC.

5. Conclusions

Accurate hybrid system identification can significantly improve data driven control, and provide insight into the system itself. For MPC with GP system model, local models can also improve the speed of optimization. Although LPC is shown to be effective for general PWS system identification, it is limited to uniformly sampled datasets and is sensitive to noise and hyperparameters. Importantly, though, it illustrates the potential of spectral clustering in hybrid system identification through the proposed similarity function.

References

F. Lauer, G. Bloch, 2019, Hybrid System Identification: Theory and Algorithms for Learning Switching Methods

H. Ohlsson, L. Ljung, 2013, Identification of switched linear regression models using sum-of-norms regularization, Automatica 49, 1045-1050

B.J. Park, Y. Kim, J.M. Lee, 2021, Design of switching multilinear model predictive control using gap metric, Computers and Chemical Engineering 150

F. Lauer, G. Bloch, 2014, Piecewise smooth system identification in reproducing kernel Hilbert space, IEEE Conference on Decision and Control 53

G. Lee, Z. Marihno, A.M. Johnson, G.J. Gordon, S.S. Srinivasa, M.T. Mason, 2017, Unsupervised Learning for Nonlinear PieceWise Smooth Hybrid Systems

U. von Luxburg, 2007, A Tutorial on Spectral Clustering, Statistics and Computing 17

C.E. Rasmusen, K.I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006

X. Wang, V.L. Syrmos, 2007, Identification of Nonlinear Dynamical System using Hierarchical Clustering Analysis and Local Linear Models, European Control Conference

N. Kazantzis, C. Kravaris, 2000, Synthesis of state feedback regulators for nonlinear processes, Chemical Engineering Science 55

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50275-X

Distillation Column Temperature Prediction Based on Machine-Learning Model Using Wavelet Transform

Hyukwon Kwon[a], Yeongryeol Choi[a,b], Hyundo Park[a,b], Kwang Cheol Oh[c], Hyungtae Cho[a], Il Moon[b], Junghwan Kim[a*]

[a]*Green Materials and Processes R&D Group, Korea Institute of Industrial Technology, Ulsan 44413, Republic of Korea*
[b]*Department of Chemical and Biomolecular Engineering, Yonsei University, Seoul 03722, Republic of Korea*
[c]*Department of Biosystems Engineering, Agriculture and Life Sciences Research Institute, Kangwon National University, Gangwon-do 24341, Republic of Korea*
kjh31@kitech.re.kr

Abstract

This study presents a machine-learning-based prediction model for distillation process operation data using wavelet transform. The process operation data collected from a distillation column contain noise due to sensor errors. Developing a machine-learning model using noisy data reduces the accuracy of the model; therefore, the data should be denoised. Denoising was achieved using wavelet transform, and a long short-term memory (LSTM) machine-learning model was developed. Wavelet transforms generally decompose data into high- and low-frequency components using wavelet functions with various frequencies. The high-frequency components are the details comprising noisy data, and the low-frequency components correspond to the approximations of the original data. The approximations were used to develop the LSTM model. Depending on the type of wavelet function used for decomposition, the denoised values varied and affected the model accuracy. Case studies were conducted using various wavelet functions to develop models with optimum prediction performances. By applying the optimal wavelet transform to the LSTM model, the prediction performance improved by 10%.

Keywords: distillation column temperature, machine learning, wavelet transform, long short-term memory.

1. Introduction

The distillation process is a representative process for improving the purity of a product in the chemical industry. The main types of data in the distillation process are related to temperature, pressure, flow rate, and liquid level, which are measured in real time. In general, additional devices are required to measure the purity of the distillation process as real-time measurements are difficult. In addition, time delays are inevitable, which pose challenges to controlling real-time purity data (das Neves et al., 2018). Product purity is often controlled indirectly by estimating other related measurable data. Product purity is mainly controlled by temperature in the distillation process; therefore, accurate data collection and temperature prediction methods are required.
The distillation process data have time-series characteristics that accumulate over time. The artificial neural network (ANN)-based prediction model performs well in mapping

and utilizing the input and output values of a complex nonlinear relationship (Joo et al., 2021). Additionally, predictions and control are easy in real time because there are slight time delays of the order of seconds to output values using the developed model (Himmelblau, 2008). However, time-series data have performance limitations with predictions that use traditional ANN-based models. In time-series data, the past information affects the future because the traditional ANNs cannot reflect this. Recurrent neural network (RNN)-based long short-term memory (LSTM) is therefore designed to remember and convey past information in the future, making it suitable for analyzing and predicting time-series data.

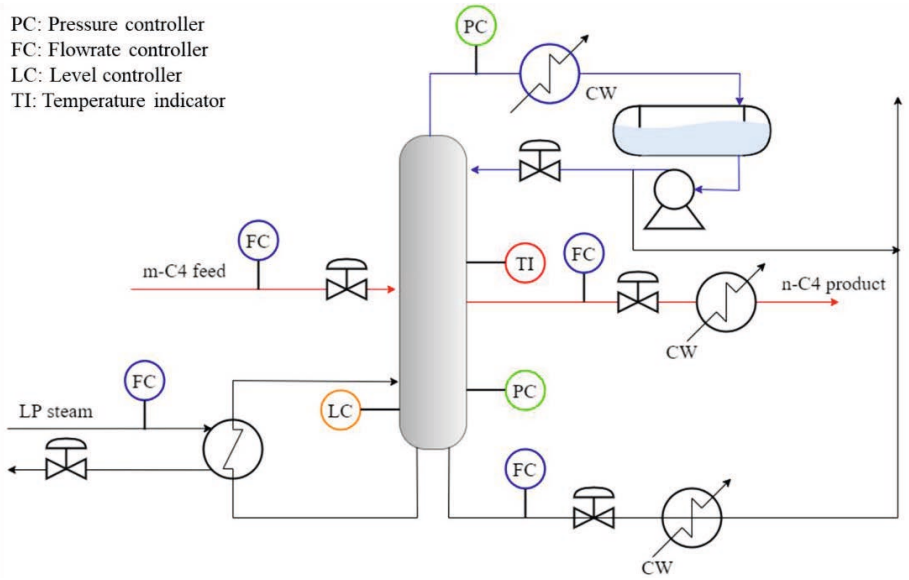
The performances of ANNs or linear and nonlinear regression models are affected by the data quality and require appropriate data preprocessing methods. Time-series data often include abnormal characteristics, such as tendency and periodicity; these characteristics are analyzed in several ways such as the autocorrelation and spectrum are mainly used. Spectral analysis is generally used to determine the periodicity of time-series data and is based on Fourier transform of the time-domain data to the frequency domain. Although the frequencies present in the signal can be analyzed after the Fourier transform, there is a disadvantage that the existence of each frequency in time is unknown because the time information is removed. Candidate (2019) reported that the wavelet transform (WT) is an approach to solving this problem. Khandelwal et al. (2015) conducted a study using ANN models and discrete wavelet transform (DWT) to predict time-series data. Kwon et al. (2021) reported that data from the distillation process could be predicted using the LSTM algorithm. The statistical technique called autoregressive integrated moving average (ARIMA) can be combined with ANNs to predict time-series data. The problem with ANNs as well as the linear and nonlinear regression models is that the predictive performances decrease for abnormal data that have not been appropriately preprocessed. The DWT can help analyze abnormal time-series data and decompose them into normal and noisy data.

In the present study, an LSTM-based prediction model with WT was developed to predict the temperature of the distillation process accurately. To improve the noise removal performance, a study was conducted on the basis function of the WT, and the most appropriate wavelet basis function was selected by measuring noise removal performance. The denoised data were then used as the input data to the LSTM model to develop the WT-LSTM model, and performance comparisons were made with the general LSTM model.

2. Methods

2.1. Description of the distillation process

The object of this study is a commercial distillation process that produces normal butane from mixed butane. The distillation column consists of 78 stages, and the raw material, i.e., mixed butane, is designed to flow into 35 stages. Mixed butane is composed of normal butane, isobutane, and pentane and produces 99% pure normal butane in 64 stages. Figure 1 shows a schematic of the distillation process and the sensor positions for data collected during the process.



2.2. Figure 1. Mixed butane splitter process diagram. Wavelet transforms and wavelet basis functions

WTs are divided into discrete and continuous types. The DWT consists of highpass and lowpass filters that separate the original signal into high- and low-frequency bands, respectively. In general, the approximation data from the lowpass filters determine the signal characteristics, and the detail data passed through the highpass filters are treated as noise. In decomposing the original signal into approximation and detail data, particular functions are used to determine the forms of the approximation data and are called as the wavelet basis functions.

The WT is defined as an extension of the wavelet basis function, which must satisfy the following two conditions:

$$\int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty \tag{1}$$

$$c_\psi = 2\pi \int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty \tag{2}$$

In Eq. (1) and (2), ψ is the wavelet basis function, and Ψ is the Fourier transform of ψ. Eq. (1) indicates that the function ψ has a finite value, and Eq. (2) indicates that $\Psi(0) = 0$ when Ψ is smoothed as an acceptance condition. Because the wavelet basis function is a simple function that needs to satisfy only the above two conditions, the types of wavelet basis functions used are so diverse that it is necessary to select suitable functions based on the data characteristics. The data collected during this process have different tendencies depending on the variable. Therefore, it is necessary to select the most suitable wavelet basis function for each variable. In this study, five wavelet basis functions are used: biorthogonal (bior), coiflet (coif), Daubechies (db), reverse biorthogonal (rbior), and symlet (sym). These five functions are the most commonly used in discrete wavelet transformations and have been useful in other signal processing studies. The degree of noise removal was evaluated for each wavelet function to select the most appropriate function for each data variable. In the process of selecting the basis function, all decomposition levels were set to five.

2.3. Long short-term memory

Traditional neural networks have an input layer, an output layer, and a hidden connected layer; however, the nodes within each layer are not connected, so they do not affect each other. This makes it unsuitable for predicting time-series data in which the past information affects the future. RNNs are ANNs for predicting timeseries data and are connected such that they convey past information, unlike other traditional neural networks. The LSTM is proposed herein to solve the vanishing gradient problem, which is one of the disadvantages of the RNN.

3. WT-LSTM model development

In this study, a combined LSTM and WT model is developed. Data used for model development were collected between 07.18.2019 and 07.23.2019 using sensors installed in the actual commercial process. Among the collected data, the feed, steam, reflux flowrate, bottom pressure, and temperature values are noisy owing to sensor errors; hence, preprocessing is performed with WT to remove noise. In the process of denoising with WT, various basis functions are applied to select the optimal noise removal function based on performance. The preprocessed data are applied at the input layer of the LSTM model for learning, and the temperature is predicted at the output layer. The number of hidden nodes in the LSTM was set to 20 with 128 batch sizes. The other parameters are summarized in Table 1.

Table 1. Parameters of LSTM model.

Parameter	Values
Optimizer	Adam
Learning rate	0.001
Activation function	ReLU
Training/test dataset ratio	70/30
Loss function	MSE
Wavelet basis functions	bior, coif, db, rbior, sym
Wavelet decomposition levels	5

The root mean-squared error (RMSE) is used to evaluate model performance and is calculated as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2} \quad (3)$$

Here, x_i and y_i are the ith values of the actual and predicted datasets respectively, and N is the total number of datasets. The RMSE is a measure of the error between the raw and predicted data, indicating that the smaller the error value, the more accurate is the performance of the prediction model.

4. Results and discussion

In this study, an LSTM model combined with a WT was developed to predict the temperature of the distillation process. Various basis functions were applied and evaluated to improve the noise removal performance of the WT. Performance comparisons with basic LSTM models were performed to verify the performance of the proposed model.

4.1. Wavelet basis functions

The selection of the wavelet basis function is important for removing noisy data using the DWT. In general, when denoising a signal using WT, greater similarity of the shape of the wavelet function to the signal indicates better noise removal. However, it is difficult to create useful wavelet functions, hence, an appropriate wavelet function is selected and used from among the existing wavelets. Daubechies, Symlets, Coiflets, biorthogonal, and other wavelets are included for decomposing signals and denoising with DWT. In this study, five wavelet basis functions were applied to each distillation process dataset for denoising and performance evaluation for optimal function selection. Table 2 shows the denoising results and RMSEs for bior, coif, db, rbior, and sym basis functions.

Table 2. Results of denoising evaluations for each variable with RMSE.

	Feed flowrate	Steam flowrate	Reflux flowrate	Bottom pressure	Temperature
bior	652.941	264.516	1828.90	0.089	0.390
coif	652.412	264.009	1823.67	0.089	0.390
db	652.457	264.859	1825.079	0.089	0.390
rbior	685.477	274.663	1914.24	0.092	0.403
sym	652.449	264.368	1824.54	0.089	0.390

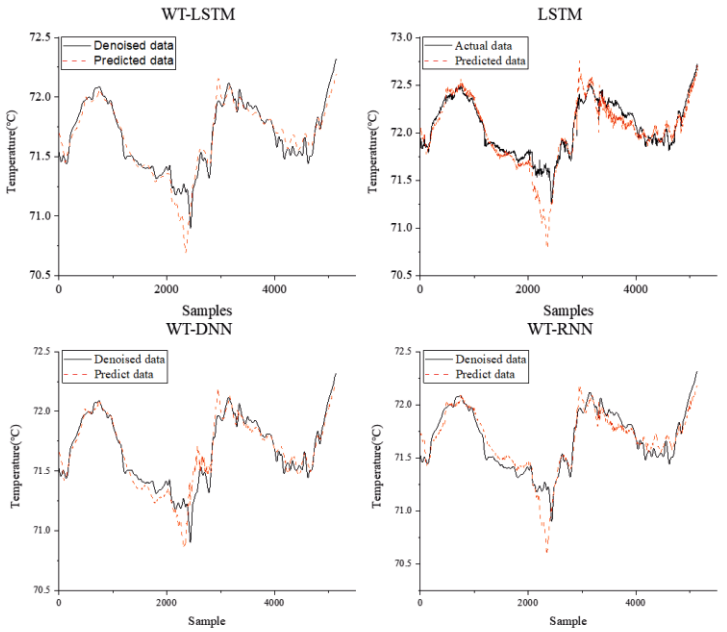


Figure 2. The comparisons between denoised and predicted temperature of WT-LSTM, LSTM, WT-DNN and WT-RNN model.

In this study, we confirmed that the coif function was most suitable for denoising data. However, the bior and db functions showed no significant differences compared to coif. Thus, we found that other functions were also suitable for denoising.

4.2. Performance evaluation

After removing data noise using the selected coif function, learning was conducted with the LSTM model to develop the WT-LSTM model. The performance of the proposed

model was compared with the basic LSTM model, WT- DNN, and WT-RNN. The learning conditions of WT-LSTM, LSTM model, and WT-RNN were set the same, and the predictive performance was evaluated with RMSE. Figure 3 shows the prediction of test set data using WT-LSTM, LSTM, WT-DNN and WT-RNN model. The RMSE of each model was 0.0966, 0.144, 0.0976 and 0.117. we confirmed that a more accurate prediction could be performed by reducing RMSE by 33# than the basic-LSTM model by removing noise with WT.

5. Conclusion

A machine-learning model combined with WT was developed herein to predict distillation process temperatures. The process data were denoised effectively using WT, and we confirmed that the coif function was most appropriate for optimal denoising performance. It was shown that the WT-LSTM model using the coif function could achieve more accurate predictions than other basis functions, with a 33% reduction in RMSE than the basic LSTM model. When using WT, the LSTM model achieved more accurate predictions. The temperature is thus calculated to be lower by approximately 0.4 ℃. Hence, the WT approach needs to be improved, and future studies will focus on models that can predict the process more accurately through supplemental corrections.

Acknowledgement

This study has been conducted with the support of the Korea Institute of Industrial Technology as “Development of AI Platform Technology for Smart Chemical Process (KITECH JH-21-0005)” and “Development of digital-based energy optimization platform for manufacturing innovation (KITECH IZ-21-0063)” and “Development of Global Optimization System for Energy Process (KITECH IZ-21-0052, IR-21-0029, EM-21-0022)”.

References

T.G. das Neves, W.B. Ramos, G.W. de Farias Neto, R.P. Brito, 2018, Intelligent control system for extractive distillation columns. Korean Journal of Chemical Engineering, 35, 826–834. https://doi.org/10.1007/s11814-017-0346-0

D.M. Himmelblau, 2008, Accounts of experiences in the application of artificial neural networks in chemical engineering. Industrial and Engineering Chemistry Research, 47, 5782–5796. https://doi.org/10.1021/ie800076s

C. Joo, H. Park, J. Lim, H. Cho, J. Kim, 2021, Development of physical property prediction models for polypropylene composites with optimizing random forest hyperparameters. International Journal of Intelligent Systems, https://doi.org/10.1002/int.22700

I. Khandelwal, R. Adhikari, G. Verma, 2015, Time series forecasting using hybrid ARIMA and ANN models based on DWT decomposition, Procedia Computer Science, 48, 173–179. https://doi.org/10.1016/j.procs.2015.04.167

H. Kwon, K.C. Oh, Y. Choi, Y.G. Chung, J. Kim, 2021, Development and application of machine learning-based prediction model for distillation column, International Journal of Intelligent Systems, 36, 1970–1997. https://doi.org/10.1002/int.22368

H. Lee, J. Lee, C. Yoo, 2019, Selecting a mother wavelet for univariate wavelet analysis of time series data, Journal of Korea Water Resources Association, 52, 575–587. https://doi.org/10.3741/JKWRA.2019.52.8.575

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50276-1

Moisture Estimation in Woodchips Using IIoT Wi-Fi and Machine Learning Techniques

Kerul Suthar[a], Q. Peter He[a*]

[a]*Department of Chemical Engineering, Auburn University, Auburn, AL 36849, USA*

qhe@auburn.edu

Abstract

For the pulping process in a pulp & paper plant that uses woodchips as raw material, the moisture content (MC) of the woodchips is a major process disturbance that affects product quality and consumption of energy, water, and chemicals. Existing woodchip MC sensing technologies have not been widely adopted by the industry due to unreliable performance and/or high maintenance requirements that can hardly be met in a manufacturing environment. To address these limitations, we propose a non-destructive, economic, and robust woodchip MC sensing approach utilizing channel state information (CSI) from industrial Internet-of-Things (IIoT) based Wi-Fi. While these IIoT devices are small, low-cost, and rugged to stand for harsh environment, they do have their limitations such as the raw CSI data are often very noisy and sensitive to woodchip packing. Thus, direct application of machine learning (ML) algorithms leads to poor performance. To address this, statistics pattern analysis (SPA) is utilized to extract physically and statistically meaningful features from the raw CSI data, which are sensitive to woodchip MC but not to packing. The SPA features are then used for developing multiclass classification models as well as regression models using various linear and nonlinear ML techniques to provide potential solutions to woodchip MC estimation for the pulp and paper industry.

Keywords: systems engineering, machine learning, feature engineering, channel state information, IIoT sensors.

1. Introduction

The US pulp and paper industry ranks the third in energy consumption among US industries. The pulping process, which converts woodchips into pulp by displacing lignin from cellulose fibers, is one of the most energy intensive processes and has been identified as a major opportunity to improve energy productivity and efficiency of the industry (Brueske et al., 2015). Currently, vast majority of the US pulp is produced by chemical pulping processes and most of them utilize continuous Kamyr digesters. For Kamyr digesters, the incoming woodchip moisture content (MC) is a major disturbance that affects the cooking performance.

Currently, the woodchip MC is not measured in real-time due to the lack of affordable, reliable, and easy-to-maintain sensors. As a result, the performance of existing control solutions is often unsatisfactory and process engineers often overcook the woodchips to ensure pulp quality, which results in significant loss of pulp yield, overuse of heat/energy and chemicals. Chemical overuse also adds burdens to the downstream processes, such as washing and evaporation, and results in increased energy and chemical usages for downstream processes as well. To address this need, this work proposes a non-

destructive, economic, and robust approach based on 5 GHz IIoT short-range Wi-Fi and use channel state information (CSI) to estimate MC in woodchips. Both classification and regression techniques are studied for MC estimation. For classification, we investigate linear discriminant analysis (LDA), support vector machine (SVM), artificial neural network (ANN), bagging with LDA, and ensemble boosting XGBoost. For regression, we study ANN, k-nearest neighbor regression (KNNR), Gaussian process regression (GPR), and support vector regression (SVR) with radial basis function (RBF) kernel.

The remainder of this work is organized as follows: Section 2 describes the experimental setup and software tools used in this study, as well as the features proposed and the modeling techniques utilized in this work. Section 3 presents results and discussions of this work, and Section 4 draws conclusions.

2. Data collection and feature engineering

2.1. Channel state information for moisture estimation

Using Wi-Fi cards such as IWL5300, it is convenient to collect CSI measurements that record the channel variation during propagation of wireless signals. After being transmitted from a source, the wireless signal is expected to experience impairments caused by obstacles before the signal reaches the receiver. CSI can reflect indoor channel characteristics such as multipath effect, shadowing, fading, and delay. In this work, we collect CSI using CSItool, which is built on IWL5300 NIC using a custom modified firmware and open-source Linux wireless drivers. The channel response of the i^{th} subcarrier can be represented as:

$$CSI_i = |CSI_i| \exp\{\angle CSI_i\} \tag{1}$$

where $|CSI_i|$ is the amplitude and $\angle CSI_i$ is the phase response of the i^{th} subcarrier.

2.2. Data description

In this work, data are collected for 20 different MC classes or levels ranging from 53.39% to 11.81% on the wet basis (see Eqn (2)). A single antenna is used on the transmitter side which is configured in injection mode to send CSI and 3 antennas are used on the receiving side to take advantage of diversity. Woodchips are places in an airtight container between the transmitter and receiver to collect data. 10,000 packets are sent from the transmitter to the receivers for each sample collection. Total mass (m_T) is measured during each experiment and oven drying method was performed after all experiments were conducted to determine the oven dry weight (m_D). m_T and m_D are then used to determine the mass of water (m_W) and MC as the following.

$$MC = \frac{m_W}{m_T} \times 100\% = \frac{m_W}{m_W + m_D} \times 100\% \tag{2}$$

The 20 different MC levels are plotted in Figure 1(a), which shows that MC levels are narrowly separated at the high MC region and even more so at the low MC region. The minimum difference between MC levels is 0.05%, which is more than sufficient for pulping process optimization and control.

2.3. Methodology and feature engineering

To address the shortcoming of raw CSI features that lead to poor classification and prediction performance, in this work, statistics pattern analysis (SPA) is utilized to

generate more robust and predictive features. In SPA, the statistics of the process variables, instead of process variables themselves, are used for modeling. This is based on the hypothesis that these statistics are sufficient and even better in capturing process characteristics than original process variables. This hypothesis has been supported in various applications, including fault detection (He et al., 2019; He & Wang, 2011, 2018; Wang & He, 2010), fault diagnosis (He & Wang, 2018), and virtual metrology or soft sensor (Shah et al., 2019, 2020; Suthar et al., 2019). SPA is selected in this work to extract robust and predictive features from raw CSI data. It is worth noting that SPA does not require preprocessing of the CSI data (e.g., outlier detection and handling, noise removal/reduction) that has been required in previous studies (Hu et al., 2019; Yang et al., 2018). A schematic for SPA based feature engineering is shown in Figure 1 (b). After a deeper exploration of candidate features and statistics, mean difference of consecutive subcarrier in CSI amplitude are chosen which leads to 87 features considering all 3 antennas on the receiving side.

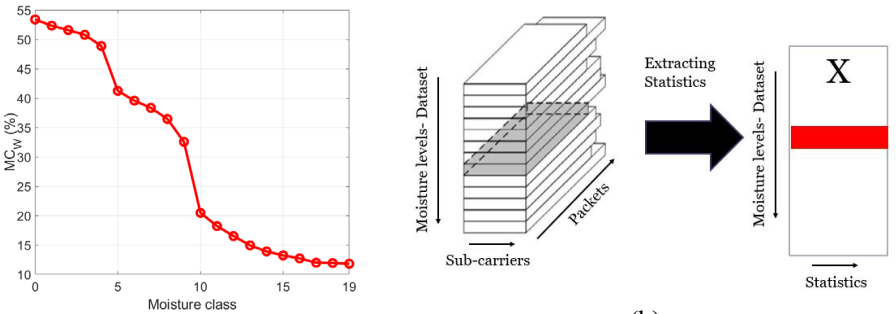


Figure 1 (a) 20 different moisture levels tested in this work; (b) SPA based feature engineering for MC estimation

3. Results and discussion

In this work, we conduct investigations from three perspectives: (1) comparing raw CSI data vs engineered features; (2) comparing the performance of different classification approaches; and (3) comparing the performance of different regression approaches. For each model, 9 samples are randomly selected as training samples from 10 shuffled samples at the same MC level for each of the 20 MC levels, which results in 180 training samples. The remaining shuffled sample for each of the MC levels is used for testing. In this work we use Monte Carlo validation and testing (MCVT) procedure 100 times for performance comparison. To assess various classification approaches, the mean and standard deviation of classification accuracy of the 100 MCVT simulations are reported. For regression approaches, the mean and standard deviation of root mean square error (RMSE) for the 100 MCVT simulations are reported.

First, raw CSI data are used for MC level classification. The results are similar across different classification techniques. Due to limited space, only results from LDA are discussed here. Figure 2 (a) shows the overall classification accuracy of all classes when the raw CSI data were used. The comparison indicates that LDA classifier using both amplitude and phase difference performs the best with 86.15% classification accuracy, followed by LDA classifier using phase difference with 83.85% classification accuracy, while the LDA classifier using amplitude alone results in the lowest classification

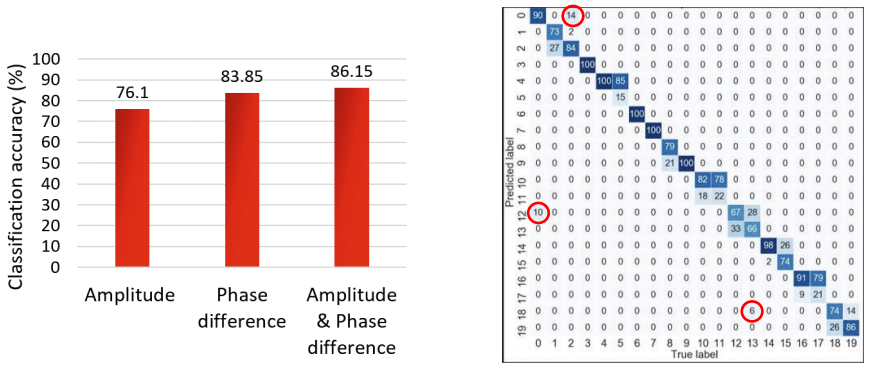


Figure 2 (a) Overall classification accuracy using different raw CSI data with LDA classifier based on 100 Monte Carlo runs. (b) Classification confusion matrix of 100 MCVT when both amplitude and phase difference are used. The far-off misclassifications (*i.e.*, the predicted class differs from the true class by more than one MC level) are highlighted by red circles.

accuracy of 76.10%. Figure 2 (b) plots the confusion matrix for the LDA classifier using both CSI amplitude and phase difference, which allows us to dig deeper into the classification results. As can be seen from Figure 2 (b), classification accuracy of individual classes ranges from 15% to 100%. It can also be seen that classification accuracy alone is not a good performance indicator. For example, the far-off misclassifications (i.e., the predicted class of a sample is off its true class by more than one level) will have worse consequences than the nearest-neighbor misclassifications (i.e., the predicted class is off true class by one level, either above or below) if they were used to control the white liquor usage or digester temperature. It can be seen from Figure 2 (b) that the classification results using raw CSI data are poor as there are samples misclassified far off their true classes. There are totally 478 misclassified samples, of which 30 are far-off misclassifications (highlighted by red circles in Figure 2 (b)). Also, the overall classification accuracy is not satisfactory.

Next the 87 rationally engineered features (i.e., the mean difference of consecutive subcarrier in CSI amplitude) are used for MC level classification and the results are summarized in Table 1. The classification accuracies shown in Table 1 indicate that all methods perform well with higher than 95% classification accuracy. The significantly improved performance compared to that of the raw CSI data demonstrates that the engineered features are more informative and characterize the MC in woodchips far better than the raw CSI data. Among all classification methods studied in this work, the bagging LDA performs the best with 98.75% average classification accuracy. The standard deviation of its classification accuracy is the lowest of 2.29%, indicating the

Table 1 Classification accuracy using engineered features

Method	Classification Accuracy	
	Mean	Std. dev.
SVM	95.50	3.79
ANN	95.85	4.15
XGBoost	96.40	3.70
LDA	97.55	2.89
Bagging (LDA)	98.75	2.29

bagging LDA is also the most robust or consistent classifier among all methods studied in this work.

Finally, we study different regression methods for MC estimation. When raw CSI data are used, all regression methods perform poorly, similar to the classification results when the raw CSI data are used. Due to limited space, they are not shown here. When the same 87 engineered features are used for regression-based MC estimation, a well-tuned ANN with two hidden layers outperforms other regression-based approaches as shown in Table 2. KNNR performs comparable to ANN while GPR and SVR with RBF kernel have relatively higher average RMSE's for 100 MCVT simulations.

Table 2 Regression for MC estimation using engineered features

Method	RMSE	
	Mean	Std. dev.
ANN	0.51	0.3921
KNNR	0.6573	0.5055
GPR	1.9223	0.5714
SVR(RBF)	2.0179	0.523

Figure 3 shows the measured vs predicted MC values for ANN and SVR(RBF). It can be seen from Figure 3(a) that the ANN predicted MC values agree very well with the actual or measured MC values. In comparison, while SVR captures the MC trend, its predictions have much higher standard deviation compared to ANN. It is worth noting for all the above-mentioned results, the models and their hyperparameters were tuned using random search followed by Bayesian optimization (Bergstra & Bengio, 2012).

4. Conclusions

In this work, we investigate the potential of an IIoT short-range Wi-Fi based woodchip MC sensing technology to overcome some limitations of the existing technologies. The proposed technology takes the advantages of IIoT devices (e.g., toughness, connectivity,

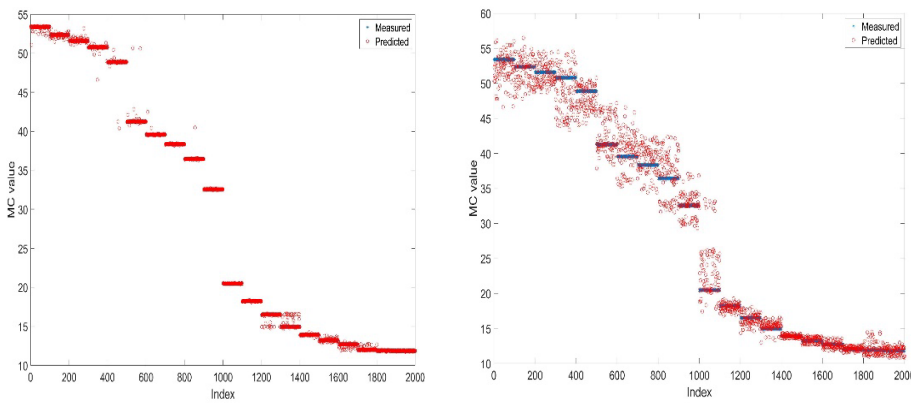


Figure 3 Measured vs predicted MC by (a) ANN and (b) SVR(RBF)

low-cost, small-size, etc.), while overcoming their shortcomings (e.g., the machine learning challenges of messy big data) through SPA-based feature engineering. We investigate the use various classification and regression approaches for the estimation of 20 different moisture levels. We demonstrate that with SPA-based features, all classification approaches studied in this work can successfully classify 20 different MC levels, some of which are separated by small margins. We also investigate the use of different regression approaches for continuous MC estimation. While SVR and GPR capture the trend of measured MC values but with relatively high RMSE's, methods including ANN and KNNR predict the moisture levels accurately. The relationship between the CSI and woodchip MC is very complex, which requires further work to get a better understanding of this relationship for further improvement of this work.

References

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*.

Brueske, S., Kramer, C., & Fisher, A. (2015). *Bandwidth Study on Energy Use and Potential Energy Saving Opportunities in US Pulp and Paper Manufacturing*. Energetics.

He, Q. P. P., & Wang, J. (2018). Statistical process monitoring as a big data analytics tool for smart manufacturing. *Journal of Process Control*, *67*, 35–43.

He, Q. P., & Wang, J. (2011). Statistics pattern analysis: A new process monitoring framework and its application to semiconductor batch processes. *AIChE Journal*, *57*(1), 107–121.

He, Q. P., & Wang, J. (2018). Statistical process monitoring as a big data analytics tool for smart manufacturing. *Journal of Process Control*, *67*, 35–43.

He, Q. P., Wang, J., & Shah, D. (2019). Feature Space Monitoring for Smart Manufacturing via Statistics Pattern Analysis. *Computers & Chemical Engineering*, *126*, 321–331.

Hu, P., Yang, W., Wang, X., & Mao, S. (2019). MiFi: Device-free wheat mildew detection using off-the-shelf wifi devices. *2019 IEEE Global Communications Conference, GLOBECOM 2019 - Proceedings*.

Shah, D., Wang, J., & He, Q. P. (2019). A feature-based soft sensor for spectroscopic data analysis. *Journal of Process Control*, *78*, 98–107.

Shah, D., Wang, J., & He, Q. P. (2020). Feature Engineering in Big Data Analytics for IoT-Enabled Smart Manufacturing–Comparison between Deep Learning and Statistical Learning. *Computers & Chemical Engineering*, 106970.

Suthar, K., Shah, D., Wang, J., & He, Q. P. (2019). Next-generation virtual metrology for semiconductor manufacturing: A feature-based framework. *Computers and Chemical Engineering*, *127*, 140–149.

Wang, J., & He, Q. P. (2010). Multivariate Statistical Process Monitoring Based on Statistics Pattern Analysis. *Industrial & Engineering Chemistry Research*.

Yang, W., Wang, X., Cao, S., Wang, H., & Mao, S. (2018). Multi-class wheat moisture detection with 5GHz Wi-Fi: A deep LSTM approach. *Proceedings - International Conference on Computer Communications and Networks, ICCCN*.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50277-3

Transfer Learning for Quality Prediction in a Chemical Toner Manufacturing Process

Shohta KOBAYASHI[a*], Masashi MIYAKAWA[a], Susumu TAKEMASA[a], Naoki TAKAHASHI[a], Yoshio WATANABE[a], Toshiaki SATOH[b], Manabu KANO[c]

[a] *Ricoh Co., Ltd., Shizuoka 410-0004, JAPAN*
[b] *MEIWA e-TEC Co., LTD., Aichi 471-0047, JAPAN*
[c] *Kyoto University, Kyoto 606-8501, JAPAN*
shohta.kobayashi@jp.rioch.com

Abstract

In chemical toner manufacturing plants, equipment and raw materials are frequently changed to improve the toner quality and productivity. These changes require reconstruction of the prediction model, which plays a key role in the automatic quality control system, and cause downtime. To reduce the downtime, we developed an efficient modelling method based on transfer learning, which can build an accurate model from small-size data obtained just after the changes. By extending Frustratingly Easy Domain Adaptation, a new heterogeneous domain adaptation technique was proposed. In addition, gaussian process regression (GPR) was adopted with bagging to improve the robustness and accuracy of the model. The proposed method showed superior performance to partial least squares regression, random forest, and GPR. Finally, the proposed prediction method was applied to a toner mass-production plant; the prediction accuracy target was satisfied for all toner qualities. As a result, a 75% reduction in plant control person-hours of the toner quality manager was achieved.

Keywords: Quality prediction, Transfer learning, Frustratingly Easy Domain Adaptation, Gaussian Process Regression, Chemical Toner Production.

1. Introduction

In recent years, automatic quality control has been used for stabilizing chemical toner quality and determining efficient operating conditions in toner plants (Khorami et al., 2017; Takahashi et al., 2020). In these plants, equipment and raw materials are often changed. As shown in Fig. 1, such a change alters the dimensions and distributions of input variables, makes it difficult to use existing prediction models, and makes it necessary to reconstruct the models. During the re-accumulation of training data, automatic quality control is forced to stop functioning, and manual quality control is required. This manual control requires many person-hours, therefore, it has been desired to construct an accurate prediction model using as short-term data as possible.

A promising approach to solve this problem is to use transfer learning. We expanded Frustratingly Easy Domain Adaptation (FEDA), which is a simple homogeneous domain adaptation method, to cope with a heterogeneous domain adaptation (HDA) problem without complex parameter tuning. The proposed method is referred to as Frustratingly Easy Heterogeneous Domain Adaptation (FEHDA). Moreover, we utilized

a combination of Gaussian Process Regression (GPR) and bagging, a type of ensemble learning, for predicting the toner quality.

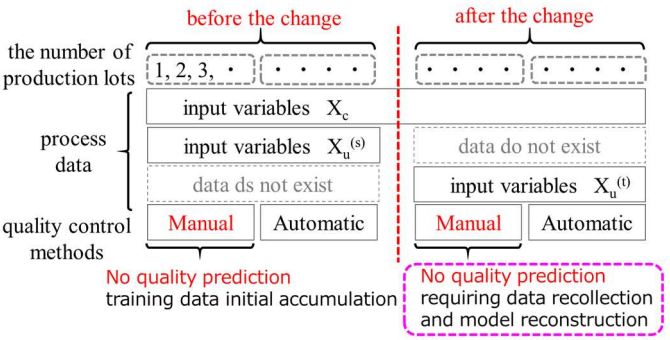


Fig. 1. The influence of changes in equipment and raw materials.

2. Chemical toner manufacturing process

The chemical toner manufacturing process treats one lot per day, and it takes seven days from raw material to final product, as shown in Fig. 2. The IoT-based manufacturing process data collection system handles several thousand variables (items), including raw material properties, equipment operation conditions, and toner quality, and stores data of several hundred lots or more.

Before the introduction of automatic quality control, toner quality was controlled manually by the toner quality manager, who determined the optimum operating condition for lot *N* based on the quality measurements of the lots whose manufacturing was finished (lot *N*-2 and older). The manual quality control consumes many person-hours and increases the risk of out-of-specification due to variations in toner quality.

The automatic quality control system currently in operation consists of a quality prediction module that predicts future toner quality and an operating condition optimization module that determines the operation amount (Takahashi et al., 2020).

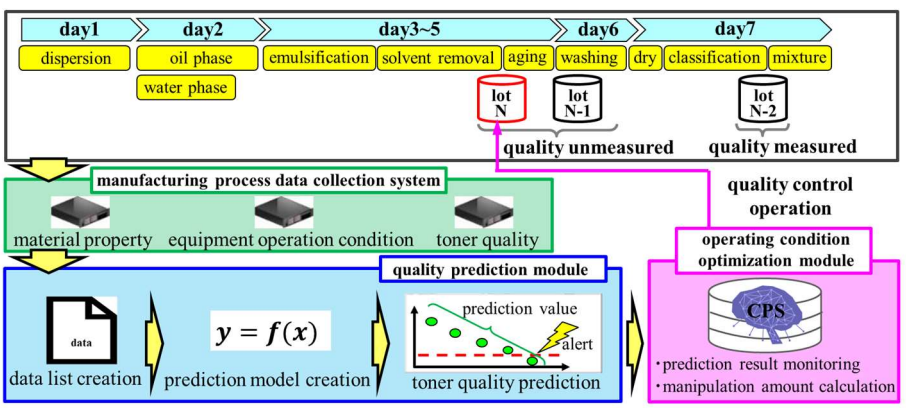


Fig. 2. Chemical toner manufacturing process and automatic quality control system (Takahashi et al., 2020).

This automatic quality control is feed-forward inferential control, which simulates the manual operation, as shown in Fig. 2. However, as mentioned in the previous section, the change of equipment and raw materials requires the data re-accumulation of at least 40 lots (days) so that the model reconstructed satisfies the accuracy target.

3. Prediction using transfer learning

3.1. Frustratingly Easy Heterogeneous Domain Adaptation

FEDA is a method of transfer learning that is easy to implement with simple feature space expansion (Daumé III, 2007). Assuming that the input variables $\boldsymbol{x}^{(s)}$ in the source domain (hereinafter referred to as "SD") and $\boldsymbol{x}^{(t)}$ in the target domain (hereinafter referred to as "TD") are K-dimensional, the input variables in both domains are expanded into $3K$-dimensional features as follows:

$$D_s = \left(\boldsymbol{x}^{(s)},\ \boldsymbol{x}^{(s)},\ 0\quad\right) \tag{1}$$
$$D_t = \left(\boldsymbol{x}^{(t)},\ 0\quad,\ \boldsymbol{x}^{(t)}\right) \tag{2}$$

The expanded feature space consists of a space with features common to both domains, a space with features unique to SD, and a space with features unique to TD. Also, $0 = (0, 0, 0, \ldots., 0) \in \mathcal{R}^K$ in Eqs. (1) and (2) is the zero vector.

In the manufacturing process, due to changes in equipment and raw materials, the configuration of the manufacturing equipment differs in both domains, which makes the location and number of installed sensors also different. Hence, when heterogeneous domain adaptation is required, FEDA cannot be used as it is. To make FEDA applicable to heterogeneous domain adaptation (HDA), heterogeneous feature augmentation (HFA) (Duan et al., 2012) was proposed. This method needs much computational time because to solve an optimization problem for finding the optimal latent space.

We propose frustratingly easy heterogeneous domain adaption (FEHDA), which is a direct and simple extension of FEDA and applicable to HDA. The proposed method does not require solving the optimization problem. We divide input variables $\boldsymbol{x}^{(s)} \in \mathcal{R}^P$ in SD into $\boldsymbol{x}_c{}^{(s)} \in \mathcal{R}^K$ that is common to SD and TD and $\boldsymbol{x}_u{}^{(s)} \in \mathcal{R}^{P-K}$ that is unique to SD. Similarly, input variables $\boldsymbol{x}^{(t)} \in \mathcal{R}^Q$ in TD is divided into the common input variables $\boldsymbol{x}_c{}^{(t)} \in \mathcal{R}^K$ and the unique input variables $\boldsymbol{x}_u{}^{(t)} \in \mathcal{R}^{Q-K}$. As shown in Fig. 3, $\boldsymbol{x}_c{}^{(s)}$ and $\boldsymbol{x}_c{}^{(t)}$ are expanded as in Eqs. (1) and (2), respectively, while $\boldsymbol{x}_u{}^{(s)}$ and $\boldsymbol{x}_u{}^{(t)}$ are placed in the space with unique features in each domain as follows:

$$D_s = \left(\boldsymbol{x}_c{}^{(s)},\ \boldsymbol{x}_c{}^{(s)},\ \boldsymbol{x}_u{}^{(s)},\ \ 0\quad,\ \ 0\ \right) \tag{3}$$
$$D_t = \left(\boldsymbol{x}_c{}^{(t)},\ \ 0\ \ ,\ \ 0\quad,\ \boldsymbol{x}_c{}^{(t)},\ \boldsymbol{x}_u{}^{(t)}\right) \tag{4}$$

3.2. Prediction Model

To build a prediction model, we propose a method that combines Gaussian process regression (GPR) and bagging. The input variables are the expanded ones in Eqs. (3) and (4). GPR can predict not only the expected values but also the standard deviations of output variables and provide the reliability of the prediction. Bagging is a form of ensemble learning that uses bootstrap sampling to construct many independent weak learners and then integrates the results of the weak learners into a prediction.

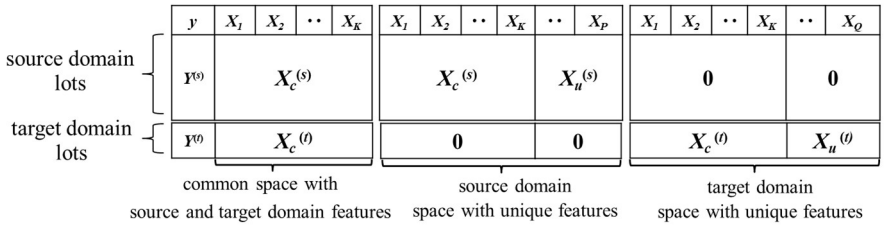


Fig. 3. Feature space expansion in Frustratingly Easy Heterogeneous Domain

Kamishima et al. (2009) proposed TrBagg, which uses bagging for transfer learning. TrBagg builds weak learners using data sampled from SD and TD. The weak learners are adopted based on the classification errors for TD. The method may cause over-fitting or require the separation of TD for validation.

The chemical toner manufacturing process produces only one lot per day. To reduce the downtime of the automatic quality control system, the number of samples after each change, which are used for reconstructing the prediction model, needs to be limited. That means the number of TD lots must be small, i.e., about 10 lots. Since TrBagg does not work well in such a situation, we did not adopt it. In the proposed method, bagging is modified by selecting only weak learners with small standard deviations of the output variables when integrating the results of the weak learners. The weak learners with small standard deviations are expected to give a more reliable prediction because it is considered to use data with high similarity to the target lot preferentially. We use sequential updating of the prediction model for each lot.

4. Comparison of prediction methods

The proposed modeling method, i.e., GPR and bagging, was compared with the typical regression methods, partial least squares regression (PLSR), random forest (RF), and GPR in two cases: 1) change of coloring materials, representing material improvement, and 2) change of production scale, representing equipment improvement. FEHDA was used in both cases, and the two most important qualities were investigated. The dimensions of the input variables are shown in Table 1.

In case 1, black and magenta toners, which were made from almost the same materials except for the coloring one, were targeted, regarding black toner as SD and magenta toner as TD. In case 2, the same color toner manufactured by equipment with different scales was targeted, regarding the large scale plant as SD and the small scale one as TD.

The prediction accuracy was evaluated using Root Mean Squared Error (RMSE). In

Table 1

The dimensions of the input variables and the number of lots in two cases: 1) change of coloring materials and 2) change of production scale.

	input variables	case (1)	case (2)	Application to a mass production plant
the number of variables	$X^{(s)}$	2821	2826	2914
	$X^{(t)}$	2831	2266	2721
	$X_c^{(s)}$, $X_c^{(t)}$	2789	1803	2598
	$X_u^{(s)}$	32	1023	316
	$X_u^{(t)}$	42	463	123
the number of lots	$X^{(s)}$, $X_c^{(s)}$, $X_u^{(s)}$	450	450	333
	$X^{(t)}$, $X_c^{(t)}$, $X_u^{(t)}$	10-109	10-109	10-39

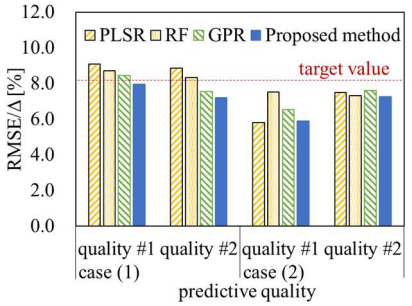


Fig. 4. Comparison of RMSE's of prediction methods.

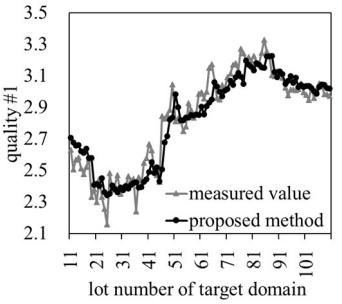


Fig. 5. Quality #1 prediction results of the proposed method for case 1.

defining the target for the prediction accuracy, the following conditions were set; first, the center of the predicted distribution of the qualities is within 50% of the process specification width Δ, and second, the probability of out-of-specification is less than 0.3% when the quality prediction value is at the upper or lower limit of the process specification width Δ. The variability was assumed to be normally distributed (Takahashi et al., 2020). Based on these conditions, the target value for prediction accuracy became $0.5\Delta \geq 6$ RMSE, i.e., $\text{RMSE}/\Delta \leq 8.3[\%]$.

Fig. 4 shows the evaluation results for the 11th to 110th lots in TD. The proposed method outperformed the other methods in both cases and also satisfied the prediction accuracy target. In particular, a more significant improvement was obtained in case 1. Fig. 5 shows the predicted and measured values for each lot of quality #1. It was confirmed that the predicted values followed the trend of the actual measured values, and there were no large errors in all lots. On the other hand, the improvement achieved by the proposed method in case 2 was smaller than that in case 1. This can be attributed to the large proportion of intrinsic variables that account for 45% of the input variables in each domain, which implies that SD contains less valid information for the transfer.

5. Application to a mass-production plant

The proposed method was applied to a mass-production plant in RICOH. There are 12 quality items to be predicted, including particle size distribution, particle shape, and charging characteristics. The numbers of variables and lots are shown in Table 1. The proposed prediction method was compared with two different methods using only TD (hereinafter, referred to as Target) and using only common input variables in SD and TD (referred to as Common). In these two methods, we used random forest, which has been used in the existing automatic quality control (Takahashi et al., 2020).

We conducted the prediction of the 12 qualities from the 11th lot to the 40th lot in TD. While Target and Common failed to achieve the prediction accuracy target for two and three quality items, respectively, the proposed method achieved the prediction accuracy target for all quality items. Besides, the proposed method outperformed Target and Common in all qualities. The prediction accuracy in RMSE of the proposed method was 11.4% higher than Target on average, and particularly 17.4% for quality #10. Compared to Common, the average improvement was 15.4%, and the best was 25.4% in quality #4.

Fig. 6 shows the predicted and measured values for each lot of quality #2. The predicted values of the proposed method follow the measured values better than those of Target

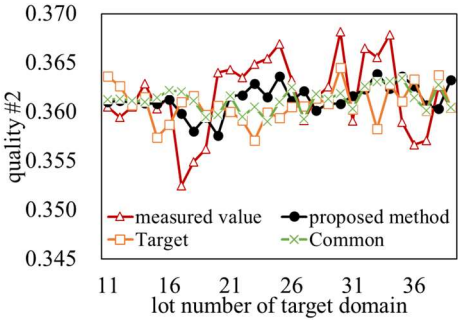


Fig.6. The prediction results of the proposed method, Target, and Common for quality #2.

and Common. In particular, the prediction errors in the initial stage for lots 18, 24, 25, and 26 are small. The prediction accuracy indices, i.e., RMSE/Δ, are 8.1% for the proposed method, 9.5% for Target, and 8.6% for Common, indicating that the proposed method satisfies the prediction target values for these small lots.

The toner qualities predicted by the transfer learning were used in the automatic quality control system based on feed-forward inference control (Takahashi et al., 2020) described in Section 2. Before applying prediction by the transfer learning, 40 lots (days) of data had to be accumulated to achieve the required accuracy target. With the proposed method, the data accumulation was reduced to 10 lots (days), and the person-hours required for monitoring and control by quality managers immediately after a change in equipment or raw materials were reduced by 75%.

6. Conclusions and future tasks

We first proposed a new transfer learning method that can cope with heterogeneous domain adaptation, i.e., FEHDA, which is simple extension of FEDA. Second, we proposed a new prediction method that combines Gaussian process regression (GPR) and bagging. Finally, the proposed method was adopted in the automatic control system of RICOH's chemical toner plant. The downtime of the automatic quality control system decreased from 40 lots (days) to 10 lots (days), and the person-hours required for manual quality control by toner quality managers have been reduced by 75%.

References

Daumé III, H., 2007. Frustratingly Easy Domain Adaptation. Proceedings of the 45th Annual Meeting of ACL 256–263.

Duan, L., Xu, D., Tsang, I.W., 2012. Learning with Augmented Features for Heterogeneous Domain Adaptation. Proceedings of the 29th International Conference on Machine Learning, ICML 2012 1, 711–718.

Kamishima, T., Hamasaki, M., Akaho, S., 2009. TrBagg: A simple transfer learning method and its application to personalization in collaborative tagging, in: Proceedings - IEEE International Conference on Data Mining, ICDM. pp. 219–228.

Khorami, H., Fgaier, H., Elkamel, A., Biglari, M., Chen, B., 2017. Multivariate Modeling of a Chemical Toner Manufacturing Process. Chemical Engineering and Technology 40, 459–469.

Takahashi, N., Miyakawa, M., Satoh, N., Watanabe, Y., Satoh, T., Kano, M., 2020. Automatic Quality Control for Chemical Toner Using Machine Learning Prediction Technology. Society of Instrument and Control Engineers 7th Multi Symposium on Control Systems 2A2-3, in Japanese.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50278-5

Towards An Automated Physical Model Builder: CSTR Case Study

Shota Kato and Manabu Kano*

Department of Systems Science, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, Japan

manabu@human.sys.i.kyoto-u.ac.jp

Abstract

Physical models are indispensable for the realization of digital twins, which are expected to enhance process design, operation, and optimization. The conventional physical model building relies entirely on experts and takes much time and effort. This arduousness has hampered the widespread use of digital twins. To overcome the difficulty and enable non-experts to build a practical physical model, we aim to develop an automated physical model builder, AutoPMoB. AutoPMoB conducts five tasks: 1) searching literature databases for documents relating to a target process, 2) converting the format of each document to HTML format, 3) extracting information required to build a physical model, 4) judging whether the extracted information is equivalent in different documents to unify the expressions, and 5) reorganizing the information to output a desired physical model. In the present study, we proposed an architecture of AutoPMoB and developed its prototype. By building a physical model of a continuous stirred-tank reactor, we have demonstrated that the prototype can automatically build a model that meets all requirements. AutoPMoB is expected to facilitate physical model building and foster the realization of digital twins.

Keywords: Artificial intelligence, First principle model, Process modeling, Natural language processing, Information extraction

1. Introduction

A digital twin is a core technology for realizing a cyber-physical system and has attracted much attention in recent years. It uses a model to reproduce the behavior of an actual plant and explore and predict unknown phenomena (Wang, 2020). Physical models based on the principles of chemistry, physics, and biology are applicable over a wide range of conditions, while statistical models should not be used outside the range of training data. In order to realize a digital twin, it is necessary to build an accurate physical model.

Conventionally, researchers and engineers with in-depth plant knowledge have surveyed the literature and built a process model that meets their demands. There are multiple pieces of literature to be investigated, and it is not easy to immediately find the equations that the desired model requires. Furthermore, when the accuracy of the model is inadequate, researchers and engineers need to improve its accuracy by trial and error. Thus, the conventional physical model building takes much toil.

To facilitate physical model building, we aim to develop an automated physical model builder, AutoPMoB. AutoPMoB extracts information of variables, formulas, experimen-

tal data, and prerequisites from documents and then integrates the information to build a desired physical model. AutoPMoB frees the engineers from laborious tasks of physical model building and provides access to information overlooked when manually processed.

In the present paper, we first propose an architecture of AutoPMoB and describe the fundamental technologies required for realizing it in section 2. We then develop a prototype of AutoPMoB and apply it for building a physical model of a continuous stirred-tank reactor (CSTR) in section 3.

2. Automated Physical Model Builder

AutoPMoB first retrieves documents concerning a target process from literature databases. Next, AutoPMoB extracts information necessary to build a physical model. AutoPMoB then unifies the notations so that the information with the same meaning is not written differently. AutoPMoB finally integrates the information to build a physical model that meets all requirements.

We can obtain the documents regarding the target process by using the existing search engine. Here, it is assumed that such documents are collected in advance. This section describes the fundamental technologies required for realizing AutoPMoB.

2.1. Document Format Conversion

The most widespread format of scientific digital documents is PDF. Extracting information directly from documents in PDF format using computers is difficult because PDF is designed for human viewing. On the other hand, documents in HTML format are relatively easy to extract information from because the information is tagged. In order to extract mathematical formulas, it is effective to represent the formulas in mathematical markup language (MathML) format (Ausbrooks et al., 2014). MathML is a standard format of web pages and is mainly used for information extraction and retrieval of mathematical expressions. Thus, AutoPMoB first converts the format of each document to HTML format. Documents in PDF format can be converted to those in HTML format with high accuracy using existing tools, such as InftyReader (Suzuki et al., 2003) and LaTeXML (Miller, 2018).

2.2. Information Extraction

Various types of information are required to build physical models, such as variables, formulas, experimental data, and prerequisites. The essential pieces of information are variables and equations.

AutoPMoB must identify the variable definitions. All variable symbols can be extracted based on their tags in MathML format, but the challenge is to extract the definition of each symbol accurately. Some scientific documents have a table describing symbols and corresponding definitions. For such documents, AutoPMoB can accurately extract the definitions using noun phrases next to the symbols in the table. On the other hand, some scientific documents do not have such a table, and AutoPMoB needs to extract variable definitions from sentences. Although some studies (Schubotz et al., 2016, 2017) proposed variable definition extraction methods, the accuracy of the existing methods is insufficient for practical use. We have been currently developing a method to extract the variable definitions accurately (Kato and Kano, 2020).

2.3. Equivalence Judgment and Unification

AutoPMoB judges whether the variable definitions and equations extracted from different documents are equivalent and unifies their expressions.

We can determine whether two noun phrases represent the same variable by calculating their similarity and checking whether it exceeds a threshold. A language model trained with a large corpus, such as BERT (Devlin et al., 2019), is known to achieve high performance on natural language processing (NLP) tasks. The model's performance varies depending on the corpus used for training. For example, Beltagy et al. (2019) released SciBERT, a BERT-based language model trained on a large corpus of scientific texts, and achieved higher performance on a range of NLP tasks in the scientific domain than BERT. To judge the equivalence of the variable definitions, a language model trained on a corpus consisting of documents relating to physical models of processes would be useful.

In order to judge the equivalence of equations, it is necessary to consider the calculation they perform. Existing computer algebra systems can judge the equivalence of two polynomials; for example, in Wolfram Language, $(x+1)^2$ and x^2+2x+1 are judged equivalent by expanding the former and then comparing the two polynomials (Wolfram Research, 2007). Physical models are mainly described by differential equations, algebraic equations, and partial differential equations. To our best knowledge, there has been no research focusing on the equivalence of equation groups consisting of multiple equations. We have been developing a method to judge the equivalence of two equation groups.

After judging the equivalence of variable definitions and equations, AutoPMoB unifies their expressions.

2.4. Integration & Scoring

AutoPMoB builds model candidates by integrating the unified equations. The desired model among the candidates varies depending on the purpose of model building. A user of AutoPMoB gives input variables of the model, and then AutoPMoB builds models whose numbers of degrees of freedom match the number of the input variables. Then, AutoPMoB scores and ranks the models because it takes time and effort to determine which one to choose when multiple models are built. The score is calculated based on the information, such as the number of documents containing the equations in the model and their citation information.

3. Case Study

3.1. Implementation

We developed a prototype of AutoPMoB using Streamlit, which is a Python library. Figure 1 shows a screenshot of the prototype. The prototype first takes HTML files as the input, extracts variables and equations from the files, and displays them. In this prototype, the variable definitions are extracted from a table in each file. A user selects variables required for the model and input variables. The prototype builds all models satisfying the requirements and shows them.

3.2. Dataset

Based on the documents (Uppal et al., 1974; Marlin, 2000), we created two files, 01.html and 02.html, including physical models of an ideal jacketed CSTR, where a first-order, exothermic, irreversible reaction (A→B) takes place. The equations consisting the models are as follows:

$$\frac{\mathrm{d}C_\mathrm{A}}{\mathrm{d}t} = \frac{F}{V}\left(C_\mathrm{A0} - C_\mathrm{A}\right) + r_\mathrm{A} \tag{1}$$

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{F}{V}\left(T_0 - T\right) + \frac{h_\mathrm{r}}{\rho C_\mathrm{p}} r_\mathrm{A} - \frac{Q}{V\rho C_\mathrm{p}} \tag{2}$$

$$-r_\mathrm{A} = k_0 \exp\left(-\frac{E}{RT}\right) C_\mathrm{A} \tag{3}$$

$$Q = \frac{aF_\mathrm{c}^{b+1}}{F_\mathrm{c} + \left(\dfrac{aF_\mathrm{c}^{b}}{2\rho_\mathrm{c} C_\mathrm{pc}}\right)} \left(T - T_\mathrm{cin}\right) \tag{4}$$

$$Q = UA\left(T - T_\mathrm{c}\right) \tag{5}$$

where the nomenclature is shown in Table 1. The model in 01.html consists of Eqs. (1)–(4), and the model in 02.html consists of Eqs. (1), (2) and (5).

3.3. Results & Discussion

Figure 1 shows the result when we selected five variables as the required ones and one variable as the input variable. This prototype built an accurate model that met our requirements. However, the prototype has several limitations. To realize AutoPMoB, we need to develop methods for 1) improving the accuracy of extracting variable definitions from sentences, 2) accurately judging the equivalence of variable definitions, and 3) scoring and ranking the models.

Table 1: Nomenclature.

Symbol	Definition	Symbol	Definition
A	heat transfer area	T	reactor temperature
C_p	specific heat of the reacting meterial	T_c	coolant temperature
C_pc	specific heat of the coolant	T_cin	inlet temperature of the coolant
C_A	reactor concentration of A	T_0	feed temperature
C_A0	inlet concentration of A	U	heat transfer coefficient
E	activation energy	V	reactor volume
F	feed flow rate	h_r	heat of reaction
F_c	coolant flow rate	ρ	density of the reacting material
k_0	reation rate constant	ρ_c	density of the coolant
R	universal gas constant	a	parameter
t	time	b	parameter

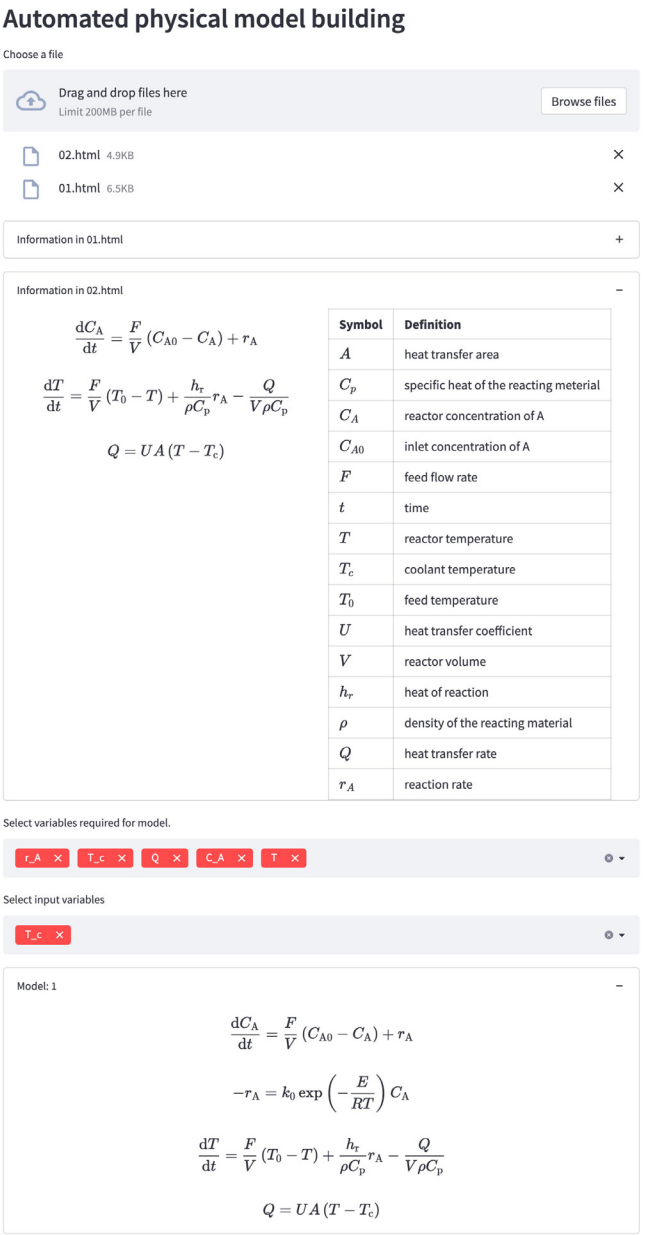


Figure 1: Screenshot of a prototype of an automated physical model builder, AutoPMoB.

4. Conclusions

To facilitate physical model building, we proposed an architecture for an automated physical model builder, AutoPMoB. AutoPMoB retrieves documents regarding a target process, extracts information such as variables and equations, and builds a physical model by

integrating the information. Furthermore, we developed a prototype of AutoPMoB and demonstrated its usefulness using documents on a CSTR. Although there remain some challenges, the realization of AutoPMoB will eliminate the barriers to physical model building and lead to the utilization of digital twins.

Acknowledgments This work was supported by JSPS KAKENHI Grant Number JP21K18849 and JST SPRING Grant Number JPMJSP2110.

References

R. Ausbrooks, S. Buswell, D. Carlisle, G. Chavchanidze, S. Dalmas, S. Devitt, A. Diaz, S. Dooley, R. Hunter, P. Ion, M. Kohlhase, A. Lazrek, P. Libbrecht, B. Miller, R. Miner, C. Rowley, M. Sargent, B. Smith, N. Soiffer, R. Sutor, S. Watt, 2014. Mathematical Markup Language (MathML) Version 3.0 2nd Edition. https://www.w3.org/TR/MathML3, (Accessed on 2021/11/06).

I. Beltagy, K. Lo, A. Cohan, 2019. SciBERT: A pretrained language model for scientific text. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). pp. 3615–3620.

J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). pp. 4171–4186.

S. Kato, M. Kano, 2020. Identifier information based variable extraction method from scientific papers for automatic physical model building. PSE Asia, Paper No. 210043.

T. E. Marlin, 2000. Process Control: Designing Processes and Control Systems for Dynamic Performance, 2nd Edition. McGraw-Hill.

B. Miller, 2018. LaTeXML The Manual—A LaTeX to XML/HTML/MathML Converter, Version 0.8.3. https://dlmf.nist.gov/LaTeXML/, (Accessed on 2021/11/06).

M. Schubotz, A. Grigorev, M. Leich, H. S. Cohl, N. Meuschke, B. Gipp, A. S. Youssef, V. Markl, 2016. Semantification of identifiers in mathematics for better math information retrieval. In: SIGIR 2016 - Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 135–144.

M. Schubotz, L. Krämer, N. Meuschke, F. Hamborg, B. Gipp, 2017. Evaluating and improving the extraction of mathematical identifier definitions. In: G. J. F. Jones, S. Lawless, J. Gonzalo, L. Kelly, L. Goeuriot, T. Mandl, L. Cappellato, N. Ferro (Eds.), Conference and Labs of the Evaluation Forum (CLEF), Dublin, Ireland. Vol. 10456. pp. 82–94.

M. Suzuki, F. Tamari, R. Fukuda, S. Uchida, T. Kanahori, 2003. Infty: an integrated ocr system for mathematical documents. In: Proceedings of the 2003 ACM symposium on Document engineering. pp. 95–104.

A. Uppal, W. Ray, A. Poore, 1974. On the dynamic behavior of continuous stirred tank reactors. Chemical Engineering Science 29 (4), pp. 967–985.

Z. Wang, 2020. Digital Twin Technology. Ch. 7, In: Industry 4.0 - Impact on Intelligent Logistics and Manufacturing.

Wolfram Research, 2007. Equal. https://reference.wolfram.com/language/ref/Equal.html, (Accessed on 2021/11/06).

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50279-7

Forward physics-informed neural networks for catalytic CO_2 methanation via isothermal fixed-bed reactor

Son Ich Ngo[a], and Young-Il Lim[a,*]

[a]*Center of Sustainable Process Engineering (CoSPE), Department of Chemical Engineering, Hankyong National University, Gyeonggi-do, Anseong-si, Jungang-ro 327, 17579 Korea*
limyi@hknu.ac.kr

Abstract

We developed physics-informed neural networks (PINNs) to solve an isothermal fixed-bed (IFB) model for catalytic CO_2 methanation. The PINN is composed of a feed-forward artificial neural network (FF-ANN) with two inputs and physics-informed constraints for governing equations, boundary conditions, initial conditions, and nonlinear reaction kinetics. The forward PINN showed excellent extrapolation performance for the IFB model. The calculation speed of the PINN surrogate model is faster significantly than a stiff ODE numerical solver. These results suggest that forward PINNs can be used as a surrogate model for chemical reaction kinetics.

Keywords: Catalytic CO_2 methanation; Fixed-bed reactor; Reaction kinetics; System identification; Machine learning; Physics-informed neural network.

1. Introduction

CO_2 methanation (Ngo et al., 2021) combining captured CO_2 with H_2 produced via water electrolysis (Kim et al., 2021) is an alternative to existing energy systems that could be integrated with renewable electricity sources. CO_2 methanation technologies could considerably reduce carbon emissions by encouraging industrial symbiosis from industries with large CO_2 footprints such as thermal power plants (Kim et al., 2021). Because CH_4 is easier to store and transport than H_2 (Ngo et al., 2021), the synergistic integration of renewable electricity with a natural gas grid is expected via CO_2 methanation (Miguel et al., 2018).

Despite of advances in first principles and empirical elucidations, artificial neural network (ANN) models in the category of data-driven models, black-box models, or surrogate models (SMs), have become an alternative functional mapping between input and output data because of their prompt predictions, automated knowledge extraction, and high inference accuracy (Abiodun et al., 2018, Gusmão et al., 2020).

Recently, ANNs and conservation equations coupled with automatic differentiation (AD) that solve ordinary differential equations (ODEs) and partial differential equations (PDEs), called physics-informed neural networks (PINNs), have been reported (Raissi et al., 2019). Because PINNs are constrained to respect any symmetries, invariances, or first-principle laws (Raissi et al., 2019), they present a potential for solving chemical engineering problems, which usually deal with complex geometries and physical phenomena. In contrast to common ANNs, PINNs do not depend on empirical data because the initial and boundary conditions are directly used to adjust the network

parameters such as weights and biases (Raissi et al., 2019). In addition, the extrapolation capability of PINNs is enhanced owing to physical constraints (Kim et al., 2020, Ngo and Lim, 2021). Nevertheless, there are few applications of PINNs in process modeling and chemical reactor design.

In this study, forward PINNs coupled with AD were developed for the solution and parameter identification of a highly nonlinear reaction rate model for catalytic CO_2 methanation in an IFB reactor. The results obtained from the PINNs were compared with those obtained using a common numerical solver of ODEs (ode15s in MATLAB). The extrapolation capability was analyzed by narrowing the collocation training domain and detaching the collocation training domain from the boundary. It was deomonstrated that the forward PINN solved fixed-bed models with highly nonlinear chemical reaction kinetics.

2. Isothermal fixed-bed reactor for CO_2 methanation

The single-tube IFB was assumed to be equipped with a heat exchanger that was able to transfer immediately the heat generated in the exothermic reactions to the coolant. The catalytic CO_2 methanation reaction, known as the Sabatier reaction, is (Ngo et al., 2020, Ngo and Lim, 2021)

$$CO_2 + 4H_2 \rightleftarrows CH_4 + 2H_2O, \Delta H_r^{298K} = -165 \text{ kJ·mol-1} \quad (1)$$

The operating conditions were set as a temperature (T) of 450 °C, a pressure (P) of 5 bar, and a volumetric flow rate (Q) of 10 Nm^3/s. The pure gas reactants were fed to the inlet at a CO_2/H_2 molar ratio of 1/4.

The mass balances for the i^{th} species ($i = CO_2$, H_2, CH_4, and H_2O) participating in the CO_2 methanation reaction in Eq. (1) are formulated as follows:

$$\frac{1}{A_t}\frac{dF_i}{dz} = \eta \nu_i r \quad (2)$$

where z (m) is the reactor tube axial position, F_i (mol/s) is the molar flow rate of a species i at position z, A_t (m^2) is the tube cross-sectional area, ν_i is the stoichiometric coefficient of species i, and r ($mol/m^3/s$) is the volumetric reaction rate. η is the effectiveness factor of the chemical reaction, which is defined as the volume-averaged reaction rate with diffusion within catalyst particles divided by the area-averaged reaction rate at the catalyst particle surface (Ngo et al., 2020).

The boundary conditions for the molar flow rate (F_i) of the species at the inlet ($z = 0$) are as follows:

$$F_i|_{z=0} = x_{i,0}F_0 \quad (3)$$

where $x_{i,0}$ and F_0 (mol/s) are the inlet mole fraction of gas species i and the total molar flow rate of the inlet gas mixture, respectivel. A reaction kinetics model proposed by (Koschany et al., 2016) for catalytic CO_2 methanation, which was tested within a wide range of Ni contents and industrial operating conditions, was adopted in this study.

$$r = \rho_{cat}(1-\varepsilon)k \cdot \frac{p_{H_2}^{0.31} p_{CO_2}^{0.16}}{1 + K_{ad}\frac{p_{H_2O}}{p_{H_2}^{0.5}}}\left(1 - \frac{p_{CH_4} p_{H_2O}^2}{p_{H_2}^4 p_{CO_2} K_{eq}}\right) \quad (4)$$

$$k = 6.41 \times 10^{-5} \exp\left(\frac{93.6}{R}\left(\frac{1}{555} - \frac{1}{T}\right)\right) \tag{5}$$

$$K_{ad} = 0.62 \times 10^{-5} \exp\left(\frac{64.3}{R}\left(\frac{1}{555} - \frac{1}{T}\right)\right) \tag{6}$$

$$K_{eq} = 137 \cdot T^{-3.998} \exp\left(\frac{158.7}{RT}\right) \tag{7}$$

where R (=8.314×10^{-3} kJ/mol/K) is the gas constant, T (K) is the temperature, p_i (bar) is the partial pressure of species i, k ($mol/g_{cat}/s$) is the reaction rate constant, K_{ad} (1/bar$^{0.5}$) is the adsorption constant, and K_{eq} is the thermodynamic equilibrium constant. The catalyst density (ρ_{cat}) was set to $2300 \times 10^3\ g_{cat}/m^3_{cat}$ (Koschany et al., 2016).

3. Forward PINN structure

The architecture of the forward PINN problem is shown in Fig. 1. The objective of the forward PINN problem is to solve the given governing equation with initial, boundary, and operating conditions. The 30,000 collocation points were used to train the governing equations over the reactor length ($0 < z \leq L$) except $z = 0$. The Dirichlet's boundary conditions fixed the value of $F_{i,0} = [97.74\ \ 378.9\ \ 0\ \ 0]$ mol/s at the reactor inlet ($z = 0$). The lower and upper bounds of η were 0 and 1, respectively.

The FF-ANN structure contained two inputs (z and η), four outputs (F_i), five hidden layers, and 128 neurons for each layer. The activation function (f_a) of hyperbolic tangent (*tanh*), was applied for each neuron. The weights ($w_{j,k}$) and biases ($b_{j,k}$) for the j^{th} hidden layer and the k^{th} neuron are adjusted to minimize the loss function (*Loss*). The AD for spatial derivatives ($\frac{dF_i}{dz}$) was calculated via the reverse accumulation mode which propagates derivatives backward from a given output (Güneş Baydin et al., 2018). The governing equations as the physics-informed part of the ANN included the reaction kinetic rate (r) in Eq. (4), the four ODEs in Eqs. (2), and the boundary conditions.

The optimized weights and biases (w^* and b^*) were obtained from the following optimization problem:

$$\{w^*, b^*\} = \underset{w,b}{\operatorname{argmin}}\{Loss = MSE_g(w,b) + MSE_b(w,b)\} \tag{8}$$

$$MSE_g(w,b) = \frac{1}{N_{train}} \sum_{j=1}^{N_{train}} \sum_{i=1}^{N_{comp}} \left| \frac{1}{A_t}\left(\frac{dF_i}{dz}\right)_j - \eta \nu_i r_j \right|^2 \tag{9}$$

$$MSE_b(w,b) = \frac{1}{N_{bnd}} \sum_{k=1}^{N_{bnd}} \sum_{i=1}^{N_{comp}} \left| F_{i,k}|_{z=0} - x_{i,0} F_0 \right|^2 \tag{10}$$

where MSE_g and MSE_b are the mean squared errors for the governing equation and boundary condition, respectively. N_{train}, N_{comp}, and N_{bnd} are the number of training data sets, species (or components), and boundary condition sampling points, respectively. The loss function (*Loss*) sums MSE_g and MSE_b.

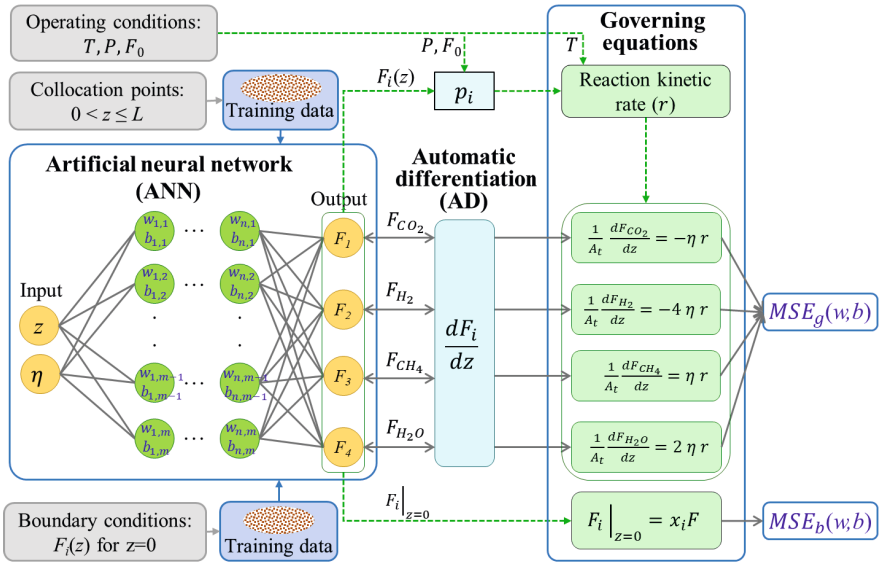


Figure 1. Structure of the physics-informed neural network (PINN) forward problem for CO_2 methanation in an isothermal fixed-bed (IFB) reactor.

An Adam optimizer (Kingma and Lei Ba, 2015) was used to solve Eq. (8), which combines a stochastic gradient descent with adaptive momentum, because of its good convergence speed (Rao et al., 2020). A mini-batch size of 128, which had a minor effect on the PINN training results, was used. The number of training epochs was set to 1,000. In the FF-ANN, the biases (b) were initialized to zeros and the weights (w) were initialized by the commonly used heuristic called Xavier's method (Xavier and Yoshua, 2010).

4. Results and discussion

Fig. 2 shows the performance of the forward PINN for 30,000 training data points in a limited range of z and 1,000 test data points in a full range of z ($0 \leq z \leq 2$) while η was fixed at 1. The collocation range of the training data starts from $z = 0$ and ends at $z = 0.5$ (Fig. 2a) and 1.0 (Fig. 2b). Even though the PINN was trained within one-sixth ($0 \leq z \leq 0.5$) of the full range, the PINN output ($F_{i,PINN}$) for the test data of the full range ($0 \leq z \leq 2$) agrees well the ODE solution ($F_{i,ODE}$) outside the training range (Fig. 2a).

Fig. 3 plots the performance of forward PINN for 30,000 training data points within the full range of $0 \leq z \leq 2$ and $0 \leq \eta \leq 1$. Fig. 3a and 3b show the interpolation for $\eta = 0.5$ and 1.0, respectively, whereas Fig. 3c demonstrats the extrapolation for $\eta = 1.5$. With the extrapolation of 150% higher than the trained bound, the PINN captures F_i with a prediction accuracy of 97.3%.

The extrapolation capability of the PINN is remarkable, unlike that of common ANNs (Abiodun et al., 2018). The accuracy of the PINN solution is closely related to the range and distribution of the training data (Jagtap et al., 2020). The forward PINN model is appropriate for solving governing equations with complex geometries or moving boundary conditions (Sun et al., 2020). In addition, numerical diffusion and round-off errors are minimized in PINNs with the aid of AD.

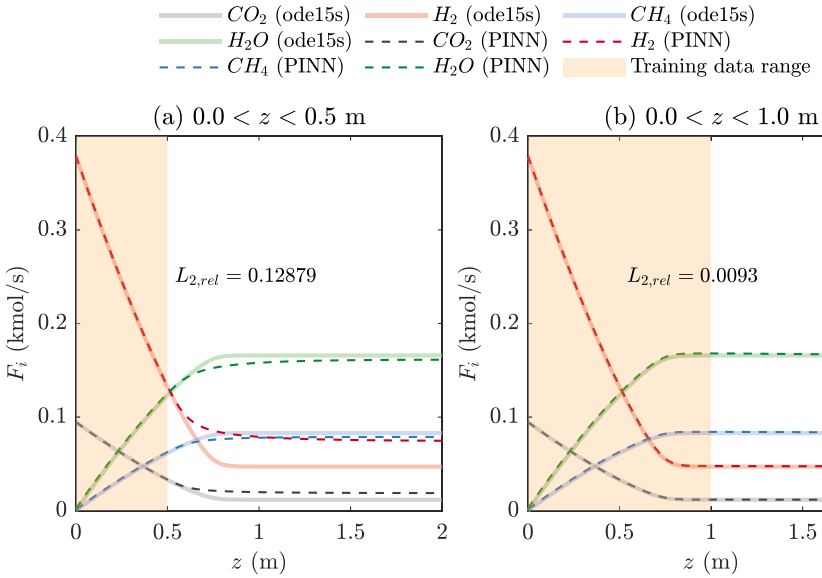


Figure 2. Performance of the forward PINN for 10,000 training data points in a limited range of the reactor length (z) and 1,000 test data points in the full range of z and $\eta = 1$.

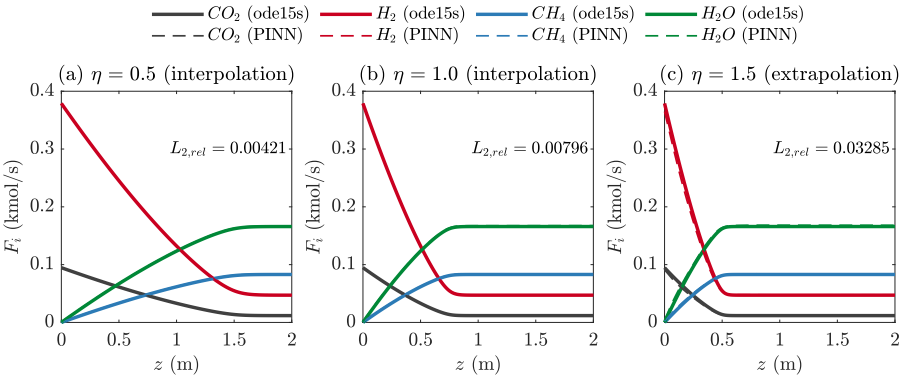


Figure 3. Performance of forward PINN for two inputs as η and z.

5. Conclusions

The physics-informed neural network (PINN) was developed for an isothermal fixed-bed (IFB) reactor model for catalytic CO_2 methanation. The PINN was composed of a feed-forward artificial neural network (FF-ANN), automatic differentiation (AD) for derivatives, and governing equations with a stiff reaction kinetic rate. The loss function of the PINN included two mean squared errors (MSEs) for the governing equations and boundary conditions. The one-dimensional reactor was initialized at a molar flow rate that was the same as the boundary condition at the reactor inlet.

The forward PINN model exhibited an excellent extrapolation performance because the PINN provided a solution satisfying physics-informed constraints. The current approach is useful for building a surrogate model for CO_2 methanation process design and optimization.

References

Abiodun, O. I., et al., 2018. State-of-the-art in artificial neural network applications: A survey. Heliyon, 4, 11, e00938.
Güneş Baydin, A., et al., 2018. Automatic differentiation in machine learning: A survey. Journal of Machine Learning Research, 18, 1-43.
Gusmão, G. S., et al., 2020. Kinetics-Informed Neural Networks. December.
Jagtap, A. D., et al., 2020. Conservative physics-informed neural networks on discrete domains for conservation laws: Applications to forward and inverse problems. Computer Methods in Applied Mechanics and Engineering, 365, 113028-113028.
Kim, J., et al., 2020. DPM: A Novel Training Method for Physics-Informed Neural Networks in Extrapolation.
Kim, S., et al., 2021. Coal power plant equipped with CO_2 capture, utilization, and storage: Implications for carbon emissions and global surface temperature. Energy and Environmental Science, in review.
Kim, S., et al., 2021. Effects of flue gas recirculation on energy, exergy, environment, and economics in oxy-coal circulating fluidized-bed power plants with CO_2 capture. International Journal of Energy Research, 45, 4, 5852-5865.
Kingma, D. P. and J. Lei Ba, 2015. Adam: A method for stochastic optimization. Iclr, 1-15.
Koschany, F., et al., 2016. On the kinetics of the methanation of carbon dioxide on coprecipitated $NiAl(O)_x$. Applied Catalysis B: Environmental, 181, 504-516.
Miguel, C. V., et al., 2018. Intrinsic kinetics of CO_2 methanation over an industrial nickel-based catalyst. Journal of CO2 Utilization, 25, 128-136.
Ngo, S. I. and Y.-I. Lim, 2021. Solution and Parameter Identification of a Fixed-Bed Reactor Model for Catalytic CO2 Methanation Using Physics-Informed Neural Networks. Catalysts, 11, 11.
Ngo, S. I., et al., 2020. Flow behaviors, reaction kinetics, and optimal design of fixed-and fluidized-beds for CO_2 methanation. Fuel, 275, 117886.
Ngo, S. I., et al., 2021. Flow behavior and heat transfer in bubbling fluidized-bed with immersed heat exchange tubes for CO_2 methanation. Powder Technology, 380, 462-474.
Ngo, S. I., et al., 2021. Experiment and numerical analysis of catalytic CO_2 methanation in bubbling fluidized bed reactor. Energy Conversion and Management, 233, 113863.
Raissi, M., et al., 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 378, 686-707.
Rao, C., et al., 2020. Physics-informed deep learning for incompressible laminar flows. Theoretical and Applied Mechanics Letters, 10, 3, 207-212.
Sun, L., et al., 2020. Surrogate modeling for fluid flows based on physics-constrained deep learning without simulation data. Computer Methods in Applied Mechanics and Engineering, 361, 112732-112732.
Xavier, G. and B. Yoshua (2010). Understanding the difficulty of training deep feedforward neural networks, PMLR. 9: 249-256.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50280-3

Hashing-based just-in-time learning for big data quality prediction

Xinmin Zhang[a]*, Jiang Zhai[a], Zhihuan Song[a], Yuan Li[b]

[a]*State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China*
[b]*Department of Information Engineering, Shenyang University of Chemical Technology, Shenyang 110142, China*
xinminzhang@zju.edu.cn

Abstract

In recent years, the just-in-time (JIT) predictive models have attracted considerable attention due to their ability to prevent degradation of prediction accuracy. However, one of their practical limitations is expensive computation, which becomes a major factor that prevents them from being used for big data quality prediction. This is because the JIT modeling methods need to update the local regression model using the relevant samples that are searched through the lineal scan of the database during online operation. To solve this issue, the present work proposes a novel hashing-based JIT (HbJIT) modeling method that is suitable for big data quality prediction. In HbJIT, a family of locality-sensitive hash functions is firstly used to hash big data into a set of buckets, in which similar samples are grouped on themselves. During online prediction, HbJIT looks up multiple buckets that have a high probability of containing similar samples of a query object through the intelligent probing scheme, uses the data objects in the buckets as the candidate set of the results, and then filters the candidate objects using a linear scan. After filtering, the most relevant samples are used to construct the local regression model to yield the prediction of the query object. By integrating the multi-probe hashing strategy into the JIT learning framework, HbJIT can not only deal with process nonlinearity and time-varying characteristics but also is applicable to large-scale industrial processes. Experimental results on real-world dataset have demonstrated that the proposed HbJIT is time-efficient in processing large-scale datasets, and greatly reduces the online prediction time without compromising on the prediction accuracy.

Keywords: Virtual sensor, soft-sensor, big data quality prediction, hashing-based just-in-time modeling.

1. Introduction

In the modern process industry, with the widespread utilization of distributed control systems and the Internet of Things, large amounts of process data have been collected. Data-driven soft-sensors are important tools in process industries for online prediction of some quality variables that generally cannot be automatically measured at all, or can only be measured sporadically, with high delay, or at high cost (Kadlec et al., 2009; Kano and Ogawa, 2010). In recent years, a wide variety of data-driven soft-sensors ranging from linear models to nonlinear models have been developed (Zhang et al., 2019; Zhang et al., 2020).

In practical applications, there is a challenging problem with soft-sensors, that is, the predictive performance of soft-sensors might deteriorate due to the time-varying characteristics of industrial processes. As reported in (Kano and Ogawa, 2010), model maintenance is thought to be one of the most important issues related to soft-sensors. A simple and effective solution to solve the problem of predictive model degradation is to use the just-in-time (JIT) learning methods. Since the JIT modeling methods have the advantages of handling process nonlinearity and time-varying characteristics, they have been extensively used in various fields (Liu et al., 2012; Jin et al., 2019). Despite the JIT methods have some successful applications, they suffer from the problem of high computational cost when used for big data quality prediction. The reason is that the JIT methods need to update the local regression model using the relevant samples that are searched through the lineal scan of the database during online operation. The calculation of relevant samples through the linear scan of the database is very time-consuming. However, the size of the data collected in the process industry is increasing vastly, and the process industry has entered the era of big data (Qin, 2014). Thus, it is important to design a new JIT modeling algorithm that is suitable for big data quality prediction.

To solve this issue, the present work proposes a novel hashing-based JIT (HbJIT) modeling method that is suitable for big data quality prediction. HbJIT is designed based on the multi-probe hashing scheme, which first hashes big data into multiple buckets, in which similar samples are grouped on themselves. To perform a quality prediction, HbJIT looks up multiple buckets that have a high probability of containing the similar samples of a query object through the intelligent probing scheme, uses the data objects in the buckets as the candidate set of the results, and then filters the candidate objects using a linear scan. After filtering, the most relevant samples are used to construct the local regression model to yield the prediction of the query object. As a fast adaptive soft-sensor, HbJIT can not only deal with process nonlinearity and time-varying characteristics but also is suitable for large-scale industrial processes. The effectiveness of the proposed HbJIT is evaluated on real-world large-scale dataset. The results demonstrate that HbJIT can significantly reduce the online prediction time without sacrificing much in terms of accuracy.

The remainder of this paper is organized as follows. Section 2 gives a brief description of the JIT modeling method and Gaussian process regression. The proposed HbJIT modeling method is presented in Section 3. Section 4 provides the experimental results on real-world large-scale dataset. Conclusion is given in Section 5.

2. Preliminaries

2.1. Just-in-time (JIT) modelling method

Generally, the prediction accuracy of soft-sensors may be degraded due to changes in process characteristics (Kano and Ogawa, 2010). The JIT modeling method can deal with changes in process characteristics as well as nonlinearity, and thus it can prevent degradation of prediction accuracy.

Different from traditional soft-sensor which builds a global model of the process in an offline manner, the JIT modeling method constructs a query-driven local model. More specifically, given a historical dataset, the JIT modeling method consists of three steps: (1) When an output estimate is required for a new query, it searches for relevant samples to the query in the reference dataset based on some similarity measures. The most popular similarity measure is the Euclidean distance. (2) A local model is built

using the relevant samples. (3) An output estimate is produced by the constructed local model, and then the constructed local model is discarded. When the next query sample arrives, one needs to follow the same steps as above to build a new local model.

2.2. Gaussian process regression (GPR)

In JIT modeling, the local model should be built using some regression methods. In this work, GPR is adopted to construct the local regression model. Given a dataset $S = \{\boldsymbol{X}, \boldsymbol{y}\} = \{(\boldsymbol{x}_1, y_1), \cdots, (\boldsymbol{x}_N, y_N)\}$, where $\boldsymbol{x} \in \mathbb{R}^M$ and $y \in \mathbb{R}$ denote any input-output pair. Let $f(\boldsymbol{x})$ denote a latent function which maps input $\boldsymbol{x}$ to output y. $\boldsymbol{f} = \big(f(\boldsymbol{x}_1), f(\boldsymbol{x}_2), \cdots, f(\boldsymbol{x}_N)\big)$ represents the function values of the input vectors. In GPR, the function $f(\boldsymbol{x})$ is regarded as a random variable following a Gaussian process (GP) prior distribution (Williams and Rasmussen, 2006). A GP is defined in terms of a positive definite kernel (or covariance) function $k(\boldsymbol{x}, \boldsymbol{x}')$ as follows

$$f(\boldsymbol{x}) \sim GP\big(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')\big) \tag{1}$$

$$m(\boldsymbol{x}) = \mathrm{E}[f(\boldsymbol{x})] \tag{2}$$

$$k(\boldsymbol{x}, \boldsymbol{x}') = \mathrm{E}\big[\big(f(\boldsymbol{x}) - m(\boldsymbol{x})\big)\big(f(\boldsymbol{x}') - m(\boldsymbol{x}')\big)\big] \tag{3}$$

Generally, $m(\boldsymbol{x})$ is zero-mean. According to the principles of GP, the distribution of $\boldsymbol{f}$ follows the zero-mean Gaussian distributions with covariance matrix $\boldsymbol{K} = \big(k(\boldsymbol{x}_i, \boldsymbol{x}_j)\big)_{ij}$:

$$p(\boldsymbol{f}|\boldsymbol{X}) = \mathrm{N}(\boldsymbol{0}, \boldsymbol{K}) \tag{4}$$

where $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ denotes the (i, j)-element of $\boldsymbol{K}$. Given a query object $\boldsymbol{x}_q$, the prediction of the function value $f(\boldsymbol{x}_q)$ (denoted as f_q) is given by

$$\bar{f}_{\boldsymbol{q}} = \boldsymbol{k}_q[\boldsymbol{K} + \delta_{\boldsymbol{\epsilon}}^{\boldsymbol{2}}\boldsymbol{I}]^{-1}\boldsymbol{y} \tag{5}$$

where $\boldsymbol{I}$ is an identity matrix and $\delta_{\boldsymbol{\epsilon}}^{\boldsymbol{2}}$ is the variance of the Gaussian noise term.

3. Hashing-based just-in-time (HbJIT) modeling method

Notice that a major drawback of the JIT modeling methods is that a high computational cost is required in order to search for the relevant samples from the database to construct the local regression model, when a query object is provided. Especially, when the database provided is large scale, the computational cost of the JIT modeling will be very high, leading to a large delay in online quality prediction. However, the amount of data collected in the process industry is exploding, and the process industry has entered the era of big data. In such a scenario, if the JIT modeling method is used for big data quality prediction, the online prediction time will be very long. To handle the computational cost challenge of the JIT modeling methods, the present work proposes a novel hashing-based JIT (HbJIT) modeling method that is applicable to big data quality prediction. HbJIT is designed based on the multi-probe LSH scheme, which first uses the hashing functions to hash the big data into a set of buckets, in which similar samples are grouped on themselves. To perform a quality prediction, HbJIT looks up multiple buckets that have a high probability of containing the nearest neighbours of a query

object through the intelligent probing scheme, uses the data objects in the buckets as the candidate set of the results, and then filters the candidate objects using a linear scan. After filtering, the most relevant samples are used to construct the local GPR model to yield the prediction of the query object. Different from the conventional JIT modeling method which calculates similar samples to the query in linear time with respect to the data size, the proposed HbJIT calculates similar samples in sub-linear time, and significantly speeds up the online quality prediction.

The proposed method is implemented in two steps. The first step is to construct the index data structure using the locality sensitive hashing (LSH) functions, and the second step is to construct the local GPR model based on the multi-probe scheme. More specifically, let $\boldsymbol{S}$ be the historical database, and D be the distance measure between two objects. A family of hash functions $\boldsymbol{H} = \{h: \boldsymbol{S} \to \boldsymbol{U}\}$ is called (r, cr, p_1, p_2)-sensitive if the following conditions are satisfied for any two data objects $\boldsymbol{x}_p, \boldsymbol{x}_q \in \boldsymbol{S}$ (Datar et al., 2004):

$$if\ D(\boldsymbol{x}_p, \boldsymbol{x}_q) \le r, \qquad then\ Pr_H[h(\boldsymbol{x}_q) = h(\boldsymbol{x}_p)] \ge p_1 \tag{6}$$

$$if\ D(\boldsymbol{x}_p, \boldsymbol{x}_q) > cr, \qquad then\ Pr_H[h(\boldsymbol{x}_q) = h(\boldsymbol{x}_p)] \le p_2 \tag{7}$$

where $c > 1$ is an approximation factor, $p_1, p_2 \in (0,1)$ represent two probability thresholds and satisfy $p_1 > p_2$. The characteristic of the LSH function is that similar objects have a higher probability of being hashed to the same bucket than distant ones. The family of LSH functions based on p-stable distributions is considered as follows

$$h(\boldsymbol{x}) = \left\lfloor \frac{\boldsymbol{a}^T \boldsymbol{x} + b}{r} \right\rfloor \tag{8}$$

where $\boldsymbol{a}$ is a d-dimensional random vector whose entries are drawn independently from a p-stable distribution, b is a real number drawn uniformly from the range $[0, r]$. In practical applications, in order to construct the index data structure with high search precision, multiple hash tables need to be built, and each hash table contains multiple LSH functions. Let L and M denote the number of hash tables and LSH functions, respectively. A concatenation of M LSH functions in a hash table can be represented as

$$g(\boldsymbol{x}_p) = \left(h_1(\boldsymbol{x}_p), \cdots, h_M(\boldsymbol{x}_p)\right) \tag{9}$$

For L hash tables, we can construct L independent copies of $g_1(\boldsymbol{x}), \cdots, g_L(\boldsymbol{x})$. During the stage of constructing the index data structure, each data object in the database is hashed into one of the hash buckets of $g_1(\boldsymbol{x}), \cdots, g_L(\boldsymbol{x})$ and stored.

During the online prediction phase, when a query $\boldsymbol{x}_q$ is coming, we need to find the most k relevant samples of $\boldsymbol{x}_q$, which is used for local GPR modeling. Generally, the basic LSH algorithm will compute $g_1(\boldsymbol{x}_q), \cdots, g_L(\boldsymbol{x}_q)$, and search all these L buckets to get a set of candidates. Then, this candidate set is pruned through the linear scan algorithm to obtain the most k relevant samples of $\boldsymbol{x}_q$. It is worth noting that in practical applicaitons, a large number of hash tables need to be built in order to obtain higher search precision. A large number of hash tables results in a memory footprint. When the space requirement of the hash tables exceeds the size of the main memory, disk I/O may

be required to look up the hash bucket, which causes a large amount of delay in the query process. One of the solutions to solve this issue is to use a multi-probe scheme (Lv et al., 2007), which can significantly reduce the memory footprint of the LSH data structure. The intuition of multiprobe LSH is that it probes multiple buckets in each hash table instead of building many different hash tables. This is realized by using an intelligently generated probing sequence which probes multiple buckets that are likely to contain the nearest neighbours of the query object in each hash table. According to the characteristics of LSH, if a data object is close to $\boldsymbol{x}_q$, but is not hashed into the same bucket as $\boldsymbol{x}_q$, then it is likely to be in a nearby bucket. Based on this principle, multiprobe LSH defines a probeing sequence $(\Delta_1, \Delta_2, \cdots \Delta_J)$ where $\Delta_1 = (\delta_m, \cdots, \delta_M)$ is a hash perturbation vector with $\delta_m \in \{-1,0,1\}$, $m = 1, \cdots, M$, and $j = 1, \cdots, J$. Given $\boldsymbol{x}_q$, the basic LSH probes the hash bucket $g_l(\boldsymbol{x}_q)$, while multiprobe LSH probes not only the bucket $g_l(\boldsymbol{x}_q)$ but also the buckets $g_l(\boldsymbol{x}_q) + \Delta_1, \cdots, g_l(\boldsymbol{x}_q) + \Delta_J$. These buckets are ordered according to the score of the perturbation vector, which is defined as

$$score(\Delta_j) = \sum_{m=1}^{M} x_m(\delta_m)^2 \tag{10}$$

where $x_m(\delta_m)$ denotes the distance from $\boldsymbol{x}_q$ to the bundary of the bucket $h_m(\boldsymbol{x}_q) + \delta_m$. For the perturbation vectors that have smaller scores, they will have higher probability of yielding data objects near to the query object $\boldsymbol{x}_q$.

4. Case studies

This section describes the experimental results of the proposed HbJIT-GPR method on a large-scale real-world blast furnace ironmaking dataset, and compared it against the GPR and JIT-GPR$_{(brute)}$ methods. JIT-GPR$_{(brute)}$ denotes the brute-force linear scan that is used to select similar samples. The experimental results are evaluated from two aspects: prediction accuracy and speed. The prediction accuracy is measured in terms of the root mean squared error (RMSE) criterion. The prediction speed is measure by the online prediction time.

The experiments were carried out in the blast furnace ironmaking process, which is a typical nonlinear time-varying process (Geerdes et al., 2020). It is difficult to accurately control the blast furnace to produce hot metal with consistent quality, because the harsh operating circumstances prevent the inside chemical heat from being directly detected. silicon content of hot metal is an important index indicating the chemical heat of molten iron. To carry on a steady operation of the blast furnace and produce hot metal with consistent quality, it is important to predict silicon content in real time.

To construct the prediction model, the collected dataset is randomly separated into a training dataset (80k samples), a validation dataset (10k samples) and a test dataset (10k samples). The dataset contains 110 process variables. The key parameters of the proposed HbJIT-GPR are selected using the grid search method in terms of the RMSE criterion on the validation set. Table 1 summarizes the prediction accuracy and prediction time for the silicon content by all methods on the testing dataset. GPR obtained the worst prediction accuracy although the online prediction time is the shortest. Compared with GPR, JIT-GPR$_{(brute)}$ has higher prediction accuracy and longer online prediction time. In comparison, the proposed HbJIT-GPR achieved the similar

prediction accuracy as JIT-GPR$_{(brute)}$, but provided shorter online prediction time. As a result, the prediction accuracy couple with the reduction in online prediction time clearly demonstrated the success of the proposed HbJIT-GPR in handling large-scale dataset.

Table 1. Prediction results for the silicon content on the testing dataset.

Methods	RMSE	Time (ms)
GPR	0.0211	121
JIT-GPR$_{(brute)}$	0.0126	460
HbJIT-GPR	0.0127	350

5. Conclusions

In this paper, a novel hashing-based JIT (HbJIT) modeling method was proposed for big data quality prediction. HbJIT is a fast adaptive soft-sensor that can not only deal with process nonlinearity and time-varying characteristics but also be applicable to large-scale industrial processes. The usefulness of the proposed method was verified through an industrial blast furnace ironmaking process. The experimental results show that the proposed method can reduce the online prediction time by a huge amount without sacrificing much in terms of accuracy.

References

Kadlec, P., Gabrys, B. and Strandt, S., 2009, Data-driven soft sensors in the process industry, Computers & Chemical Engineering, 33(4), 795-814.

Kano, M., Ogawa, M., 2010. The state of the art in chemical process control in japan: Good practice and questionnaire survey. Journal of Process Control, 20, 969–982.

Zhang X., Kano M., Matsuzaki S., 2019. A comparative study of deep and shallow predictive techniques for hot metal temperature prediction in blast furnace ironmaking. Computers & Chemical Engineering, 130, 106575.

Zhang X., Kano M., Tani M., et al., 2020, Prediction and causal analysis of defects in steel products: Handling nonnegative and highly overdispersed count data, Journal of chemical engineering of Japan, 95, 104258.

Liu, Y., Gao, Z., Li, P., Wang, H., 2012. Just-in-time kernel learning with adaptive parameter selection for soft sensor modeling of batch processes. Industrial & Engineering Chemistry Research 51, 4313–4327.

Jin, H., Pan, B., Chen, X., Qian, B., 2019. Ensemble just-in-time learning framework through evolutionary multiobjective optimization for soft sensor development of nonlinear industrial processes. Chemometrics and Intelligent Laboratory Systems 184, 153–166.

Qin, S.J., 2014. Process data analytics in the era of big data. AIChE Journal 60, 3092–3100.

Williams, C.K. , Rasmussen, C.E. , 2006. Gaussian processes for machine learning. The MIT Press 2 (3), 4 .

Buhler, J., 2001. Efficient large-scale sequence comparison by locality-sensitive hashing. Bioinformatics 17, 419–428.

Datar, M., Immorlica, N., Indyk, P., Mirrokni, V.S., 2004. Locality-sensitive hashing scheme based on p-stable distributions, in: Proceedings of the twentieth annual symposium on Computational geometry, pp. 253–262.

Geerdes, M., Chaigneau, R., Lingiardi, O., 2020. Modern Blast Furnace Ironmaking: An Introduction (2020). Ios Press.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50281-5

Physics-Constrained Autoencoder Neural Network for the Prediction of Key Granule Properties in a Twin-Screw Granulation Process

Chaitanya Sampat[a] and Rohit Ramachandran[a,*]

[a]*Chemical and Biochemical Engineering, Rutgers, the State University of New Jersey, Piscataway, NJ 08823, USA*
rohitrr@soe.rutgers.edu

Abstract
With the advancement of digitization of industrial manufacturing, there has been an increase in the application of machine learning methods to model these processes. These data-driven models are multivariate in nature and on occasion may not deliver the accuracy that can be obtained from first-principle models. The statistical approach in data-driven models is completely data-dependent and may give erroneous or undesired results due to noisy and incomplete database. Though accurate, first-principle models are often slow to simulate and lack the ability to predict data in real-time (Chen et al., 2020). Thus, to obtain real-time process predictions with accuracy similar to first-principle models, there is a need to develop data-driven models with first principle-based process constraints within their framework. In this study, several experimental datasets for twin-screw granulators (TSG) were considered. The data for 13 different TSGs was collected from previously published studies. The collected data was sorted for process parameters, material properties and geometric conditions of the study. An autoencoder neural network was developed to model these processes. The output from this model not only predicted the data well but also showed granule growth characteristics with the output properties obeying first-principle laws. The encoding section of the neural network helped find correlated inputs creating a reduced order model and captured information about the underlying physics of the process.
Keywords: Physics constrained neural network; autoencoders; twin screw granulation; Physics informed neural networks; PINN; PCNN

1. Introduction
Wet granulation is the process of agglomeration of fine powder into larger granules by adding a liquid binder. These granules help achieve desired quality attributes which can aid in improved flow, better dissolution rates, and better compression characteristics (Iveson et al., 2001). Wet granulation find application in various powder processing industries like mineral processing, agricultural products, detergents, food, and pharmaceuticals. It is an important unit operation in downstream oral dosage manufacturing in the pharmaceutical industry to more uniform distribution and dissolution characteristics. Previously wet granulation has been performed in a batch manufacturing scenario where, powder was mixed using an impeller and a liquid binder was sprayed using a liquid binder. This high-shear granulation can produce less compressible granules and operate in a very narrow range (Kumar et al., 2013). These challenges were overcome by converting this batch process to continuous manufacturing process.
Twin-screw granulation (TSG) is a widely used continuous wet granulation process. This equipment consists of a barrel which contain 2 co-rotating screws along parallel axes helping in the transfer of material along its length (Seem et al., 2015). These

screws are made up of smaller several screw elements which can help alter the flow of the material along the axis and can aid the mixing, breakage and other mechanisms which can affect the CQAs of the final granules. TSG can also support higher production volume compared to a batch granulator. TSGs have a larger design space due to the large number of independent operating parameters. This results in a large design of experiments which needs to be performed for optimization of the process to obtain the desired granule critical quality attributes (CQAs). Performing large number of experiments in early-stage process development when large amounts of active pharmaceutical ingredient (API) may not be available. Thus, there is a need for development of generalized models that can predict the outcome of the TSG. This model would need to be trained on a large data set of experiments which incorporates the effects of various independent operating parameters on the final granule CQAs and process outputs.

Neural networks with their dense structure have proven to be able to capture complicated relationships between inputs and outputs accurately. These neural networks can also be used to create reduce order models for faster prediction of these processes. Recently, to improve the prediction of neural networks for more complex physical processes, physical information about the process has been added to supplement its training (Mao et al., 2020; Raissi et al., 2019). Other studies have focused on constraining outputs of the neural networks with physics-based boundaries to make better informed models which, have the ability to accurately predict process outcomes(Zhu et al., 2019). These physics-based boundaries when incorporated into the loss these neural networks, help the model learn the underlying physics of the process leading to accurate predictions and more reliability under uncertain process conditions (Sampat and Ramachandran, 2021). These physics-based boundaries can be added to both the representation loss as well as the supervised loss, which leads to the addition of an extra loss function to the training. They have also resulted in neural networks requiring less data to train, which is especially useful with TSGs as this would reduce the amount of experimental data required.

In this work, a physics-constrained autoencoder (PCSAE) network was developed to create a reduced order model to represent a complete TSG process. Experimental data from 13 previously published literature was collected for various operating process parameters, process outputs and granule CQAs. The boundary conditions for each output were determined and were added the loss function of the developed PCSAE network. Sensitivity analysis was also performed on the PCSAE to determine whether it was able to capture the physical information about the process.

2. Methods

2.1. Data collection and completion methods

Twin-screw granulators with a wide design space have a large number of process parameters and geometry which can be varied. These variations when combined with changes in formulation can lead to an almost infinite combinations which can make the development of a general model for TSG very complicated. To incorporate all these effects a detailed data collection is required. In this study, data was collected from 9 different previously published experiments encompassing changes in formulation, process parameters and geometry(Dhenge et al., 2013, 2012; Kumar et al., 2016; Meier et al., 2017; Meng et al., 2019; Mundozah et al., 2020; Shirazian et al., 2017; Vercruysse et al., 2012). A total of 227 data points were collected for the creation of the model. Granule growth within a granulation process can be inferred from the process

outcomes and critical quality attributes (CQAs) of the granules obtained. Some of the process outcomes of the TSG process commonly studied are residence time distribution (RTD), mixing, torque inside the system, while granule size distribution (GSD) and granule density / porosity are the commonly studied granule quality attributes (Seem et al., 2015). Table 1 lists all the input parameters and outputs collected from each of the sources to develop the PCSAE model for a TSG process. In some literature, outputs had been reported in figures and each figure was processed individually for relevant data. The data from each plot was extracted using WebPlotDigitizer (Rohatgi, 2021). The data was split in the ratio of 3:1 for training and validation.

Table 1: Inputs and output monitored for the development of the PCSAE model

Input Parameters			**Output Parameters**
Geometry	**Process**	**Material**	
Number of CE and KE (nCE,nKE)	L/S ratio	Initial PSD	Granule size distribution
Staggering angle (SA) of KE	Screw Speed	Binder viscosity	Torque
L/D Ratio	Feed Rate	% API in powder	Mean Residence Time
Granulator diameter	Temperature		
Liquid addition position			

In this study, a multivariate linear regression was used since the torque and MRT values for a TSG are dependent on several inputs instead of only a single input. The regression model ($\boldsymbol{Y} = B\boldsymbol{X} + X_i$) was trained using the *sklearn* (Pedregosa et al., 2011) package in Python. The regression model used the existing data for torque and MRT for training. Here, $\boldsymbol{Y}$ is the response matrix of size $n \times p$, $\boldsymbol{X}$ is the matrix containing all predictors with size of $n \times (q+1)$. $\boldsymbol{B}$ is a $(q+1) \times p$ matrix of fixed parameters, X_i is the intercept matrix of size $n \times p$. Here n represents the number of observations, q are the number of inputs or predictors and p represents the number of responses or outputs. This model is often referred to as deterministic regression imputation. Such an imputation can add a bias to the predictions. To remove such biases, uncertainty can be added back to these models.

2.2. Development of Physics-constrained supervised auto-encoders (PCSAE)

An auto-encoder (AE) is a neural network which output are the same as the inputs and during its training for reconstruction, they can extract underlying attributes which can enable accurate predictions (Le et al., 2018). Single-layer AEs with linear activation functions are equivalent to principal component analysis, moreover non-linear auto-encoders have found to extract key attributes (Vincent et al., 2010). A supervised auto-encoder (SAE) is an AE with the addition of a supervised loss on the representation layer. A single linear layer SAE would perform like a partial least square method. The addition of a supervise loss to the AE better directs the representation learning.

The PCSAE model for this study was developed in Python v3.7.6 using Keras (Chollet et al., 2015). Keras is a wrapper used for machine learning package Tensorflow (Martin Abadi et al., 2015) developed by Google. The network had 12 input nodes which were divided into three separate groups as shown in Table 1. This helped create the 3 different reduced dimensional bottleneck layers representing each of the group individually. This bottleneck layer was then used for both reconstruction of the inputs as well as training the outputs with the physical constraints. The output physical constraints were obtained using physics-based boundaries. Maximum value boundaries

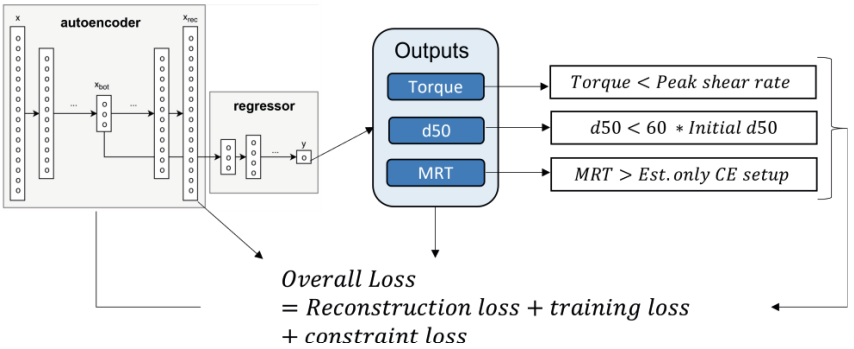


Figure 1: Physics-constrained supervised auto-encoder (PCSAE) model

for the median granule diameter and torque were determined using an empirical correlation and peak shear rate respectively, while a minimum boundary value was determined for the mean residence time (MRT). The minimum value was based on a screw configuration consisting only of conveying elements, which are known to aid conveying of material with little to no back-mixing. No physical constraints were introduced for the reconstruction. Figure 1 contains a schematic detailing the PCSAE model used. A single encoding layer with four nodes was used for the three individual inputs layers, a single decoding layer was used for reconstruction with eight nodes. Four layers were used for prediction of the outputs of the TSG with four nodes in each layer. All layers used the *'tanh'* activation function. The 'Adam' optimizer was used for optimization of the PCSAE with a learning rate of 0.008.

3. Results

3.1. Performance of the PCSAE

The total loss for the system was calculated as the summation of the reconstruction loss, training loss and the error due to the physical boundary constraints. These losses help aid the training of the system and prevent over-fitting of the model. For the PCSAE trained model no over-fitting was observed. The coefficient of determination (R^2) for prediction and reconstruction of outputs and inputs were 0.64 and 0.86 respectively. These values indicate that PCSAE was accurate to reconstruct the inputs to the model while the prediction of the outputs may not always be accurate. Figure 2(a) represents a parity plot for the predicted values of the outputs v/s the actual experimental values and it can be observed that some of the points are away from the $x = y$ line, indicating low accuracy. Figure 2(b) illustrates the parity plot for the reconstruction of the inputs and with an even spread across the $x = y$ line. The accuracy for the output prediction can be

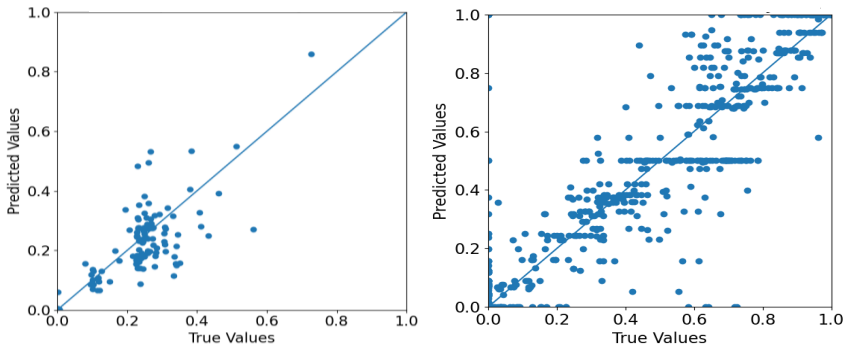


Figure 2: Parity plots for (a) Output prediction (b) reconstruction.

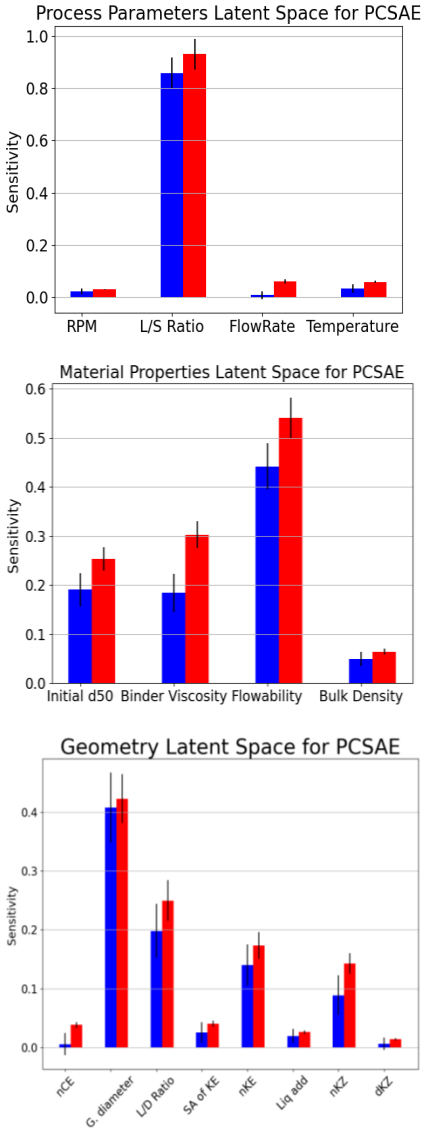


Figure 3: Sensitivity analysis of each input on the reduced dimension layer. The blue bars represent direct effects while the red bars indicate the total effect of the inputs.

improved in several ways including a deeper neural network for regression of the outputs, optimizing the hyperparameters of the neural network, improving the boundaries conditions.

3.2. Sensitivity analysis of PCSAE

To understand the effect of individual inputs on the outputs it is necessary to study their effects on the individual reduced dimension nodes. These nodes in-turn used to predict the outputs as well as reconstruct the inputs. It is vital for a model to capture the physics of the system well such that it considers the effect of each input. In this study, a Sobol sensitivity analysis was performed by varying the inputs across the range of values found in the literature. This sensitivity was compared against a supervised autoencoder without physical boundary constraints, and it was found that the PCSAE's sensitivity captured more physical information about the process than the non-physics constrained autoencoder. Figure 3 shows the sensitivity of the inputs on the reduced dimensions nodes. The effect of L/S ratio and RPM seems to the highest from the process parameters, while the contribution of different material properties seems to be almost equal, and the effect of staggering angle is the most prominent from the geometry parameters. These effects have been studied in literature and are in close accordance with the observed results. The effects of the inputs in the normal autoencoder system were observed to be skewed and did not align with physical observations.

4. Conclusions

Twin-screw granulation is a complicated process with an infinite number of combinations possible for its operation. With the help of developed physics-constrained autoencoder model, we were able to incorporate all the effects into a single model. This robust modelling framework is required could reduce the number of dimensions of the inputs to 3 segregated latent spaces for better process understanding. This framework which has been trained on several experimental datasets was able to capture the underlying physics of the system with accuracy of ~65%. The model was able to

identify key inputs affecting the outputs which may not be captured using a regular autoencoder. The overall performance of the model can further be increased by optimizing the neural network structure and including more datasets with larger variations in the inputs. This model can further be used to reduce experimentation by supplementing the design of experiments. Prediction of the latent spaces could be used to assess the granule growth regimes and identify experiments which would yield desired granule CQAs resulting in material and cost saving. This framework could also be adapted to different unit operations with changes in boundary conditions for desired outputs for better cost saving during process development.

References

Chen, Y., Yang, O., Sampat, C., Bhalode, P., Ramachandran, R., Ierapetritou, M.G., 2020. Digital Twins in Pharmaceutical and Biopharmaceutical Manufacturing : A literature review. Processes 8, 1–33.

Chollet, F., others, 2015. Keras. Available at: https://keras.io

Dhenge, R.M., Cartwright, J.J., Hounslow, M.J., Salman, A.D., 2012. Twin screw wet granulation: Effects of properties of granulation liquid. Powder Technol. 229, 126–136.

Dhenge, R.M., Washino, K., Cartwright, J.J., Hounslow, M.J., Salman, A.D., 2013. Twin screw granulation using conveying screws: Effects of viscosity of granulation liquids and flow of powders. Powder Technol. 238, 77–90.

Kumar, A., Dhondt, J., Vercruysse, J., De Leersnyder, F., Vanhoorne, V., Vervaet, C., Remon, J.P., Gernaey, K. V., De Beer, T., Nopens, I., 2016. Development of a process map: A step towards a regime map for steady-state high shear wet twin screw granulation. Pow. Tech. 300, 73–82.

Mao, Z., Jagtap, A.D., Karniadakis, G.E., 2020. Physics-informed neural networks for high-speed flows. Comput. Methods Appl. Mech. Eng. 360, 112789.

Martin Abadi, and others, 2015. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems.

Meier, R., Moll, K.P., Krumme, M., Kleinebudde, P., 2017. Impact of fill-level in twin-screw granulation on critical quality attributes of granules and tablets. Eur. J. Phar. Biop. 115, 102–112.

Meng, W., Román-Ospino, A.D., Panikar, S.S., O'Callaghan, C., Gilliam, S.J., Ramachandran, R., Muzzio, F.J., 2019. Advanced process design and understanding of continuous twin-screw granulation via implementation of in-line process analytical technologies. Ad. Pow. Tech. 30, 879–894.

Mundozah, A.L., Yang, J., Omar, C., Mahmah, O., Salman, A.D., 2020. Twin screw granulation : A simpler re-derivation of quantifying fill level. Int. J. Pharm. 591, 119959.

Pedregosa, F.and others, 2011. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2835--2830.

Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J. Comput. Phys. 378, 686–707.

Rohatgi, A., 2021. Webplotdigitizer: Version 4.5.

Sampat, C., Ramachandran, R., 2021. Identification of granule growth regimes in high shear wet granulation processes using a physics-constrained neural network. Processes 9.

Seem, T.C., Rowson, N.A., Ingram, A., Huang, Z., Yu, S., de Matas, M., Gabbott, I., Reynolds, G.K., 2015. Twin screw granulation - A literature review. Powder Technol. 276, 89–102.

Shirazian, S., Kuhs, M., Darwish, S., Croker, D., Walker, G.M., 2017. Artificial neural network modelling of continuous wet granulation using a twin-screw extruder. Int. J. Pharm. 521,102–109.

Vercruysse, J., Córdoba Díaz, D., Peeters, E., Fonteyne, M., Delaet, U., Van Assche, I., De Beer, T., Remon, J.P., Vervaet, C., 2012. Continuous twin screw granulation: Influence of process variables on granule and tablet quality. Eur. J. Pharm. Biopharm. 82, 205–211.

Zhu, Y., Zabaras, N., Koutsourelakis, P.S., Perdikaris, P., 2019. Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data. J. Comput. Phys. 394, 56–81.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50282-7

CSTR control with deep reinforcement learning

Borja Martínez[a], Manuel Rodríguez[a*] and Ismael Díaz[a]

[a]*Department of Chemical Engineering, Universidasd Politécnica de Madrid, José Gutierrez Abascal 2, Madrid 28006, Spain*
manuel.rodriguezh@upm.es

Abstract

In this paper we have applied the use of the deep reinforcement learning (DRL) for process control, to explore its applicability. The main objective is to develop a controller based on the deep reinforcement learning methodology in order to keep the level and composition of a continuous stirred-tank reactor under control.

Keywords: Deep Reinforcement learning; process control.

1. Introduction

The Oil&Gas and the process industry use mostly decentralized PID control and in some units Model Predictive Control. Both are mature technologies and well established in industry practices. But still, many PIDs are badly tuned and the costs of developing a model based predictive control are high. This classic control is difficult to implement and requires a lot of resources in complex processes that are not easy to model.

Today, machine learning has had a new outburst and specifically reinforcement learning. The use of neural networks with reinforcement learning, in what is called deep reinforcement learning has shown astounding results in some domains, such as games, (where the machine called AlphaGo created by DeepMind defeated the world champion of the game Go), self-driving cars, medicine o process control.

This paper is focused in the last one, specifically for the case of a continuous stirred-tank reactor with the aim of improving its classical process control, since this could have implications, such as replacing existing process control technology, mitigating the limitations of Model Predictive Control or helping to manage controller settings.

The structure of this paper has two distinct parts: the first one presents a brief explanation of the deep reinforcement learning methodology and its application in process control (section 2), and the second one shows the case study in which the work has been carried out, with the corresponding results and conclusions obtained (sections 3 to 5).

2. Reinforcement Learning and Deep Reinforcement Learning

Reinforcement learning is one of the three methodologies, together with supervised learning and unsupervised learning, that make up Machine Learning (Bishop, 2006). Unlike the other two, reinforcement learning does not use a set of labelled data, but rather the agent learns a task by interacting with the surrounding environment and evaluating the actions performed (Sutton and Barto, 2018). Schematically, the process carried out in reinforcement learning is illustrated in Fig 1.

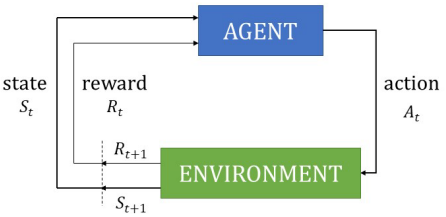


Fig. 1. Reinforcement Learning loop

The figure represents a cycle between the agent and the environment that starts with the observation of an action at time instant t, performed by the agent on the environment. This produces a series of changes at future instants (t+1) for both the reward and the new state of the environment, so that the cycle would continue with a new action by the agent (A_{t+1}).

The aim of all this is for the agent to be able to learn the relationship between the actions executed and the states obtained, known as policy π (A_t=a | S_t = s), and to maximise the long-term value of the rewards according to the value function v_π(s):

$$v_\pi(s) = E\{R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} \dots | S_t = s\} \quad (1)$$

The parameter γ ϵ [0, 1], denoted as discount factor, determines the behaviour of v_π(s) by prioritising long-term rewards.

Through this value function, the algorithms update their parameters iteratively with the intention of improving the policy associated with this function. In such a way that if the agent performs good actions, so that its policy is good and improves, it means that the values of the associated function will be greater at each iteration.

It is therefore conceivable that this function may reach a maximum value when the resulting policy is optimal. Optimal Control Theory supports this premise by means of Bellman's equation:

$$v_*(s) = \max_a \sum_{s',r} p(s', r | s, a) [r + \gamma v_*(s')] \quad (2)$$

Here the term v_* refers to the optimal policy value function, and *p(s', r |s, a)* is the transition probability. This factor indicates the probability that the environment transitions to a new state *s'* and offers a reward *r*, when the environment is in the previous state *s* and the action *a* has been executed.

Once the optimal value function is known, the associated optimal policy π_*(s) can be found using the transition model (Shin, J. et al., 2019). The learning procedure is based on the Actor-Critic methodology, whereby according to the existing policy, an action is chosen to be performed on the environment (Actor), and subsequently evaluated, based on the reward issued by the environment (Critic). After this evaluation, the parameters governing the Actor policy are readjusted for immediate future actions.

This process can be used with neural networks (deep reinforcement learning, DRL) to eliminate the need to store all state and value pairs and allow the agent to estimate state values using an approximation function. Within DRL there are numerous algorithms that differ mainly in their architecture to optimise the Actor policy. Examples are: Proximal Policy Optimization (PPO), Deep Deterministic Policy Gradient (DDPG), Soft-Actor Critic (SAC) or Twin Delayed DDPG (TD3).

2.1. Twin Delayed DDPG: TD3 algorithm

The TD3 algorithm is a model-free, online, off-policy reinforcement learning method, in which the agent follows the Actor-Critic methodology to achieve an optimal policy that maximizes the value of the expected cumulative long-term reward (MathWorks, 2021).

As the name indicates, TD3 is an extension of the DDPG algorithm. This approach assumes that, for environments with continuous action spaces, the Bellman optimisation function with which the policy is learned is differentiable with respect to the action argument. Therefore, a gradient-based learning rule is established for a policy μ(s) that exploits this fact (Spinning Up, 2021).

Since this estimation of the value function can imply an erroneous learning by the agent, the TD3 algorithm aims to solve this by means of three adjustments: (1) learning two Critic ("twin") functions to form the targets in the Bellman error loss functions, (2) updating the Actor parameters less frequently than Critic ("delayed") and (3) adding noise to the action chosen by Actor with the intention of preventing the policy from exploiting errors made in the value function estimates.

2.2. Deep Reinforcement Learning for process control

The control of industrial processes is a rather complex task, largely due to the non-linearity of the processes and the fact that in many cases there is more than one control loop. This has motivated the development of new control techniques that adapt more efficiently to the process in question (Robayo, F., et al., 2015).

The use of neural networks stands out among these techniques because they are non-linear models that can represent systems based only on the input and output data of the system, and they are highly adaptive by adjusting their parameters to changing operating conditions (Morcego, B., 2000).

This technique has been applied in cases such as the gasoline blending process (Yu, W., et al., 2004) where the use of recurrent neural networks is proposed to model the process, without the need to know the equations that define it, the system of interconnected tanks (Robayo, F., et al., 2015) where a neural controller based on an inverse model is developed or the case of a predictive controller based on neural networks for the control of the water level in a steam generator (Parlos, A., 2001).

The main objective pursued with the application of deep reinforcement learning in process control is to ensure that the value of a desired variable is the one established by its set point, while at the same time complying with the constraints of the process.

It is important to emphasize that the action space to be performed by the agent is continuous, rather than discrete, as in other areas where reinforcement learning has been successful.

In this framework, the agent represents the controller and the environment represents the process, so that the interaction between the agent and the environment is achieved through the actions (control actions) that the agent performs depending on the state of the environment it receives. In addition, the reward system is added to evaluate the quality of

the chosen action, according to the variation that the state of the environment has undergone (Spielberg, S.P.K., et al., 2017).

3. Case study

This section presents the continuous stirred-tank reactor (CSTR) unit that has been controlled using the TD3 algorithm.

In this reactor, the first order reaction (A → B) takes place isothermally, while it receives the input of two flows with different concentrations of reactant, and the product is obtained by gravity.

This work focuses on the control of two variables: the level of content inside the reactor and the concentration of the product. Moreover, it is intended to be able to work with different setpoints defined for each variable.

3.1. Implementation of the process

An agent has been designed to be responsible for performing two simultaneous actions on two other variables of the process, which have a direct impact on the variables to be controlled. These manipulated variables are the opening of the valve located in one of the inlet flows and the reactor temperature.

The environment, which as shown in Fig.1 is the other major player in the reinforcement learning loop, has been defined following the correct mass balance of the CSTR. The state of this environment, which contains the information that the agent receives from it, is defined by the variables that define each change of state: the instantaneous measurements of the variables to be controlled (level and product concentration), the absolute errors made in each control variable with respect to their setpoints, and the integrals and derivatives of these errors.

The most important hyperparameters of the reinforcement learning algorithm used by the agent in this work are shown in the following Table 1.

Table 1. Hyperparametros del agente para la unidad CSTR

Hyperparameter	**Value**	**Note**
Batch size n	64	
Replay memory size	10^6	Older transitions are replaced
Policy and Critic learning rate	$3 \cdot 10^{-4}$	Step size for ADAM
Policy hidden layers	2	1st: 400 neurons; 2nd: 200 neurons
Critic hidden layers	2	1st: 800 neurons; 2nd: 400 neurons
Hidden activation function	Relu	
Output activation function	Tanh	Only for policy
Loss function	MSE	
Target update rate (τ)	0.003	
Discount factor (γ)	0.99	
Exploration noise	0.2	Ornstein-Uhlenbeck process
Warm-up time	1,000	Timesteps until training starts

4. Simulation results

The resulting learning curve after application or the algorithm is shown in Fig.2 This curve represents the evolution of the reward values obtained throughout the episodes used to simulate the TD3 algorithm.

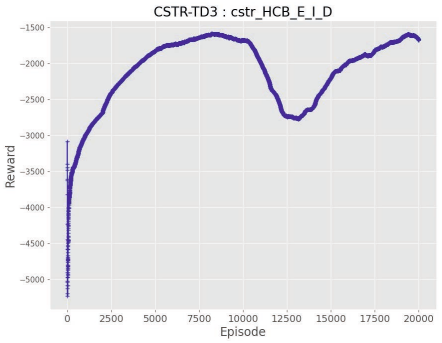


Fig. 2. Learning curve

The curve has been obtained for a simulation of 20,000 episodes. It starts with low reward values, as expected, and then grows progressively to high reward values as the episodes progress. It should be note that once a maximum reward value has been obtained, the learning curve declines from episode 10,000 onwards, which can be justified by a process of exploration by the agent seeking other possible solutions since the rewards already obtained do not continue to improve. This explanation is plausible because in episode 12,500, the curve is back on track and grows until it reaches the maximum values already obtained.

To complete the analysis of the learning curve, the TD3 algorithm is validated. This algorithm has been trained by achieving different set points for each control variable that are randomly generated episode after episode.

The Fig.3 show the response of the agent when the set point for each control variable is specified. In addition, the evolution of the reactant inside the CSTR is represented.

It can be seen that the evolution over time of the variables to be controlled is quite satisfactory, as the agent manages to achieve the objective of maintaining the control variables at their set points, when these are quite far apart.

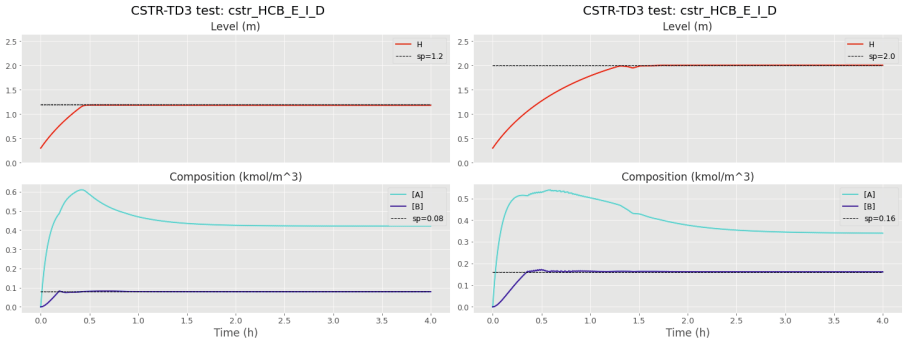


Fig. 3. Validation of the TD3 algorithm.

5. Conclusions and further work

The application of the TD3 algorithm to control the level and composition of a continuous stirred-tank reactor has generated favourable results, although the learning process is complex due to the difficulty of creating a multivariable and adaptive controller to be applicable to several set points. To solve this, the research will continue with the application of other deep reinforcement learning algorithms and configurations.

In addition, this work aims to give an advance with respect to PID controllers, since in this case the controller aims to cover a range of defined set points, without the need to be retuned. Furthermore, it has been shown that the reward function is important to obtain different results and it requires further work because a single function to achieve the control of two variables may be the cause of the learning problem, and it could be improved with the possibility of using two controllers (agents), so that each agent is in charge of controlling a single variable, with its corresponding reward function.

References

Bishop, C.M. , 2006. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag New York. Inc. Secaucus, NJ, USA.

MathWorks (2021) Retrieved from https://es.mathworks.com/help//reinforcement-learning/ug/create-agents-for-reinforcement-learning.html

Morcego, B. , 2000. Estudio de redes neuronales modulares para el modelado de sistemas dinámicos no lineales. Universitat Politècnica de Catalunya.

Parlos, A., Parthasarathy, S., & Atiya, A. (2001). Neuro-Predictive Process Control Using On-Line Controller Adaptation. IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY, VOL. 9, NO. 5, 741-755.

Robayo, F., Barrera, A., & Polanco, L. , 2015. Desarrollo de un controlador basado en redes neuronales para un sistema multivariable de nivel y caudal. *Revista Ingeniería y Región*, 14(2), 43-54.

Shin, J. , Badgwell, T. A. , Liu, K., Lee, J. H., 2019. Reinforcement Learning –Overview of recent progress and implications for process control. Computers and Chemical Engineering. 127, 282–294.

Spielberg, S.P.K. , Gopaluni, R.B. , Loewen, P.D., 2017. Deep Reinforcement Learning Approaches for Process Control. 6th International Symposium on Advanced Control of Industrial Processes 201 - 206.

Spielberg, S.P.K. , Tulsyan, A. ,Lawrence, N. P. , Loewen, P.D., Gopaluni, R.B. , 2019. Deep Reinforcement Learning for Process Control : A Primer for Beginners. AICHE Journal 2019.

Spinning Up (2021). Retrieved from https://spinningup.openai.com/en/latest/index.html

Sutton, R.S. , Barto, A.G. , 2018. Reinforcement Learning: An Introduction. MIT Press.

Yu, W., Moreno, M., & Gómez, E. (2004). Modelling of Gasoline Blending via Discrete-Time Neural Networks. IEEE International Joint Conference, (págs. 1291-1296 vol 2).

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50283-9

Application of machine learning and big data for smart energy management in manufacturing

Manu Suvarna[a], Pravin P.S[a], Ken Shaun Yap[b], Xiaonan Wang[a,c*]

[a] *Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore, 117585.*
[b] *Singapore Institute of Manufacturing Technology, 08-04, Innovis, Singapore, 138634.*
[c] *Department of Chemical Engineering, Tsinghua University, Beijing 100084, China*
chewxia@nus.edu.sg

Abstract

Within the *Industry 4.0* context, platforms such as cyber-physical production system (CPPS) offer numerous opportunities for smart energy management in manufacturing. In this study, we demonstrate the application of big data and machine learning (ML) to foster such practices for real manufacturing environments by taking the Model Factory (MF) in Singapore as a testbed. We first used supervised learning algorithms to predict machine-specific load profiles via energy disaggregation at the MF shop floor. Here, the light gradient boosting machines had the best predictive performance with a mean absolute error and root mean squared error of 0.035 and 0.106 (units in Watts). We then coupled unsupervised learning with mathematical optimization to devise an optimal energy scheduling plan for facility management at the MF. When applied for day-ahead scheduling, the data-driven optimizer showed cost benefits of 14% in comparison to the current existing conditions. The study successfully demonstrated the application of big data and ML in the drive towards smart manufacturing practices.

Keywords: smart manufacturing, energy disaggregation, light gradient boost, k-mean clustering, data-driven optimization

1. Introduction

The emergence of the fourth industrial revolution in recent years, commonly referred to as *Industry 4.0* has challenged and disrupted conventional manufacturing norms. Platforms such as cyber physical production systems (CPPS) and technologies such as Internet of Things (IoT), Artificial Intelligence (AI), digital twins etc. within the *Industry 4.0* framework, are transforming global manufacturing practices (Suvarna et al., 2021; Tao et al., 2018). The growing popularity of these technologies have sparked an interest on their potential application to reduce the energy consumption of manufacturing industries.

The energy consumed in a typical manufacturing setting could stem from the power required by machines in the shop floor, process equipment, office buildings and facility management. Some of the common strategies employed to minimize the energy consumption in manufacturing include lean management principles (six-sigma) for machine performances, optimization of production planning and scheduling, and eluding power peaks during prolonged production (Suvarna et al., 2020; Tan et al., 2021). While recent works have proposed conceptual means to apply CPPS and AI for the above mentioned, there is dearth in literature on application of these technologies to real-world case studies. To this aim, we show how data-driven analytics, can be applied to real-world manufacturing practices via two case studies, 1) machine learning based energy disaggregation of individual machines in a shop floor, and 2) day-ahead energy scheduling via data-driven optimization.

Methodology

1.1. Machine specific energy disaggregation

The Model Factory (MF) at the Singapore Institute of Manufacturing Technology (SIMTech) which is an actual production environment was used as a test bed for this study (Tan et al., 2021). Four machines from the MF, namely, laser welder (LW), laser trimmer (LT), oven 1 and oven 2 were selected. For each of the machines, the following information were logged: timestamp, individual electrical load profile (Watts) and operational states (1 = *off*, 2 = *production* and 3 = *idle*) at frequency of 1 minute for a duration of 15 months spanning from October 2017 to December 2018. This resulted in the extraction of 600,000 data points an ideal representation of industrial big data (IBD).

The primary objective was to disaggregate the central power supply to the machine-specific load-profiles. As such, the *total load* was defined as the input feature whereas the *individual load profiles* for each of the four machines were defined as the target labels (Tan et al., 2021). Tree-based supervised algorithms including extreme gradient boost (XGBoost), light gradient boosting machines (LightGBM), and deep learning algorithms including ensemble regular and bidirectional long short-term memory (EnLSTM) and (EnBLSTM) were evaluated for their predictive performance for the task.

During the modelling process, the entire data was first split into training and test data with temporal specifications. The 12-month period from Oct 2017 - Sept 2018 was labelled as training set while the last 3 months from Oct - Dec 2018 was labelled as the test set. The training data was subjected to hyperparameter tuning to find the optimum combination of hyperparameters for the various algorithms evaluated, using the Bayesian optimization strategy followed by k-fold cross validation (where k = 3). Once the best hyperparameters were identified on the training set, they were also used on the test data for each of the algorithms. The best performing algorithm was identified based on the terms of lowest MAE and RMSE scores on the test dataset. These are calculated as described in equations (1) and (2).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}(y_{act,i} - y_{pred,i}) \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_{act,i} - y_{pred,i})^2}{n}} \quad (2)$$

where, $y_{act,i}$ and $y_{pred,i}$ are the actual and predicted values of the target variables and *n* is the total number of data points.

1.2. Day-ahead energy scheduling via data-driven optimization

Currently the MF sources all its energy requirements for the operations (shop floor and facility management) from the central power grid. To this cause, we proposed the implementation of a *hypothetical hybrid grid* – comprising of solar panels and waste to energy (WTE) plant, in addition to the central power grid. For this case study, we focused only on the energy consumption of the refrigeration system, which is part of the technical building services in the MF. The energy consumed by the refrigeration system E_{com}, subjected to uncertainty in ambient temperature was modelled as:

$$E_{com} = \dot{m}_{com}(h_{com,disc} - h_{com,suct})/\eta_{com,mec} \quad (3)$$

where, $\dot{m}_{com}$ is the mass flow rate of the working fluid (kg/s), $h_{com,disc}$ is the specific enthalpy of the discharge fluid (J/kg), $h_{com,suct}$ is the specific enthalpy of the suction fluid (J/kg), $\eta_{com,mec}$ is the mechanical efficiency of the compressor (-).
The energy consumption of the refrigeration system is simulated with the operating conditions presented in **Table 1** as per (ASHRAE, 2020). The T_{amb} is obtained from the scenario generation method (discussed below).

Table 1. Relevant operational data for refrigeration system

Parameters	**Values**
Working fluid	R123
Degree of superheat	27.8°C
Degree of subcooling	0°C
Saturated temperature of evaporator	7.2°C
Mechanical efficiency of the compressor	85%
Cooling capacity	40 kW
Compressor speed	50 Hz

The hybrid grid proposed in this study is subjected to uncertainty on both the supply and demand side. At the demand side, it has to meet the energy requirements by the refrigeration system which is significantly influenced by ambient temperature. On the supply side, it is subjected to the solar power availability (which is a function of the solar irradiance) and constantly changing price of the main power grid. Thus, the operation of the hybrid grid (comprising of solar panels, WTE and central grid) was formulated as a stochastic optimization for day-ahead energy scheduling; with the objective to minimize the total operating cost given as:

$$\min \sum_{t=1}^{NT} (C_S^t \, E_S^{t,s} + C_{WTE}^t \, E_{WTE}^t + C_M^{t,s} \, E_M^t \,) \tag{4}$$

The optimization problem is devised to ensure that the energy delivered by the hybrid grid should satisfy the energy requirements to the air refrigeration system at all times. This constraint is formulated as follows:

$$ED^{t,s} \leq \sum_e^E E_S^{t,s} + E_{WTE}^t + E_M^t \tag{5}$$

The formulation also ensures that the amount of energy delivered by each energy sources should be always between the lower and upper bounds of the designed capacity of the energy sources. This constraint is formulated as follows:

$$E_{S,min}^t \leq E_S^{t,s} \leq E_{S,max}^t \tag{6}$$

$$E_{WTE,min}^t \leq E_{WTE}^t \leq E_{WTE,max}^t \tag{7}$$

$$E_{M,min}^t \leq E_M^t \leq E_{M,max}^t \tag{8}$$

where, *NT* is the total number of time slots in the horizon, *t* is the time period, *s* denotes stochastic parameters, $e \in E$ denotes energy sources, C_S^t and $E_S^{t,s}$ is the cost associated and energy availability from solar at time period *t,* C_{WTE}^t and E_{WTE}^t is the cost associated and energy availability WTE at time period *t*, $C_M^{t,s}$ and E_M^t is the cost associated and energy availability from power grid at time period *t,* $ED^{t,s}$ is the energy demand by the refrigeration system at time period *t*, $E_{e,min}^t$ is the lower bound/capacity of the e^{th} energy source and $E_{e,max}^t$ is the upper bound/capacity of the e^{th} energy source.

The decision variables are the energy contribution from solar ($E_S^{t,s}$), WTE (E_{WTE}^t) and the power grid (E_M^t) respectively, while all the other variables in the problem formulation described from equations (1)-(8) are the given parameters.

Conventional mathematical optimization approaches assume that the uncertainty set/scenarios are perfectly given *a priori*, which is then modelled via probability distribution function (PDF) or user defined uncertainty via sample average approximations or Monte Carlo simulation (Ning and You, 2019). In contrast, we use a data-driven approach in the form of unsupervised learning to create the scenarios. The application of unsupervised learning on big data for variable of interest (on which the scenarios are created), presents benefits such as reduced scenarios with greater confidence, thereby

resulting in enhanced optimization performance and faster computation (Ning and You, 2019; Tao et al., 2018). Thus, in this study, the hourly historical data (2015-2018) of solar radiance, ambient temperature and electricity pricing was sourced from relevant weather and electricity board in Singapore. The details of this historical data in the form of upper and lower bounds is presented in **Table 2**.

Table 2. The upper and lower bounds for relevant parameters in the stochastic optimization which determines the range uncertainty values

Parameters	Bounds	Solar	WTE	Mains
Price ($/kWh)	Lower	0.00	0.05	0.07
	Upper	0.07	0.11	0.11
Hourly energy availability (kWh)	Lower	0.00	0.00	0.00
	Upper	3.64	3.98	6.81

This data was then subjected to K-means clustering to create the scenarios. K-means partitions an N-dimensional population into k sets on a basis of a sample (Mehar et al., 2013). It is an unsupervised classification method to solve problems when no labels are available. In the dataset $\boldsymbol{D} = \{x_t\}_{t=1}^{n}$, the number of clusters is K, the natural goal is to seek a partition of the dataset $D_1 \cup \ldots \cup D_K$, as well an associated set of cluster centroids $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)$, such that the sum of Euclidian distances between features and each centroid is as small as possible. Initial $\boldsymbol{\mu}$ are randomly selected. While finding the global minimum of distance, there are two related steps: (1). Minimize the distance with respect to regulating the partitioned dataset D_K for fixed centroid μ_K; (2). Minimize the distance with respect to regulating the centroid μ_K for fixed partitioned dataset D_K. The K-means algorithm is optimized by alternating between these two steps until converges. Eventually, all points in the dataset locate around one of the K cluster centroids that achieve stable state. When centroids do not change, clusters are fixed, which means K scenarios are produced. Each probability of corresponding scene is the percentage of the number of labels in this cluster among the total count of labels in the dataset

The stochastic optimization was solved using the CPLEX solver. Its performance was compared with a base case analysis i.e., the actual energy consumed by the refrigeration system on 5th of September 2019. The base case is the actual working condition of the refrigeration system, and it uses only the central power grid as the current available energy source.

2. Results

2.1. Energy Disaggregation

In the energy disaggregation studies, it was observed that the tree-based algorithms (LightGBM and XGB) had lower MAE while the deep learning algorithms (EnLSTM and EnBLSTM) had lower RMSE to the tree-based counterparts. However, the performance of all the models were very comparable and the prediction accuracies were in close proximities to each other. Given the ease in tuning tree-based models, and the combined model balance in terms of MAE and RMSE, it was revealed that performance of order was LightGBM > EnBLSTM > XGBoost. Here it is worth a mention that, in the case of structured datasets as used in this study, tree-based models have very comparable performances to that of deep learning algorithms as they can fit the hyperparameters to the input features, which is a natural extension to their workflow. Moreover, both LightGBM and XGB are ensemble tree models and as such sum the predictions of many decision trees into a final one and thus closely compete to deep learning algorithms where the latter also use a multitude of neurons for their prediction performance. The average predictive performance of all the algorithms for the 4 machines is shown in **Table 3**. The actual (Act) v/s predicted (Pred) plot of energy disaggregation for LW and LT is show in **Figure 1**.

Table 3. Comparative evaluation of algorithms for energy disaggregation study

Ranking	Algorithm	MAE (Watts)	Algorithm	RMSE (Watts)
1	LightGBM	0.035	EnBLSTM	0.100
2	XGBoost	0.036	EnLSTM	0.103
3	EnBLSTM	0.039	LightGBM	0.106
4	EnLSTM	0.041	XGBoost	0.106

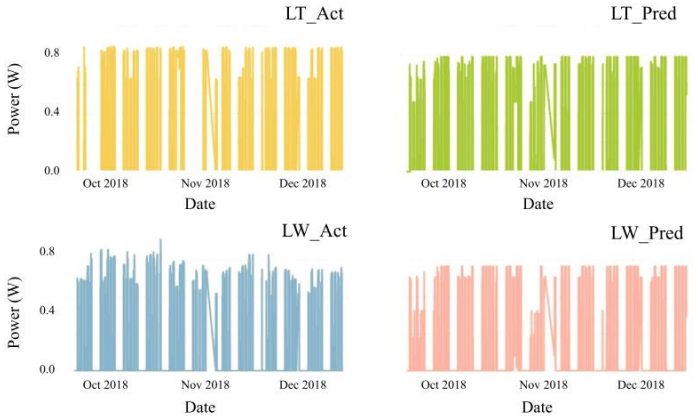


Figure 1. Actual v/s predicted plots of energy disaggregation for LW and LT for the best performing LightGBM model. Sourced with permission from (Tan et al., 2021)

2.2. Data-driven stochastic optimization for day-ahead energy scheduling

The energy demand for day ahead scheduling under uncertainty was first compared with the base case, where it was seen that the MAE and RMSE of energy demands between the two was 0.25 and 0.27 kWh respectively. With an average energy demand throughout the day between 6.5 and 6.8 kWh, the energy demand projected by the day-ahead scheduling deviated from the base case by approximately 4%.

On subjecting the day-ahead energy demand to stochastic optimization, it was identified that that the hybrid grid suitably met the demand by optimally distributing the three energy sources with the objective to minimize daily operating cost. Specifically, it was observed that the solar panel contributed significantly to the hybrid grid during the morning and afternoon hours, i.e., almost 100% of its available energy to meet energy demand thereby minimizing the overall cost. On the other end, the early morning and the late-night hours were purely met by the optimal combination of WTE, and power mains based on their hourly cost distribution. In addition to the distribution of the individual energy sources in the hybrid grid, its performance was also compared to the base case condition for cost saving evaluation. Here it was realized that the actual energy consumed by the refrigeration system using power mains only for the entire day operation was 14.18 SGD. In contrast, the overall cost incurred by the hybrid grid inclusive of all the uncertainties was found to be 12.41 SGD, concluding that adopting a hybrid grid could potentially lead to cost savings by 14% with day-ahead energy scheduling even under uncertainty. The comparative results of the base case and stochastic optimized model (hybrid grid) along with potential cost savings are shown in **Figure 2**. The data-driven optimization presented in this study

essentially captures significant uncertainty in the data, as it is trained on big data (hourly interval data for 3 years) – which also make the optimization more reliable.
One observed aspect in the creation of the uncertainty set via the k-means clustering algorithm was\s the variation of the uncertainty set on repeated implementation. This is due to the randomized component in k-means clustering. Although the results obtained from the k-means clustering varied in every iteration, the expected value of the uncertainty sets over every iteration remained the same. Hence, the final energy schedule resulting from the optimization problem did get affected by the random initialization in k-means clustering and gave consistently similar distribution mix and cost savings.

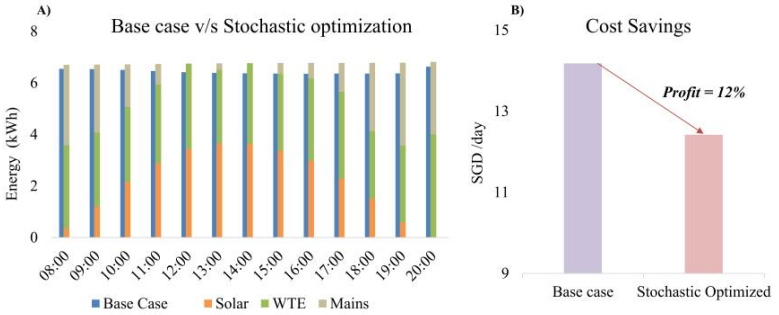


Figure 2. A) Comparative evaluation of the base case energy consumed with respect to hybrid grid B) 14% savings in terms of cost is observed with data-driven optimization under uncertainty

3. Conclusion

In this study, we first used various supervised algorithms to disaggregate the central load of the MF to four individual machines in the shop floor, where the LightGBM was found to be the most accurate in terms of the predictive as well generalization ability – as it was trained on a big data. In another case study, we showcased the application of unsupervised learning for creating scenarios essential to solve stochastic optimization. The approach is effective as was observed from the fact that this data-driven optimization resulted in cost savings of 14% for day-ahead energy scheduling for facility management. Both the approaches are purely data-driven and cross deployable in any manufacturing setting provided is availability of historical data in the plant.

References

ASHRAE, 2020. 2020 ASHRAE Handbook -- HVAC Systems and Equipment. ASHRAE.

Mehar, A.M., Matawie, K., Maeder, A., 2013. Determining an optimal value of K in K-means clustering, in: 2013 IEEE International Conference on Bioinformatics and Biomedicine.

Ning, C., You, F., 2019. Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming. Computers & Chemical Engineering 125, 434–448.

Suvarna, M., Büth, L., Hejny, J., Mennenga, M., Li, J., Ng, Y.T., Herrmann, C., Wang, X., 2020. Smart Manufacturing for Smart Cities—Overview, Insights, and Future Directions. Advanced Intelligent Systems 2, 2000043.

Suvarna, M., Yap, K.S., Yang, W., Li, J., Ng, Y.T., Wang, X., 2021. Cyber–Physical Production Systems for Data-Driven, Decentralized, and Secure Manufacturing—A Perspective. Engineering.

Tan, D., Suvarna, M., Shee Tan, Y., Li, J., Wang, X., 2021. A three-step machine learning framework for energy profiling, activity state prediction and production estimation in smart process manufacturing. Applied Energy 291, 116808.

Tao, F., Qi, Q., Liu, A., Kusiak, A., 2018. Data-driven smart manufacturing. Journal of Manufacturing Systems, Special Issue on Smart Manufacturing 48, 157–169.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50284-0

Adaptive least-squares surrogate modeling for reaction systems

Robert E. Franzoi[a], Brenno C. Menezes[a*], Jeffrey D. Kelly[b], Christopher L. E. Swartz[c]

[a]*Division of Engineering Management and Decision Sciences, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar*
[b]*Industrial Algorithms Ltd., 15 St. Andrews Road, Toronto M1P 4C3, Canada*
[c]*Department of Chemical Engineering, McMaster University, 1280 Main St W, Hamilton ON L8S 4L7, Canada*
bmenezes@hbku.edu.qa

Abstract

Surrogate modeling has been increasingly used to predict the behavior of a given system as an alternative to complex formulations that often lead to time consuming solutions and convergence issues. Surrogates are addressed herein to replace complex formulations for reactor systems within optimization problems. An adaptive sampling algorithm explores the solution space by iteratively building surrogates. Latin Hypercube Sampling is used for the experimental design (input data), and a first principles reaction formulation calculates the output data. Then, discrete least-squares regression minimizes the deviation between the surrogate response and the function being approximated. An optimization problem based on a reaction system is formulated, in which complex first principles equations are successfully replaced by the surrogates. The results indicate highly accurate predictions and near optimal solutions. Therefore, the surrogates can replace the rigorous model without significant loss in the solution quality and objective function. This methodology can potentially provide several benefits and improvements for real-time applications and for integrated optimization environments, in which the use of complex or rigorous models is not suitable.

Keywords: Surrogate modeling, adaptive sampling, optimization, data-driven, machine learning.

1. Introduction

Commercial tools for rigorous simulation have been widely used to provide highly accurate solutions for industrial problems. However, they are typically not suitable for large-scale optimization applications due to the expensive computational burden and convergence issues arising from their rigorous high-fidelity formulation. An alternative to overcome these mathematical and computational challenges derived from a detailed and complex modeling is the use of surrogate models. Several benefits have been reported in the use of surrogates in multiple applications, and the interest in developing reduced-size formulations for industrial applications has recently increased. Surrogates are built using data generated from the original model or any reliable source, and several aspects are important when designing a surrogate building strategy. First, there is a trade-off between model accuracy and computational tractability. In general, surrogates should be as accurate as possible given the availability of time and effort for

their application (Mencarelli et al., 2020). Second, their functional form should be selected considering the problem characteristics, availability of data, and dimensionality (Hüllen et al., 2019). Third, the design of experiments is chosen to generate the samples needed to train the surrogates (Simpson, 2001). Fourth, a performance method is selected to measure the fit from the surrogate to the data set (Alizadeh et al., 2020).

Previous literature on the topic has shown significant benefits in the use of surrogates in optimization problems that involve complex models and the use of rigorous simulation tools (Yang et al., 2016; Franzoi et al., 2020; Franzoi et al., 2021). In this work, we are particularly interested in building simple surrogate models for replacing complex first principles equations from reaction systems. The surrogate functions are required to have high accuracy to achieve high-quality predictions and to lead to near optimum solutions, while providing simplified formulations that are faster to solve and easier to converge.

The contribution of this work relies on the implementation of an adaptive sampling framework that iteratively builds accurate and small in size surrogates for reaction systems. Each iteration consists of bounds tightening, sampling selection, surrogate building, system optimization with surrogates, and system simulation with rigorous model. Latin Hypercube Sampling (LHS) (McKay, 1979) is used for experimental design with data generated from a rigorous formulation. The surrogates are multivariable second-order polynomial functions and include constant, linear, and bilinear terms. Discrete least squares regression minimizes the deviations between the surrogate response and the function being approximated. The proposed methodology can introduce useful applications in further embedding the simplified (yet accurate) reaction formulation into further optimization decision-making environments (e.g., scheduling, control), which can be extended to other processes and applications as well.

This paper is structured as follows. In Section 2 we present the problem statement and the mathematical formulation. In Section 3 we introduce the surrogate model building framework. The case study and the respective results are presented and discussed in Section 4. Finally, we highlight the main findings of this work in Section 5.

2. Problem statement

The case study chosen to illustrate the surrogate model methodology is proposed by (Williams and Otto, 1960). The Williams-Otto plant is considered at steady state operation and it is illustrated in Figure 1. The problem is formulated using the unit-operation-port-state-superstructure (UOPSS) formulation (Kelly, 2005), in which the mass balance consistency is ensured throughout the process.

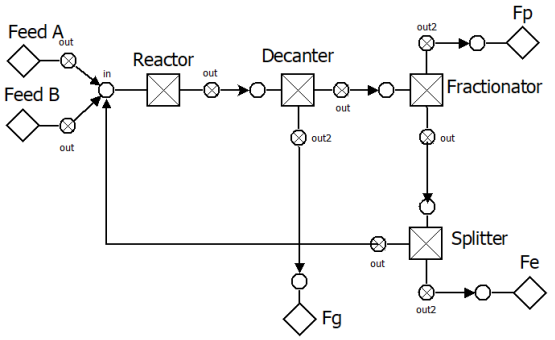


Figure 1: Williams-Otto plant flowsheet.

Although small in size, this problem is highly nonlinear due to the complex reaction system and the reflux rate that typically imposes convergence issues in the optimization. Hence, the system is very sensitive, and inaccurate formulations often lead to infeasibilities, which provides a good case study for testing and tuning surrogate model building strategies.

Two feed streams and a recycle stream enter the continuous stirred tank reactor, in which there are three exothermic reactions, with an Arrhenius temperature dependence. Eqs.(1) to (6) calculate the outlet flows of each individual component.

$$F_A^{reactor,out} = F_A^{reactor,in} - W_r f_1 e^{-\left(\frac{E_1}{T_R}\right)} Xr_A^{reactor} Xr_B^{reactor} \tag{1}$$

$$F_B^{reactor,out} = F_B^{reactor,in} - W_r f_1 e^{-\left(\frac{E_1}{T_R}\right)} Xr_A^{reactor} Xr_B^{reactor} + W_r f_2 e^{-\left(\frac{E_2}{T_R}\right)} Xr_B^{reactor} Xr_C^{reactor} \tag{2}$$

$$\begin{aligned} F_C^{reactor,out} = F_C^{reactor,in} &+ 2W_r f_1 e^{-\left(\frac{E_1}{T_R}\right)} Xr_A^{reactor} Xr_B^{reactor} - 2W_r f_2 e^{-\left(\frac{E_2}{T_R}\right)} Xr_B^{reactor} Xr_C^{reactor} \\ &+ W_r f_3 e^{-\left(\frac{E_3}{T_R}\right)} Xr_C^{reactor} Xr_P^{reactor} \end{aligned} \tag{3}$$

$$F_E^{reactor,out} = F_E^{reactor,in} + 2W_r f_2 e^{-\left(\frac{E_2}{T_R}\right)} Xr_B^{reactor} Xr_C^{reactor} \tag{4}$$

$$F_G^{reactor,out} = F_G^{reactor,in} + 1.5W_r f_3 e^{-\left(\frac{E_3}{T_R}\right)} Xr_C^{reactor} Xr_P^{reactor} \tag{5}$$

$$F_P^{reactor,out} = F_P^{reactor,in} + W_r f_2 e^{-\left(\frac{E_2}{T_R}\right)} Xr_P^{reactor} Xr_E^{reactor} - 0.5W_r f_3 e^{-\left(\frac{E_3}{T_R}\right)} Xr_C^{reactor} Xr_P^{reactor} \tag{6}$$

In the above, W_r is the reactor total mass content, f_1, f_2, f_3 are the frequency factors, E_1, E_2, E_3 are the activation energies, and $Xr^{reactor}$ are the mass fractions of each component inside the reactor. The reaction optimization problem maximizes the profit from revenue with products P and E by subtracting feedstock costs and treatment cost of by-product G in Eq.(7). Additional details of the problem and its mathematical formulation can be found in previous works (Chaudhary et al., 2009).

$$\text{Max } Z = 8400\,(0.6614\,F_p + 0.0150\,F_e - 0.0441\,F_a - 0.0661\,F_b - 0.0220\,F_g) \tag{7}$$

A mixed-integer quadratic programming (MIQP) technique determines optimizable surrogates to correlate variations of independent X variables to dependent Y variables. Seven input variables (reactor temperature, flow, and five inlet compositions) and six output variables (reactor outlet compositions) are used. We choose the form of each equation to account for linear, bilinear, and quadratic coefficients, as shown in Eq.(8), in which DV and IV are the sets for the dependent and independent variables, respectively.

$$Y_i = a_i + \sum_{j \in IV} b_{ij} X_j + \sum_{j \in IV} \sum_{k \in IV} c_{ijk} X_j X_k \qquad \begin{matrix} \forall\ i \in DV, \\ \forall\ j \le k \end{matrix} \tag{8}$$

The coefficients a_i, b_{ij} and c_{ijk} are the parameters to be estimated within the surrogate model building strategy that minimizes the least squares error in Eq.(9), where y_{ip} and Y_{ip} are the real and the estimated value for the dependent variable i at point p.

$$\text{Minimize} \sum_{p=1}^{n} (y_{ip} - Y_{ip})^2 \tag{9}$$

The model is subject to Eqs.(10) to (13), that limit the values of the coefficients in the surrogates, where M is a large enough number, z_0, z_j, and z_{jk} are auxiliary binary variables, and K is the maximum number of coefficients.

$$-Mz_0 \le a_i \le Mz_0 \qquad \forall\ i \in DV \tag{10}$$

$$-Mz_j \le b_{ij} \le Mz_j \qquad \forall\ j \in IV, i \in DV \tag{11}$$

$$-Mz_{jk} \le c_{ijk} \le Mz_{jk} \qquad \forall\ \{(j,k) \in IV, j \le k\}, i \in DV \tag{12}$$

$$z_0 + \sum_{j \in IV} z_j + \sum_{j \in IV, k \in IV, j \le k} z_{jk} \le K \qquad z_0, z_j, z_{jk} \in \{0,1\} \tag{13}$$

3. Surrogate model building: an adaptive sampling algorithm

Three decisions should be made upon designing the surrogate model building strategy, which concern the surrogate functional form, data generation, and quality of fit. To keep the model simple, we choose to build the surrogate as second-order polynomial functions, which have been often used due to their robust performance and computational efficiency (Yang et al., 2016). The LHS technique randomly samples points for the input variables (building the independent X data set); then, a first principles model calculates the output variables for each point to build the dependent (Y) data set. Lastly, least squares regression is employed within an MIQP formulation using Eqs.(9) to (13) to find the optimal coefficients that form the surrogate.

Building global surrogates for the entire solution space may introduce difficulties due to the wide search space in the optimization, which potentially leads to convergence issues. Therefore, an adaptive sampling framework is used to build locally accurate surrogates within their respective trust regions in an iterative fashion. Our algorithm selects the sampling bounds for each independent variable, and one surrogate is built for each dependent variable at each iteration, until a convergence criterion is met.

The algorithm is implemented in Python 3 using Microsoft Visual Studio 2019 in the industrial modeling and programming language (IMPL) platform. The optimizations are carried out by CPLEX 12.10.0, and the non-linearities are handled by a sequential linear programming strategy. The machine used was an Intel Core i7 with 2.90 GHz and 16 GB RAM. The following steps explain how the algorithm works.

1) Generate data set: Data are generated by sampling points for the input variables using LHS and evaluating each point using the rigorous model to calculate the output variables.

2) Update sampling survey: At each iteration, the sampling survey is updated around the incumbent optimal values of each variable. For that, a parameter δ is introduced, so that the sampling survey of an independent variable Xr is defined as $(Xr^{opt} - \delta) \le Xr \le (Xr^{opt} + \delta)$. We initially set $\delta = 0.10$, which is updated upon some shrinkage criteria.

3) Build surrogate model: The surrogate model comprises a group of six surrogate functions, one for each output variable. These functions are obtained through MIQP optimizations to identify their optimal coefficients.

4) Solve optimization problem using the Surrogate Model (SM): The surrogates are embedded in the reaction optimization problem, which is solved to local optimality.

5) Simulate system using the First Principles (FP) model: After optimizing the problem using the incumbent surrogate model, the optimal values of the decision variables (F_A, F_B, T_R) are fixed, and the system is simulated using the rigorous reaction model.

6) Convergence criteria: The algorithm terminates if there is no improvement in three sequential objective functions (with 0.01% tolerance) or at a maximum of 20 iterations.

7) Final surrogate model: The best surrogate is selected based on the objective function.

4. Example

The proposed case study is the Williams-Otto plant shown in Figure 1. Our framework iteratively builds surrogates for the reactor unit and solves the reaction optimization problem using the surrogates. The results are presented in Figure 2 (left plot). The line with circular markers represents the SM objective function, the line with square markers represents the FP objective function, and the straight line is the best objective function found by optimizing the model using the rigorous blending formulation. As the framework moves across iterations, there are improvements in the optimization search space chosen to build the surrogates, and a smooth convergence is achieved, reaching a high-quality solution within 11 iterations (beyond which there is no significant improvement in the surrogates). The best objective function found for this instance is \$1683.81, while the best surrogate model leads to an objective function of \$1683.71.

Although achieving highly accurate surrogates, reducing the size of the model would be recommended for larger problems to keep its simplicity. Thus, additional tests were performed considering different numbers of terms. For that, we run the framework multiple times by setting the maximum number of terms to be $K = \{8, 12, 18, 24, 30, 36\}$. Figure 2 (right plot) shows the objective function of the surrogates built within the framework for each scenario.

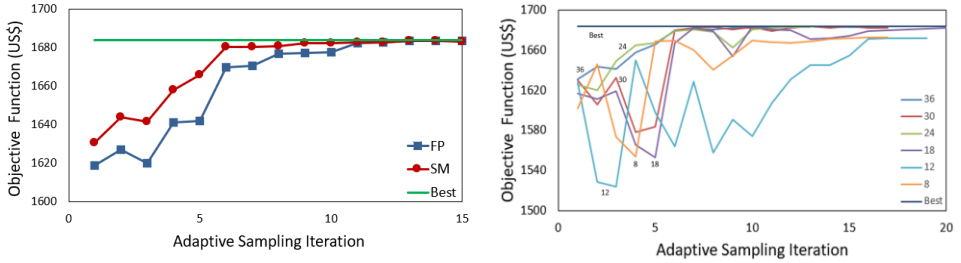


Figure 2: Adaptive sampling algorithm results.

All scenarios performed well, achieved high-quality predictions, and provided good solutions. The best performance is achieved by using 36 coefficients (0.01% lower than the best solution known), while the least accurate performance was achieved by using $K = 8$ coefficients, which resulted in a gap of 0.75%. Using less than 8 coefficients does not provide sufficiently accurate results. These results demonstrate that although a large number of terms can provide higher accuracy, sufficiently accurate surrogates of smaller size can also be built, which provides benefits in terms of reduced size, lower computational effort, and better convergence within simulation/optimization problems. In most instances, the non-improvement termination criterion is met prior to reaching the maximum number of surrogates. This provided an average reduction of over 20 % in the computational time with no loss in the objective function.

The methodology employed herein has shown to be efficient in terms of accuracy when building surrogates for highly nonlinear and complex systems, such as the Williams-Otto plant. The nonlinear Arrhenius-based equations were efficiently approximated by second-order polynomial functions, and no convergence issues are detected when embedding the surrogates in the original reaction system. The adaptive sampling algorithm successfully explores the optimization search space to find more promising regions to build the surrogates within a few iterations of the method. High-quality solutions and smooth convergence are achieved in the case study tested.

5. Conclusions

Surrogate modeling has been increasingly used to predict the behavior of a system as an alternative to complex formulations that often lead to time consuming solutions and to both convergence and calibration difficulties. In this work, an adaptive sampling algorithm iteratively explores the solution space, whereby the incumbent surrogate is embedded into an optimization problem to assure feasibility and to collect feedback for the following iteration. The methodology is applied to a reaction system network and the surrogates are built to predict the reactor outputs within optimization environments. The results indicate that the surrogates are properly built, have high accuracy, and can effectively replace the first principles model in the optimization without significant loss in the objective function. The effectiveness of the method is also demonstrated in building smaller surrogates by limiting the maximum number of coefficients, which also provides high-quality predictions. We believe this methodology is appropriate for other reaction systems and can be useful for handling data-driven black-box nonlinear formulations. Moreover, several benefits and improvements can be achieved for time-limited applications and for integrated optimization environments, in which the use of complex or rigorous models might not be suitable.

References

R. Alizadeh, J. K. Allen, and F. Mistree, 2020, Managing computational complexity using surrogate models: a critical review, Research in Engineering Design, 31(3), 275-298.

M. N. R. Chaudhary, 2009, Real time optimization of chemical processes, Doctoral dissertation, Curtin University.

R. E. Franzoi, T. Ali, A. Al-Hammadi, B. C. Menezes, 2021, Surrogate Modeling Approach for Nonlinear Blending Processes, 1st International Conference on Emerging Smart Technologies and Applications (eSmarTA), 1-8.

R. E. Franzoi , B. C. Menezes, J. D. Kelly, J. A. W. Gut, I. E. Grossmann, 2020. Cutpoint temperature surrogate modeling for distillation yields and properties, Industrial and Engineering Chemistry Research, 59 (41), 18616–18628 .

G. Hüllen, J. Zhai, S. H. Kim, A. Sinha, M. J. Realff, and F. Boukouvala, 2019, Managing Uncertainty in Data-Driven Simulation-Based Optimization, Computers and Chemical Engineering, 136, 106519.

J. D. Kelly, 2005, The Unit-Operation-Stock Superstructure (UOSS) and the Quantity-Logic-Quality Paradigm (QLQP) for Production Scheduling in the Process Industries, Multidisciplinary International Scheduling Conference Proceedings: New York, United States, 327.

J. D. Kelly, B. C. Menezes, 2019, Industrial Modeling and Programming Language (IMPL) for Off- and On-Line Optimization and Estimation Applications. In: Fathi M., Khakifirooz M., Pardalos P. (eds) Optimization in Large Scale Problems, Springer Optimization and Its Applications, 152, 75-96.

M. D. McKay, R. J. Beckman, and W. J. Conover, 1979, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics, 21, 239-245.

L. Mencarelli, A. Pagot, and P. Duchêne, 2020, Surrogate-based modeling techniques with application to catalytic reforming and isomerization processes, Computers and Chemical Engineering, 135, 106772.

T. W. Simpson, D. K. Lin, and W. Chen, 2001, Sampling strategies for computer experiments: design and analysis, International Journal of Reliability and Applications, 2(3), 209-240.

T. J. Williams, and R. E. Otto, 1960, A generalized chemical processing model for the investigation of computer control, Transactions of the American Institute of Electrical Engineers, Part I: Communication and Electronics, 79(5), 458-473.

L. Yang, S. Liu, S. Tsoka, and L. G. Papageorgiou, 2016, Mathematical programming for piecewise linear regression analysis, Expert systems with applications, 44, 156-167.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50285-2

Machine Learning and Inverse Optimization Approach for Model Identification of Scheduling Problems in Chemical Batch Plants

Hidetoshi Togo[a], Kohei Asanuma[b], Tatsushi Nishi[a*]

[a]*Graduate School of Natural Science and Technology, Okayama University, 3-1-1 Tsushima-naka, Kita-ku, Okayama City, Okayama 700-8530, Japan*
[b]*Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama-cho, Toyonaka City, Osaka 560-8531, Japan*

nishi.tatsushi@okayama-u.ac.jp

Abstract

Scheduling problems are widely used in recent production systems. In order to create an appropriate modeling of a production scheduling problem more effectively, it is necessary to build a mathematical modeling technique that automatically generates an appropriate schedule instead of an actual human operator. This paper addresses two types of model estimation methods for weighting factors in the multi-objective scheduling problems from input-output data. The one is a machine learning-based method, and the other one is the parameter estimation method based on an inverse optimization. These methods are applied to multi-objectives parallel machine scheduling problems. The accuracy of the proposed machine learning and inverse optimization methods is evaluated. A surrogate model that learns input-output data is proposed to reduce the computational efforts. Computational results show the effectiveness of the proposed method for weighting factors in the objective function from the optimal solutions.

Keywords: Inverse Optimization, Machine Learning, Multi-Objective Optimization, Production Scheduling, Weighting Factors, Model Identification

1. Introduction

Scheduling problems are widely used in chemical batch plants in current production systems. In recent years, the real-world scheduling problem is so large and it becomes so complicated. Therefore, it is required to aid the decision-makers to model the problem that enables the efficiency and flexibility of production systems. Data-driven optimization methods have expected to build appropriate optimization model from historical data. For multi-objective scheduling problem, a mathematical model that reflects the human operator's decision making is required based on the selection of multi-criteria optimization. However, it is not easy to set the appropriate weighting factors that indicate the importance of each objective function. If the weighting factors do not reflect the worker's intention, the desired solution cannot be obtained. Therefore, human operators must manually fine-tune the schedule.

Some related works have been addressed for estimating weighting factors in the multi-objective scheduling problems (Watanabe et al. (2002), Kobayashi et al. (2018)). Mat-

suoka et al. (2019) developed a machine learning approach for the identification of the objective function for parallel machine scheduling problems. Asanuma and Nishi (2020) addressed machine learning and inverse optimization approach for estimating weighting factors from historical data (Asanuma and Nishi (2020)). The exact solution algorithm is adopted to solve the inverse optimization problem. Togo et al. (2021) reported an approximate solution approach for the inverse optimization problem. However, the applicability of the practical scheduling data has not been studied in conventional works. Even if the desired schedule is obtained, the situation surrounding production environment changes. Therefore, it is necessary to repeatedly correct the weighting factors for multi-objective scheduling problem. This paper presents an inverse-optimization approach for model identification of production scheduling problem using historical data. The proposed approach is applied to real data of chemical batch plants. Various approximate solutions have been proposed for solving scheduling problems and they have advantages and disadvantages (Kise et al. (1995)). If the approximate solution method is used, the solutions are sometimes different from those obtained by the exact solution method. In this paper, we propose machine learning and inverse optimization methods to estimate weighting factors in the objective function. In order to apply to large scale problems, a simulated annealing method is proposed to derive near-optimal solutions for the multi-objective parallel machine scheduling problems. In the machine learning method, we try to extract features to improve the estimation accuracy. From the result of numerical experiments, it is confirmed that the estimation accuracy is improved by adding the feature of the errors in due date setting for each machine in the machine learning for the parallel machine scheduling problem. This paper consists of the following sections. Section 2. explains the problem definition. Section 3. introduces our proposed approach for estimating weighting factors in the objective function. Section 4. provides computational experiments. Section 6. states our conclusion and future works.

2. Problem description

2.1. Estimation problem for weighting factors in the objective function

We consider a parallel machine scheduling problem in a chemical batch plant. The plant consists of several batch units and several tanks. The plant configuration, the demand and recipe information are given. The plant has a number of daily schedules that are created by human operators. The problem treated in this study is to estimate of the objective functions and weighting factors for the scheduling problem by using the input and output data. Given a set of problem instance data and the solutions of the scheduling problem under the condition that the weighting factors of the scheduling problem are unknown, the problem is to estimate appropriate weighting factors of the objective function.

2.2. Production scheduling problem

We consider a parallel machine scheduling problem. This scheduling problem is the determination of the allocation of jobs to multiple machines and the processing order of jobs under the condition that each job is processed by a single machine. The following constraints are considered.

Constraints
1. Each machine has no idle time.
2. One machine can only handle one job at a time.
3. Each job cannot be interrupted or divided once the processing has started.
Two types of objective functions (the sum of delivery delays f_1 and the sum of setup costs f_2) are considered in this study.

3. Proposed approach

The outline of the proposed approach is shown in Fig. 1. In this approach, historical data is used to estimate the weighting factors of the objective function. The details are explained in the following sections.

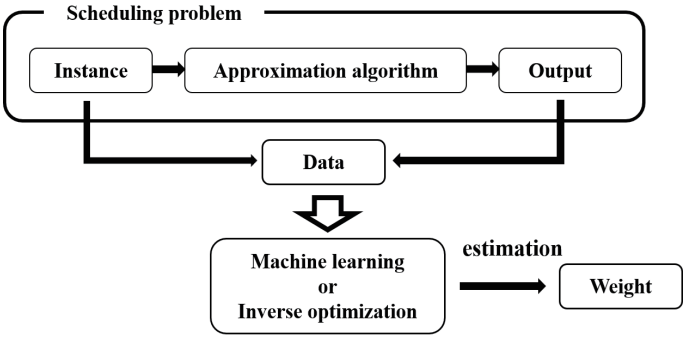


Figure 1: Outline of the proposed method

3.1. Machine Learning Algorithm

We use a supervised learning method of machine learning. First, we prepare a large number of problem examples of scheduling problems and a large set of outputs are obtained by solving them exactly via an approximate solution algorithm such as simulated annealing or genetic algorithm. In the case study, the weight, delivery date, processing time, label, and setup cost of each job are given as the parameters. In addition, the output includes the starting time of each operation and the type of machine that performs the processing. The weighting factor of each objective function actually used is taken as the correct answer. Preprocessing is applied to the prepared problem examples and output, and the feature is extracted. The square error between the estimated value of the obtained weighting coefficient and the weighting factor of the correct answer is obtained. The machine learning model trained from the process is evaluated. A recurrent random forest is used in the machine learning method. Random forest is learning using a large number of decision trees that improve generalization performance by using the representative values of these output results as the overall output (Breiman et al. (2001)).

In order to develop an accurate prediction model in a random forest, it is necessary to extract effective features. It is difficult to improve the prediction accuracy by using only

the input / output data. On the contrary, if unnecessary features or meaningless features are included, noise may occur during learning, which may significantly increase the learning time or lower the prediction accuracy. From here, we describe the data generation method for use as features.

We consider the following features in the random forest.
Feature 1 The value of the objective function
Feature 2 Rank correlation coefficient
Feature 3 The error of delivery date setting for each machine
Feature 4 The sum of the processing completion times of each machine

3.2. Inverse Optimization Approach

The inverse optimization approach is to solve the problem that, given the result of an optimization problem, determines the input to the optimization problem that generates the correct result. To solve the problem, the forward optimization problem is repeatedly solved for the given input parameters. It can be treated as a problem of determining the input parameter that minimizes the loss function of the output. In order to reduce the computation time, a surrogate model that can represent forward optimization model has been used. In this study, we use a neural network that is often used as a surrogate model. During the learning process, the data such as each parameter and the weighting factor of the problem instance, and each objective function value is used as an output. The accuracy and the computation time performance of the replaced neural network is the forward optimization part are evaluated in the computational experiments.

4. Computational experiments

4.1. Randomly generated instances

We consider a parallel machine scheduling problem with 5 machines and 50 jobs. 100 problem examples are generated by setting parameters at random. Schedules are generated by changing the weighting factors of the objective function by using the approximate solution method. Random forest is used for machine learning. Four types of features (Feature 1 only, Features 1 and 2, Features 1 and 3, Features 1 and 4) are utilized. Table 1 shows the mean square error of the proposed machine learning method and the inverse optimization method. The computational results show the effectiveness of using Features 1 and 3. The results of the inverse optimization are more effective than those of the machine learning method.

Table 1: MSE of the machine learning method and inverse optimization method

Feature 1 only	Features 1, 2	Features 1, 3	Features 1, 4	Inverse optimization
4.78e-3	4.75e-3	4.20e-3	4.33e-3	6.66e-4

5. Application to petrochemical batch plant

The proposed approach is applied to a lubricant oil production plant. The lubricant process consists of proportion mixing and additive oil mixing process, filling process into several batch tanks from feed oil. The feed oil is obtained by hydrogenation and solvent desulfurization for low, medium and high viscosity oils which are extracted from distilation columns from feed crude oils. The batch plant treated in this study is the filling process. There are six parallel filling units in the batch plant. The scheduling problem for filling process of the lubricant oil production plant is regarded as a parallel machine scheduling problem with several practical constraints.

5.1. Case study

For the input data, the amount of oil to be filled and the filling speed for each tank are given. A week of job data is utilized. For the input data, production schedules are generated by changing the weighting factor by 0.1 for each objective function. For the output data, the starting time and ending time of each filling operation, allocation to the tank, and each objective function value are obtained. The weighting factors of each objective function are estimated from these input / output data. Two types of objective functions (the sum of delivery delays f_1 and the sum of setup costs f_2) are considered in the problem for filling process of the lubricant oil production plant. In machine learning, the value of the objective function (Feature 1) and the value of the objective function and the sum of the processing completion times of each machine (Features 1, 4) are used in the random forest.

In the inverse optimization, the weighting factors are updated so that the gradient between the result obtained with the given weighting factor and the correct result becomes smaller. The algorithm finishes when it is repeated a certain number of times. Table 2 shows the estimation results when the weighting factor of f_1 and f_2 are (0.2, 0.8), (0.4, 0.6), (0.9, 0.1). The estimation result of (0.9, 0.1) of inverse optimization was far from the correct answer. In the actual schedule data, the sum of setup costs increased significantly when the weight of f_2 changed from 0.2 to 0.1 and the accuracy of the proposed method is not good. For randomly generated instances, the changes in the objective function value were little. Due to the sudden increase in the objective function value, the weights updates were increased. Then the MSE of the proposed method is not better in those situations.

Table 2: Weighting factors estimation results for machine learning (ML) and inverse optimization

	ML using Feature 1	ML using Features 1, 4	Inverse optimization
(0.200, 0.800)	(0.252, 0.748)	(0.222, 0.778)	(0.266, 0.734)
(0.400, 0.600)	(0.402, 0.598)	(0.408, 0.592)	(0.370, 0.630)
(0.900, 0.100)	(0.842, 0.159)	(0.847, 0.153)	(0.991, 0.009)

Table 3 shows the mean square error (MSE) of the proposed machine learning method and inverse optimization method. The results show the effectiveness of using feature the sum of the processing completion times of each machine. Comparing with the three methods, the

machine learning provides better results. The inverse optimization is more susceptible to the range of change in the objective function value than machine learning. It is considered that the cause of the actual schedule is that the range of change of the objective function value is different from that of the randomly generated schedule.

Table 3: MSE of the machine learning (ML) and inverse optimization method

	ML using Feature 1	ML using Features 1,4	Inverse optimization
MSE	5.96e-3	4.90e-3	2.61e-2

6. Conclusion and future works

We have studied the application of machine learning and inverse optimization method for estimating weighting factors from thepractical petrochemical scheduling problem. The effectiveness of our proposed approach to real chemical batch plants has been confirmed from the computational results.

References

K. R. Baker, J.C.Smith, A Multiple-Criterion Model for Machine Scheduling. *Journal of Scheduling* **6**(1): 7–16.

S. Watanabe, T. Hiroyasu and M. Miki (2002). Neighborhood Cultivation Genetic Algorithm for Multi-objective Optimization Problems. *Journal of Information Processing* **43**(10): 183–198.

H. Kobayashi, R. Tachi and H. Tamaki (2018). Introduction of Weighting Factor Setting Technique for Multi-Objective Optimization. *Journal of the Institute of Systems, Control and Information Engineers* **31**(8): 281–294.

K. Asanuma and T. Nishi (2020). Estimation of weights of multi-objective production scheduling problems - an inverse optimization approach -, *Proceedings of 2020 International Conference on Industrial Engineering and Engineering Management*.

H. Togo, K. Asanuma, T. Nishi, Estimating weighting factors using approximate solutions of multi-objective scheduling problems, *Proceedings of International Symposium on Scheduling 2021*: 30–33.

H. Kise (1995). Simulated Annealing Method for Scheduling Problems. *Operations Research* **5**(8): 268–273.

Y. Matsuoka, T. Nishi, K. Tierney, Machine learning approach for identification of objective function in production scheduling problems, Proceedings of 2019 IEEE International Conference on Automation Science and Engineering, 679-684 (2019)

L. Breiman (2001). Random Forest. *Kluwer Academic Publishers* (45): 5–32.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50286-4

Decision-Focused Surrogate Modeling with Feasibility Guarantee

Rishabh Gupta[a], Qi Zhang[a*]

[a]*Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN 55455, USA*

qizh@umn.edu

Abstract

Surrogate models are commonly used to reduce the computational complexity of solving difficult optimization problems. In this work, we consider decision-focused surrogate modeling, which focuses on minimizing decision error, which we define as the difference between the optimal solutions to the original model and those obtained from solving the surrogate optimization model. We extend our previously developed inverse optimization framework to include a mechanism that ensures feasibility (or minimizes potential infeasibility) over a given input space. The proposed method gives rise to a robust optimization problem that we solve using a tailored cutting-plane algorithm. In our computational case study, we demonstrate that the proposed approach can correctly identify sources of infeasibility and efficiently update the surrogate model to eliminate the found infeasibility.

Keywords: surrogate modeling, learning for optimization, inverse optimization, feasibility guarantee.

1. Introduction

A common strategy for solving difficult optimization problems, especially in real-time applications, is to develop surrogate models of reduced computational complexity. In particular, data-driven surrogate modeling methods have become very popular with the opportunity to leverage recent advances in machine learning. Here, one uses the original model to generate data, which are used to fit the surrogate model that can then be embedded in the optimization problem. A key challenge in surrogate modeling is the balance between model accuracy and computational efficiency. As a result, much of the research effort in this area has focused on developing surrogate models that have simple functional forms or specific structures such that the optimization problems are easier to solve using the surrogate models [Cozad et al., 2014, Zhang et al., 2016].

The vast majority of existing surrogate modeling methods construct models that are given as systems of equations, which represent all or part of the equality constraints of the original optimization model [Bhosekar and Ierapetritou, 2018]. The goal of these surrogate modeling algorithms is to minimize the prediction error with respect to the original systems of equations. However, as we found in our recent (not yet published) work, a low prediction error in this kind of surrogate models does not necessarily lead to a low *decision error*, which is defined as the difference between the optimal solutions of the original and the surrogate optimization models. Yet arguably, decision accuracy is what the user

primarily cares about once the optimization model is deployed as a decision-making tool. We developed a data-driven inverse optimization approach to construct surrogate models that take the form of simpler optimization models and directly minimize the decision error; hence, we refer to it as *decision-focused surrogate modeling*.

Decision-focused surrogate modeling focuses on the set of optimal solutions rather than the larger set of feasible solutions. As such, it is prone to generating surrogate models that violate constraints in the original model. In this work, we address this issue by extending our inverse optimization framework to construct surrogate models with feasibility guarantees. We propose a robust optimization approach where we treat the set of possible inputs as an uncertainty set, and we develop a tailored cutting-plane algorithm to solve the resulting extended inverse optimization problem. Results from our computational case study show that using the proposed approach, we can construct surrogate optimization models with feasibility guarantees without substantial sacrifice of decision accuracy.

2. Mathematical Formulation

We consider an original optimization problem of the following general form:

$$\begin{aligned} \underset{x \in R^n}{\text{minimize}} \quad & f(x,u) \\ \text{subject to} \quad & g(x,u) \leq 0, \end{aligned} \tag{1}$$

which is a, possibly nonconvex, nonlinear program (NLP). Here, x and u denote the decision variables and model input parameters, respectively. We assume that solving problem (1) requires more time than what is allowed in our desired online application; however, we can solve it offline to generate data in the form of (u_i, x_i)-pairs, where x_i is the optimal solution to problem (1) given the input u_i.

Given a set of data points $\mathcal{I}$, the goal is to generate a surrogate optimization model that is easier to solve but achieves the same or almost the same optimal solutions as the original model. We postulate a surrogate optimization model of the following form:

$$\begin{aligned} \underset{x \in R^n}{\text{minimize}} \quad & \hat{f}(x,u;\theta) \\ \text{subject to} \quad & \hat{g}(x,u;\omega) \leq 0, \end{aligned} \tag{2}$$

where $\hat{f}$ and $\hat{g}$ are parameterized by θ and ω, respectively, and are constructed to be convex in x, which renders problem (2) a convex NLP.

The decision-focused surrogate modeling problems attempts to directly learn an optimization model from data that are assumed to be optimal solutions to this model. As such, it gives rise to a data-driven inverse optimization problem (IOP) [Gupta and Zhang, 2021], which can be formulated as follows:

$$\underset{\theta \in \Theta,\, \omega \in \Omega,\, \hat{x}}{\text{minimize}} \quad \sum_{i \in \mathcal{I}} \|x_i - \hat{x}_i\| \tag{3a}$$

$$\text{subject to} \quad \hat{x}_i \in \underset{\tilde{x} \in R^n}{\arg\min} \left\{ \hat{f}(\tilde{x}, u_i; \theta) : \hat{g}(\tilde{x}, u_i; \omega) \leq 0 \right\} \quad \forall\, i \in \mathcal{I}, \tag{3b}$$

where $\hat{x}_i$ denotes the solution predicted by the surrogate model. The objective is to determine the surrogate model parameters θ and ω that minimize the decision error defined in (3a) as the difference between the optimal solution to the original problem x_i and $\hat{x}_i$ across the given data set. Constraints (3b) state that for each $i \in \mathcal{I}$, $\hat{x}_i$ is the optimal solution to the surrogate optimization model with input u_i.

One potential issue with the IOP formulation (3) is that a predicted solution $\hat{x}_i$ is not guaranteed to be feasible in the original model (1). In addition, assuming that the input u can be chosen from a set $\mathcal{U}$, the optimal solution to the surrogate model is not guaranteed to be feasible in (1) for all $u \in \mathcal{U}$ even if $\hat{x}_i$ is feasible in (1) for all $i \in \mathcal{I}$. Hence, to ensure feasibility, we add the following constraints to problem (3):

$$\left.\begin{array}{c} \bar{x} \in \underset{\tilde{x} \in R^n}{\arg\min} \left\{ \hat{f}(\tilde{x}, u; \theta) : \hat{g}(\tilde{x}, u; \omega) \leq 0 \right\} \\ g(\bar{x}, u) \leq 0 \end{array}\right\} \quad \forall\, u \in \mathcal{U}, \tag{4}$$

which state that given a surrogate model defined by θ and ω, the optimal solution to the surrogate model for any $u \in \mathcal{U}$, $\bar{x}$, also has to satisfy the original constraints $g(\bar{x}, u) \leq 0$.

3. Solution Strategy

The extended IOP is a bilevel semi-infinite program. To solve this problem, we propose a cutting-plane algorithm that iterates between a master problem and a cut-generating separation problem. The master problem is formulated as follows:

$$\begin{array}{lll} \underset{\theta \in \Theta,\, \omega \in \Omega,\, \hat{x},\, \bar{x}}{\text{minimize}} & \displaystyle\sum_{i \in \mathcal{I}} \|x_i - \hat{x}_i\| & \\ \text{subject to} & \hat{x}_i \in \underset{\tilde{x} \in R^n}{\arg\min} \left\{ \hat{f}(\tilde{x}, u_i; \theta) : \hat{g}(\tilde{x}, u_i; \omega) \leq 0 \right\} & \forall\, i \in \mathcal{I} \\ & \bar{x}_j \in \underset{\tilde{x} \in R^n}{\arg\min} \left\{ \hat{f}(\tilde{x}, u_j; \theta) : \hat{g}(\tilde{x}, u_j; \omega) \leq 0 \right\} & \forall\, j \in \mathcal{J} \\ & g(\bar{x}_j, u_j) \leq 0 & \forall\, j \in \mathcal{J}, \end{array} \tag{5}$$

which is a relaxation of the extended IOP since the semi-infinite constraints (4) have been replaced by a finite number of constraints defined over a discrete set $\mathcal{J}$. For each $j \in \mathcal{J}$, we have a specific input u_j and the corresponding predicted solution $\bar{x}_j$. If the optimal solution to (5) satisfies constraints (4), then it is also optimal for the extended IOP. Otherwise, we solve the following separation problem for each constraint function g_k to identify inputs for which the solutions of the surrogate model violate the original constraints:

$$\begin{array}{ll} \underset{u \in \mathcal{U},\, \bar{x}}{\text{maximize}} & g_k(\bar{x}, u) \\ \text{subject to} & \bar{x} \in \underset{\tilde{x} \in R^n}{\arg\min} \left\{ \hat{f}(\tilde{x}, u; \theta) : \hat{g}(\tilde{x}, u; \omega) \leq 0 \right\}. \end{array} \tag{6}$$

If the optimal value of (6) is greater than zero (or some defined feasibility threshold ϵ), we add the corresponding input u to the set $\mathcal{J}$ and re-solve problem (5). By doing so, we iterate between the master and the separation problems until no more constraint violations can be found, which indicates that we have solved the extended IOP.

Both problems (5) and (6) are bilevel optimization problems. To solve them, we first reformulate them into single-level problems by replacing the lower-level problems with their KKT conditions, which is possible since the surrogate optimization model is designed to be convex. The resulting formulations generally do not satisfy common regularity conditions, which makes their direct solution using standard NLP solvers difficult. Instead, we solve an exact penalty reformulation, which we do not describe here in detail due to space limitations. Note that while a local solution to problem (5) is usually enough to provide good results, problem (6) has to be solved to global optimality to guarantee feasibility.

4. Computational Case Study

In our case study, we consider the heat exchanger network shown in Figure 1, which is adopted from Biegler et al. [1997]. Here, the inlet temperature of stream H2, T_5, has a nominal value of 583 K but is subject to random disturbances. Whenever there is a change in T_5, we optimize the operation of the heat exchanger network by solving the following NLP in which we can adjust the cooling duty Q_c and the heat capacity flowrate F_{H2}:

$$\underset{Q_c,\, F_{H2}}{\text{minimize}} \quad 10^{-2}\, Q_c + 4\,(F_{H2} - 1.7)^2 \tag{7a}$$

$$\text{subject to} \quad 0.5\, Q_c + 165 \geq 0 \tag{7b}$$

$$-10 - Q_c + (T_5 - 558 + 0.5\, Q_c)\; F_{H2} \geq 0 \tag{7c}$$

$$-10 - Q_c + (T_5 - 393)\; F_{H2} \geq 0 \tag{7d}$$

$$-250 - Q_c + (T_5 - 313)\; F_{H2} \geq 0 \tag{7e}$$

$$-250 - Q_c + (T_5 - 323)\; F_{H2} \leq 0 \tag{7f}$$

$$Q_c \geq 0,\; F_{H2} \geq 0, \tag{7g}$$

which is nonconvex due to the bilinear term in constraint (7c).

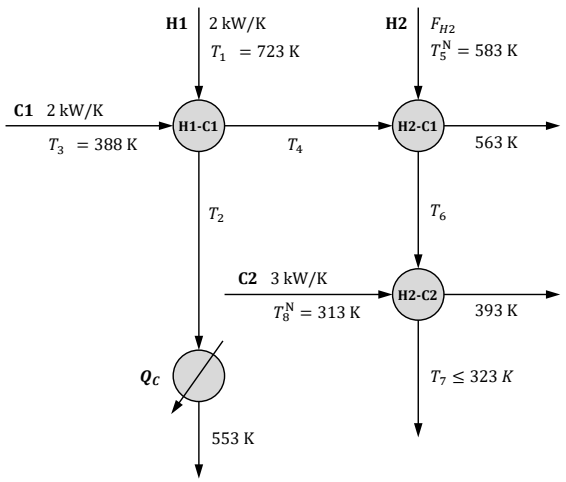


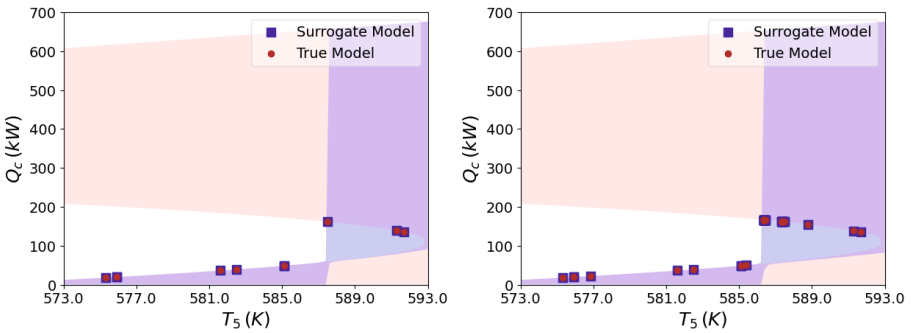
Figure 1: Given heat exchanger network.

We employ the proposed decision-focused surrogate modeling approach to replace the bilinear term $Q_c\, F_{H2}$ in constraint (7c) with the following approximation:

$$Q_c\, F_{H2} \rightarrow a(T_5)\, Q_c + b(T_5)\, F_{H2}, \tag{8}$$

where a and b are some functions of the input parameter T_5. This change, together with estimating the objective function $\hat{f}$ as a convex quadratic function and keeping all linear constraints, results in a surrogate convex QP for problem (7) that is much easier to solve.

We obtain the initial surrogate model by randomly sampling ten values of T_5 in the range [573 K, 593 K] and solving problem (3) with the corresponding global optimal solutions of (7). Here, we assume a and b in (8) to be cubic polynomials in T_5. The result is depicted in Figure 2a, which shows, for each chosen T_5, the true optimal Q_c and the Q_c obtained from solving the surrogate optimization model. In addition, it shows the sets of feasible Q_c for the original (red area) and surrogate (blue area) models. One can observe that while the feasible regions are quite different, there is very good agreement in the true and predicted optimal solutions, which can be attributed to the decision-focused nature of our approach.



(a) Iteration 1 (b) Iteration 25

Figure 2: Comparison between the original model and the surrogate optimization model.

Next, we solve the extended IOP to minimize the violation of constraint (7c) at the optimal solutions of the surrogate model. We perform 25 iterations of the proposed cutting-plane algorithm. Figure 3 shows the maximum constraint violation, which is the optimal value of problem (6) solved for constraint (7c), and the corresponding violated input temperature T_5 that is then added to set $\mathcal{J}$ in problem (5) at each iteration. One can see that as the algorithm progresses, violations across the entire input range are detected until from iteration 13 onward, the algorithm only detects constraint violation in the region around $T_5 = 586.3$ K. This can be explained by Figure 2b, which shows all training data points accumulated over the 25 iterations and the feasible regions of the true and surrogate models. We see that for $T_5 \geq 586.3$ K, part of the feasible region of the surrogate model is infeasible in the true model. While the surrogate model achieves a very good fit for almost all optimal solutions in this region, there seems to be always some point at $T_5 \approx 586.3$ K that is infeasible, which is where we see a "transition" in the feasible region of the surrogate model. This indicates that the proposed cubic approximation of constraint (7c) is not sufficient to achieve feasibility across the entire input range, resulting in the algorithm focusing

on minimizing infeasibility by repeatedly sampling the area around 586.3 K. However, our algorithm correctly identifies the main source of infeasibility. In this particular case, the result instructs a simple remedy of the problem, which is to create two surrogate models, one for $T_5 < 586.3$ K and one for $T_5 \geq 586.3$ K. Then, with the same training data points, solving the corresponding IOPs directly returns two surrogate optimization models whose optimal solutions are feasible for the entire input space.

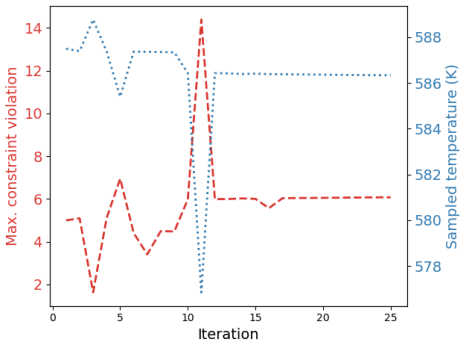


Figure 3: Progression of the cutting-plane algorithm.

5. Conclusions

In this work, we developed a decision-focused surrogate modeling approach that generates surrogate optimization models with feasibility guarantees. This is achieved by combining concepts from inverse optimization and robust optimization, and solving the resulting problem using a tailored cutting-plane algorithm. A computational case study considering a heat exchanger network example demonstrates the ability of the proposed approach to effectively identify and eliminate sources of infeasibility.

References

A. Bhosekar and M. Ierapetritou, 2018. Advances in surrogate based modeling, feasibility analysis, and optimization: A review. *Computers and Chemical Engineering*, 108:250–267.

L. T. Biegler, I. E. Grossmann, and A. W. Westerberg, 1997. Systematic methods for chemical process design.

A. Cozad, N. V. Sahinidis, and D. C. Miller, 2014. Learning Surrogate Models for Simulation-Based Optimization. *AIChE Journal*, 60(6):2211–2227.

R. Gupta and Q. Zhang, 2021. Decomposition and adaptive sampling for data-driven inverse linear optimization. *INFORMS Journal on Computing*, forthcoming.

Q. Zhang, I. E. Grossmann, A. Sundaramoorthy, and J. M. Pinto, 2016. Data-driven construction of Convex Region Surrogate models. *Optimization and Engineering*, 17(2): 289–332.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50287-6

Grade transition optimization by using gated recurrent unit neural network for styrene-acrylonitrile copolymer process

Shi-Chang Chang[a], Chun-Yung Chang[a], Hao-Yeh Lee[a]*, I-Lung Chien[b]

[a] *Dept Chemical Engineering, National Taiwan University of Science and Technology, No.43, Sec. 4, Keelung RD., Da'an Dist.,10607 Taipei City, Taiwan;*
[b] *Dept Chemical Engineering, National Taiwan University, No.1, Sec. 4, Roosevelt Road, Da'an Dist.,10617 Taipei City, Taiwan;*
haoyehlee@mail.ntust.edu.tw

Abstract

The melt index (MI) of polymer products is an important quality reference for the product properties. However, MI cannot be measured in real-time, and the current value of MI can only be obtained by laboratory analysis after several hours, which leads to unsatisfactory quality control results. To solve the problem, this paper adopts the styrene-acrylonitrile (SAN) copolymers process as a target process and uses the Gated Recurrent Unit (GRU) to establish a MI dynamic prediction model for different grades of SAN copolymer to estimate the current and future MI values, which ultimately improve the MI quality control performance. In addition, to solve the quality fluctuation caused by the difficulty of fine tune the chain modifier feed flow during the grade transition. Therefore, this paper also combines the GRU dynamic model and a virtual controller to provide recommended operating values for the chain modifier to reduce the transient time during grade transition. The simulation results in this paper show that the predicted value of MI is in agreement with the actual measured value. In addition, the recommended value of the chain modifier feed flow rate in comparison to actual manual control can significantly reduce about 28.6 hours of the grade transition time.

Keywords: GRU, soft sensor, melt index, control, grade transition.

1. Introduction

SAN resin is composed of 70~80 wt% styrene (SM) and 20~30 wt% acrylonitrile (AN). The property of polymer products is usually adjusted according to market demand. To meet the different final product of physical or chemical properties requirements, the MI of polymer needs to be adjusted, and each MI value corresponds to each grade. The transfer of polymer products from one grade to another is called grade transition. Because no sensor can measure the MI value, it is through low-frequency manual sampling and laboratory analysis. Therefore, it will take about 4 to 8 hours or even 1 day to have one MI measurement data. This large delay measurement not only makes MI more difficult to control but also requires additional manpower for quality analysis. Therefore, a model needs to be established to simulate the dynamic behavior of the process to estimate the accurate real-time MI value, also known as a soft sensor. However, due to the complexity of the copolymer polymerization reaction and the lack of a complete reaction kinetic formula, it is difficult to establish a first-principle model of the system, so this paper uses a data-driven GRU (Cho et al., 2014) model is used to estimate the dynamic behavior of

MI of different grades of SAN copolymers. In addition, various fine-tune policies are often based on the operator experiences to adjust the feed flow rate of the chain modifier during the grade transition. It will cause quality fluctuations. Therefore, this paper designs a virtual velocity form of PI controller through the GRU dynamic model and calculates the recommended value of chain modifier feed flow rate in real-time, which can ultimately reduce transient time, the waste of raw materials, the output of secondary products.

Neural networks have found widespread usage in modeling the complex and dynamic behavior of various polymer processes. Lee et al. (2009) used the first-principle EVA copolymerization reactor model to generate data, then established an artificial neural network (ANN) model to estimate the MI of the EVA process, and used PI controller for controlling melting index. Noor et al. (2010) reviewed various cases of ANNs used in polymer process simulation and emphasized the advantages of ANNs for fitting highly nonlinear systems. Jumari and Yusof (2017) used ANN and the first-principle model to simulate the MI of the polypropylene process, and compared the prediction performance of the two models. Compared with the past, this paper uses more novel deep learning modeling technology and GRU model to try to obtain a more accurate MI prediction value and combined with the virtual controller to provide the recommended value of the chain modifier feed flow rate. Finally, the model and virtual controller are applied to the SAN copolymer process.

2. Process description

The production process of SAN is shown in **Figure 1**. Fresh SM and AN monomers after being mixed with the recycled monomers in a mixing tank flow into the reactor for polymerization. An additional amount of chain modifier is also added into the feed stream before entering the reactor. The material inside the reactor is mixed at a constant mixing speed, while the reactor temperature is controlled through an external cooling utility. The reactor effluent is preheated and enters the devolatilizer from the top to separate unreacted monomers, solvent, and chain modifier. The molten resin is then passed through an extruder to obtain the final product, SAN. The gas phase material from the devolatilizer is first condensed, followed by the extraction of acrylonitrile by fresh styrene. The extracted acrylonitrile is then routed to the recycle tank. The sampling result of the recycle tank concentration is used to determine the amount of fresh monomer addition and the recycle flow to the mixing tank. After the initial adjustment to the flow of fresh monomers during grade transition, the chain modifier is mainly used for fine-tuning the MI to achieve new grade specifications. Low product MI means high molecular weight and low fluidity which necessitate increasing the chain modifier, and vice versa.

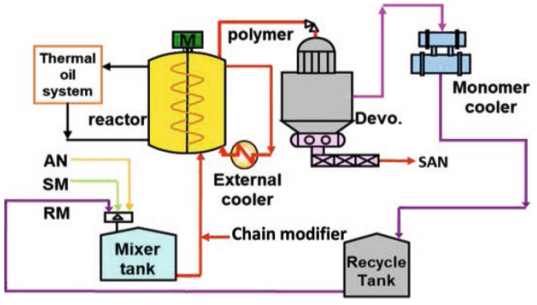


Figure 1. SAN process flowsheet

3. Data-driven GRU dynamic model

3.1. Data collection and pre-processing

All the data in this paper is obtained from the plant distributed control system (DCS) and the quality measurement results. The DCS data sampling time is 10 minutes, and the quality data (MI) is 4 to 8 hours. There is a total of 6 product grades in the data, corresponding to grade A~F, with each grade having its unique MI. Since each process variable has different units, the obtained data is first normalized between 0 and 1 to prevent the convergence problem of the neural network.

The data is divided into 8 data sets after excluding the abnormal data from a total of 48,541 DCS data sets and 2,105 MI measurement points. The training set accounts for 51%, the validation set accounts for 27%, and the testing set accounts for 22% of all data sets. The training set and the validation data set contain all grades due to the model training requirements. Due to the insufficiency of the data, only 4 types of grades that are frequently produced are covered in the testing data set.

3.2. Variable selection

This paper attempts to quantitatively analyze the importance of all the variables affecting MI. Due to the nonlinear characteristics of the system, this paper uses eXtreme Gradient Boosting (XGBoost) (Chen and Guestrin, 2016) for analysis, and combines chemical engineering background knowledge to select the input variables of the model. There are 15 DCS measurement variables in the SAN process, for which the importance analysis result is shown in **Figure 2**. The red bar represents the selected variable while the values on the y-axis signify the importance score. It is important to note that variables *f10* and *f13* which correspond to the total feed flow and reactor temperature, respectively, were ranked low despite their importance in estimating the MI**.** The reason is that the feed flow is dictated by the market demand, and in the case of fixed demand the feed flow usually remains constant. The temperature, on the other hand, is controlled by the temperature controller. Therefore, it is difficult for the XGBoost model to analyze the dynamic relationship between these two variables and MI.

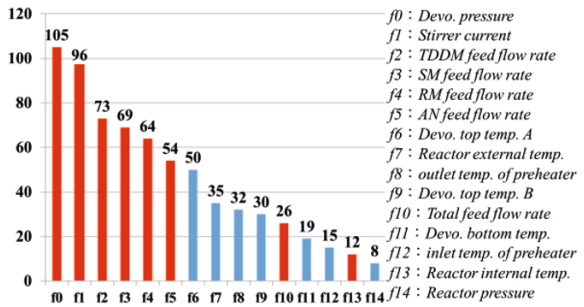


Figure 2. XGBoost variable importance analysis result

3.3. MI simulation method and results

This paper adopts Bayesian optimization (BO), a more efficient optimizer, as the hyperparameter tuning method. However, to prevent the model from overfitting, the

hidden layer of the model is fixed at 3 layers, and the early stopping method is added. Furthermore, the model training in this paper uses RAdam (Liu et al., 2019) combined with Lookahead (Zhang et al., 2019) optimizer, also known as Ranger optimizer. After selecting the hyperparameters, the iterative method based on Mini-Batch is used to optimize the parameters of the GRU model with the Ranger optimizer and the mean square error (MSE) as the objective function. An epoch represents one iterative loop over the entire training dataset. After completing an epoch training, the model is verified and it is decided whether to stop the training process (Early Stop) according to the validation result. The simulation results of each grade of the training set are found to be in good agreement to the actual MI having Mean Absolute Percent Error (MAPE) of each data set within 5%; As illustrated in **Figure 3**, the MAPE of the test set is 3.629%.

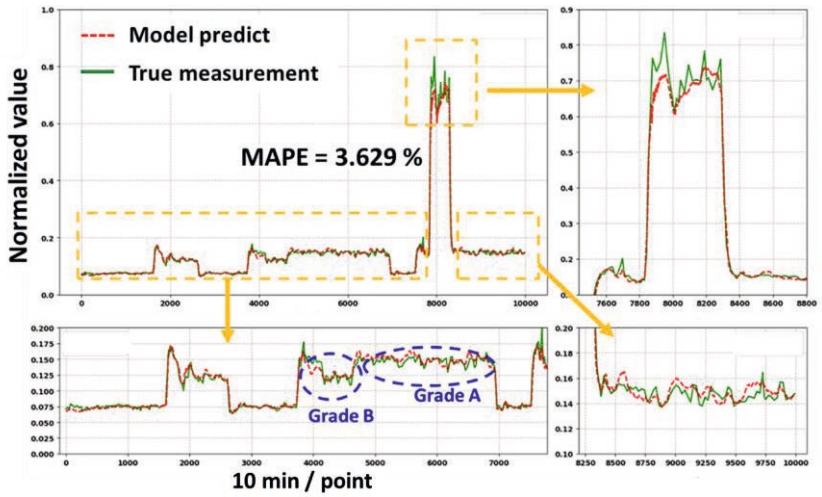


Figure 3. MI simulation result of the test data set

3.4. MI real-time prediction

To avoid data storage and calculations from increasing with time, this paper uses a rolling algorithm for online prediction. Since the GRU model in this paper considers time delay, it can predict the output response in the future. As a demonstration in **Figure 4**, it combines the concept of GRU input time step and time delay so that it can predict not only the current output y(t) but also the future output y(t+1) and y(t+2).

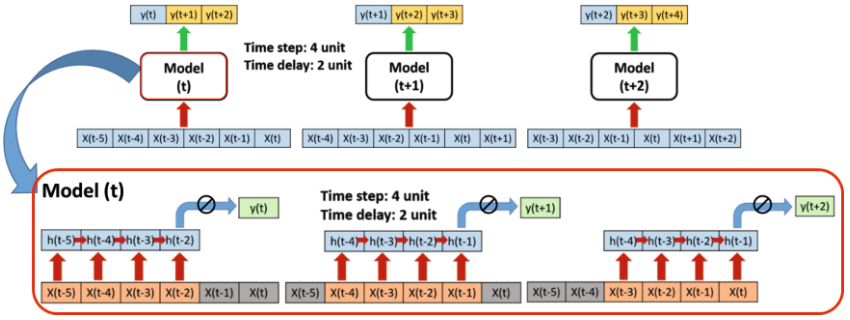


Figure 4. Model online rolling prediction

4. Improvement of grade transition

4.1. Controller design

In this paper, the Relay auto-tuning method (Åström and Hägglund, 1984) is used to determine the PI controller parameters, and the controller parameters are tuned according to a different grade. When the Relay auto-tuning method is used for tuning, the controller is temporarily replaced with a relay, and the chain modifier feed flow is passed through the relay to convert its feedback value into an up-and-down oscillation**.** Other input variables remain at the steady-state values corresponding to the current tuning grade. The dynamic parameters between chain modifier flow and MI obtained as a result of the relay feedback test are used to calculate the ultimate gain and period. Finally, the parameters of the controller are calculated through the Ziegler-Nichols tuning relation (Ziegler and Nichols, 1942).

The controller parameter of each grade does not apply for the situation of grade transition. Therefore, during the grade transition, the MI setpoint is set to the target MI of the next grade, while the virtual controller parameters are taken to be the average of current and next grades. If the current MI prediction value is already within the control range of the next grade, the controller parameters are set to the values of the next grade.

4.2. Grade transition result

Figure 5 is the result of the grade A to grade B transition when the virtual controller and the MI prediction model form a closed loop. The transition from grade A to grade B is a case of shifting from high MI to low MI. Therefore, the controller reduces the chain modifier flow to a minimum at the beginning of the transition, causing the MI to drop rapidly. The first yellow arrow in **Figure 5** represents the time point when the MI simulation result reaches the product set point, while the second yellow arrow represents the time point when the measured MI reaches the set point. It is observed that the transition time reduces by 28.6 hours for the MI simulation when compared to actual data. Therefore, after adopting the recommended value of chain modifier flow, the MI simulation result during grade transition is found to be smoother, reaching the set value faster. In addition, transitions between the other products have similar results.

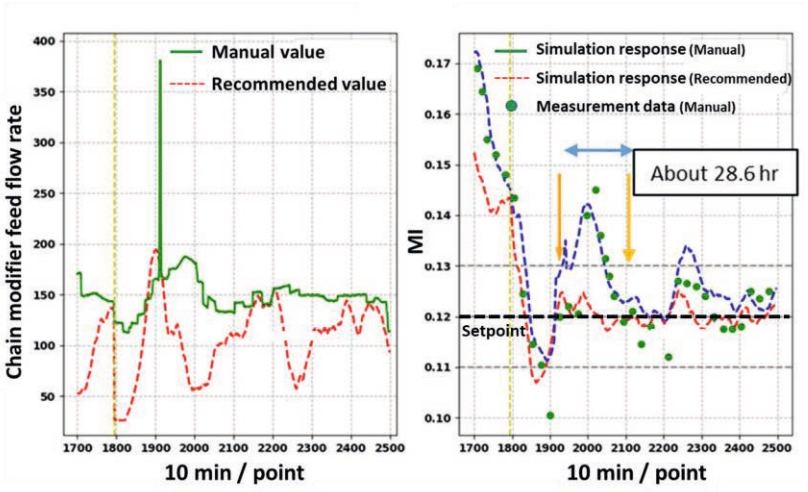


Figure 5. Chain modifier feed flow rate and MI simulation result (grade A to B)

5. Conclusions

In this paper, a GRU dynamic prediction model for the SAN production process is established to estimate the current and future MI values and combine with virtual PI controller provide the recommended value of chain modifier feed flow rate and consequently reduce the transient time during the grade transition.

The GRU model is used to build the MI prediction model having MAPE of each training and validation dataset within 5%. The MAPE of the test dataset is 3.629%, which represents good simulation results. The model also shows an accuracy of about 95% when tested online. The actual MI values are also found to be in good agreement with the predicted values from this model.

The model is then used to build a virtual controller that can suggest the values for chain modifier flow rate. The offline simulation results show that the MI simulation results are closer to the set value, and the product transitional time can be shortened by up to 28.6 hours. The results of this research can reduce substandard products in the process of product transition in polymer plants to reduce costs. Moreover, this research has already been implemented in an actual plant, and indeed shorten the time for grade transition.

References

K. J. Åström, T. Hägglund, 1984, Automatic tuning of simple regulators with specifications on phase and amplitude margins, Automatica, *20(5)*, 645-651.

T. Chen, C. Guestrin, 2016, Xgboost: A scalable tree boosting system, In proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 785-794.

K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, 2014, Learning phrase representations using RNN encoder-decoder for statistical machine translation, arXiv preprint arXiv:1406.1078.

N. F. Jumari, K. M. Yusof, 2017, Comparison of melt flow index of propylene polymerisation in loop reactors using first principles and artificial neural network models, Chemical Engineering Transactions, *56*, 163-168.

H. Y. Lee, T. H. Yang, I. L. Chien, H. P. Huang, 2009, Grade transition using dynamic neural networks for an industrial high-pressure ethylene–vinyl acetate (EVA) copolymerization process, Computers & Chemical Engineering, *33*, 1371-1378.

L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao, J. Han, 2019, On the variance of the adaptive learning rate and beyond, arXiv preprint arXiv:1908.03265.

R. M. Noor, Z. Ahmad, M. M. Don, M. H. Uzir, 2010, Modelling and control of different types of polymerization processes using neural networks technique: A review, The Canadian journal of chemical engineering, *88(6)*, 1065-1084.

M. R. Zhang, J. Lucas, G. Hinton, J. Ba, 2019, Lookahead optimizer: k steps forward, 1 step back, arXiv preprint arXiv:1907.08610.

J. G. Ziegler, N. B. Nichols, 1942, Optimum settings for automatic controllers, Trans. ASME, 64, 759–768.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50288-8

Development of Estimating Algorithm for Biodegradation of Chemicals Using Clustering and Learning Algorithm

Kazuhiro Takeda[a*], and Kazuhide Kimbara[a]

[a] *Department of Engineering, Shizuoka University, Shizuoka 432-8561, Japan*
**takeda.kazuhiro@shizuoka.ac.jp*

Abstract

Chemical substances should be assessed for biodegradation in environment. The biodegradation tests usually require 28 days of continuous testing and expensive costs. To reduce the cost, some software has been proposed to estimate the biodegradation of chemicals. Although, the software has not enough performance. Therefore, we develop a new estimation algorithm for the biodegradation of chemicals using clustering and machine learning algorithms. The combination of the Birch clustering algorithm and the XGboost learning algorithm is proposed to estimate the biodegradation of chemicals. Using 4200 real tested chemicals, the proposed algorithm was examined. The Birch algorithm might collect chemicals that are clear relationship between the explanatory variables and biodegradation.

Keywords: Biodegradation test; Chemical descriptor; Machine learning; Clustering.

1. Introduction

Chemicals should be assessed for biodegradation in environment. The collection of ready biodegradability tests is defined by the OECD (Organization for Economic Co-operation and Development) (OECD, 1992). The OECD 301C (MITI (Ministry of international Trade and Industry, Japan) I) test, one of the biodegradation tests defined by the OECD, usually requires 28 days of continuous testing and expensive costs. To reduce the cost, some software has been proposed to estimate the biodegradation of chemicals (Tunkel, 2000; Jaworska, 2002). Many tools for estimating the biodegradation of chemicals have been reviewed (Singh, 2021). The CATALOGIC (Laboratory of Mathematical Chemistry, 2021) is one of the most well-known software for estimating the biodegradation of chemicals, with 94% correct answer rate for not readily biodegradable products and 77% correct answer rate for easily decomposable products (NITE, 2020). Therefore, we develop a new estimation algorithm for the biodegradation of chemicals using clustering and machine learning algorithms.

2. Methods

Since biodegradation tests are affected by various factors such as bacteria and enzymes, it is a very difficult task to estimate the biodegradation of chemical substances. This paper proposes a new estimation algorithm for the biodegradation of chemicals using clustering and machine learning algorithms. This section briefly describes the different types of clustering and machine learning algorithms.

2.1. Clustering algorithm

There are several algorithms for clustering which automatically generate clusters from unlabelled data. The clustering algorithms such as K-Means (Sculley, 2010), Meanshift (Comaniciu, 2002), Ward (Ward, 1963), and OPTICS (Ester, 1996) use distances between points to make clusters. The K-Means algorithm clusters the data by trying to separate the samples into n groups of homoscedasticity, minimizing a criterion known as inertia or sum of squares within the cluster. The mini-batch K-Means algorithm is a variant of the K-Means algorithm. The Meanshift algorithm aims to find blobs with a smooth density of samples. The Ward algorithm is one of the hierarchical clustering algorithms. The Ward algorithm minimizes the sum of the squares of the differences in all clusters. The OPTICS algorithm creates a reachability graph to determine cluster membership. The DBSCAN algorithm (Schubert, 2017) is similar with the OPTICS algorithm but uses distances between nearest points. The Affinity propagation (Kettani, 2014) and the Spectral clustering algorithm (David, 2017) use graph distance. The Affinity propagation algorithm creates a cluster by sending a message between a pair of samples until it converges. The Spectral algorithm performs low-dimensional embedding of affinity matrices between samples. The Agglomerative clustering algorithm (Kettani, 2014) and Birch algorithm (Zhang, 1996) use any pairwise distance. The Agglomerative clustering algorithm objects use a bottom-up approach to perform hierarchical clustering. The Birch algorithm builds a tree called the clustering feature tree for the given data. The Gaussian mixture algorithm (Duda, 1973) is a probabilistic algorithm that assumes all the data points are generated from a mixture of a finite number of gaussian distributions with unknown parameters.

2.2. Machine learning algorithm

Many kinds of machine learning algorithms for regression have been proposed. Two well-known algorithms are examined. One is linear and the other is non-linear. The Lasso algorithm (Kim, 2007) is a linear model that estimates sparse coefficients and effectively reducing the number of features. On the other hand, the XGboost algorithm (Tianqi, 2016) works as Newton Raphson in function space unlike gradient boosting that works as gradient descent in function space, a second order Taylor's approximation is used in the loss function to make the connection to Newton Raphson method.

2.3. Objective data

The predictive performances by several clustering and machine learning algorithm were evaluated by the biodegradation data for 4200 chemicals, which were tested under MITI I 301C test condition. The explanatory variables for estimating the biodegradation were chemical descriptors that contain 2D and 3D structural and charge information generated by the Gaussian software (Gaussian, 2019) using the PM6 method. The chemical descriptors were generated from a SMILES specification (SMILES, 2016) of each chemical by using of the AlvaDesc software (Alvadesc, 2019). After cleaning the generated data, 4293 chemical descriptors were obtained.

3. Results and Discussions

The result of clustering the explanatory variables by the explanatory variable was demonstrated in Fig. 1. The performance of various clustering algorithms was indicated in scores. The score is coefficient of determination for validation data by the Lasso for the largest cluster by each clustering algorithm. As shown in Fig.1, the mini-batch K-Means and the Birch algorithms were higher performance than other algorithms.

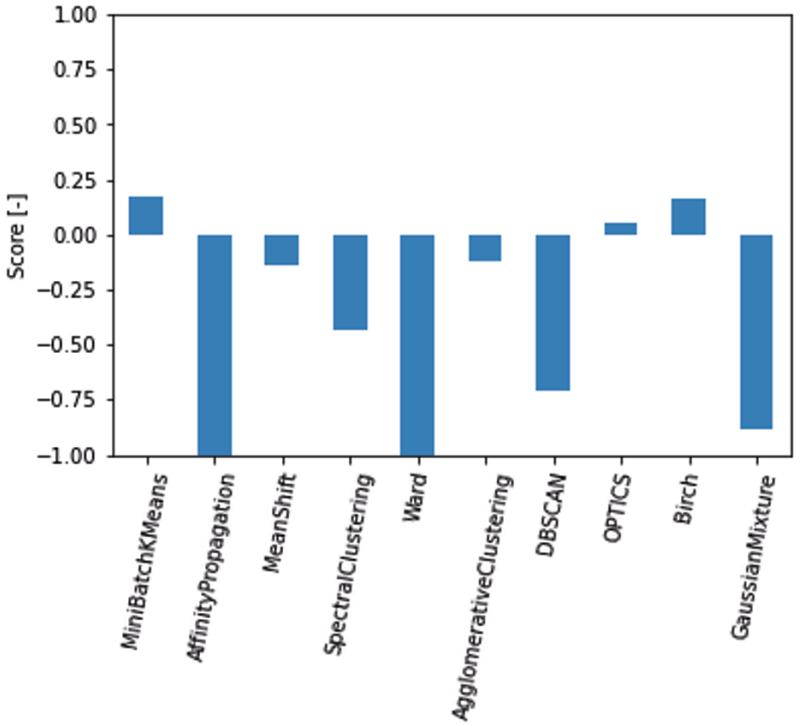


Figure 1 Classification scores of various clustering algorithms.

Then, clusters generated by the mini-batch K-Means and the Birch algorithms were examined. The other clustering algorithm cannot generate large clusters to train. Table 1 shows the evaluation results for various datasets of chemicals in or out of cluster by the mini-batch K-Means algorithm, where RMSE is rooted mean square error for training data, RMSE_V is rooted mean squared error for validation data, R^2 is coefficient of determination for training data, Q^2 is coefficient of determination for validation data, classR^2 is correct answer rate for readily biodegradable product in training data, and classQ^2 is correct answer rate for not readily biodegradable product in validation data. The results were averages of three trials. The biggest cluster generated by the algorithm contained 3718 chemicals. The Lasso or the XGboost algorithm applied to the biggest cluster or the other chemicals. The XGboost algorithm took lower RMSE and RMSE_V, and higher R^2, Q^2, classR^2, and classQ^2 than those of the Lasso algorithm. However, due to overfitting of the training data, the differences between the training data results and the validation data results of the XGboost algorithm were larger than those of the Lasso algorithm.

Table 1 Evaluation results for various datasets of chemicals in/out cluster by the mini-batch K-Means algorithm.

Used Data (number of chemicals)	Learning Algorithm	RMSE	RMSE_V	R^2	Q^2	classR^2	classQ^2
1st cluster (3718)	Lasso	0.550	0.706	0.700	0.491	0.907	0.856
	XGboost	0.006	0.660	1.000	0.567	0.999	0.862
Except for 1st cluster (482)	Lasso	0.314	0.817	0.888	0.325	0.980	0.888
	XGboost	0.000	0.727	1.000	0.510	1.000	0.911

Table 2 shows the evaluation results for various datasets of chemicals in or out of cluster by the Birch algorithm. The biggest cluster generated by the algorithm contained 1878 chemicals. The differences between the training data results and the validation data results of the Birch algorithm were smaller than those of the mini-batch K-Means algorithm. The results of 1st cluster were better than the other clusters results. It is indicated that the 1st cluster might be able to collect chemicals that are clear relationship between the explanatory variables and biodegradation.

Table 2 Evaluation results for various datasets in/out cluster by the Birch clustering algorithm.

Used Data (number of chemicals)	Learning Algorithm	RMSE	RMSE_V	R^2	Q^2	classR^2	classQ^2
1st cluster (1878)	Lasso	0.494	0.733	0.761	0.429	0.960	0.895
	XGboost	0.127	0.671	0.984	0.547	0.992	0.909
Except for 1st cluster (2332)	Lasso	0.755	0.787	0.431	0.378	0.781	0.780
	XGboost	0.181	0.684	0.967	0.534	0.986	0.832
2nd cluster (1239)	Lasso	0.679	0.755	0.528	0.456	0.810	0.801
	XGboost	0.477	0.696	0.773	0.513	0.914	0.823
3rd cluster (955)	Lasso	0.727	0.812	0.469	0.341	0.806	0.764
	Xgboost	0.047	0.768	0.998	0.428	0.997	0.800
4th cluster (39)	Lasso	0.900	1.037	0.179	-0.080	0.716	0.722
	XGboost	0.000	1.093	1.000	-0.232	1.000	0.667

Table 3 shows comparing among various cluster results. The cluster results of the mini-batch K-Means algorithm were a little better than those of whole sample chemicals. On the other hand, the cluster results, especially classQ^2, of the Birch algorithm were better

than those of whole sample chemicals. The classR^2 and classQ^2 of proposed combination algorithm were higher than those of the CATALOGIC. As control experiments, 1800 chemicals were randomly sampled. The results of the control experiment led to the Birch algorithm results not being due to sample size.

Table 3 Evaluation results for various clustering and learning algorithm.

Clustering Algorithm (number of chemicals)	Learning Algorithm	RMSE	RMSE_V	R^2	Q^2	classR^2	classQ^2
Whole sample (4200)	Lasso	0.639	0.711	0.594	0.488	0.895	0.866
	XGboost	0.134	0.649	0.982	0.579	0.992	0.881
1st cluster by mini-batch K-Means (3718)	Lasso	0.550	0.706	0.700	0.491	0.907	0.856
	XGboost	0.006	0.660	1.000	0.567	0.999	0.862
1st cluster by Birch (1878)	Lasso	0.494	0.733	0.761	0.429	0.960	0.895
	XGboost	0.127	0.671	0.984	0.547	0.992	0.909
Random sampling (1800)	Lasso	0.730	0.767	0.471	0.366	0.850	0.836
	XGboost	0.223	0.670	0.944	0.525	0.983	0.872

4. Conclusions

The combination of the Birch clustering algorithm and the XGboost learning algorithm is proposed to estimate the biodegradation of chemicals. The mini-batch K-Means and the Birch algorithms were higher performance than other algorithms. Then, clusters generated by the mini-batch K-Means and the Birch algorithms were examined. The Lasso or the XGboost algorithms applied to the biggest cluster or the other chemicals. The correct answer rate for not readily biodegradable products and correct answer rate for easily decomposable products of proposed combination algorithm were higher than those of the CATALOGIC. The Birch algorithm might be able to collect chemicals that are clear relationship between the explanatory variables and biodegradation. The cluster results of the Birch algorithm were better than the other results. In future works, the characteristics of chemicals contained in the Birch algorithm will be investigated.

Acknowledgements

This study was conducted under the METI contract research “Research on the Introduction of Weight of Evidence Approach to the Evaluations of Biodegradation and Bioaccumulation of Chemicals (2020)”. Deliverables from the cooperative research between Shizuoka University and NITE” Development of Biodegradation Predictive QSAR Method Using AI (2019-2022)” were used in this study. The authors wish to thank Dr. M.Takatsuki (Waseda Univ.) for helpful discussions.

References

AlvaDesc, 2019, https://www.alvascience.com/alvadesc/

J. F. Brendan et.al., 2007, Clustering by Passing Messages Between Data Points, Science Feb.

T. Chen, et al., 2016, XGBoost: A Scalable Tree Boosting System, https://dl.acm.org/doi/epdf/10.1145/2939672.2939785

D. Comaniciu, et.al., 2002, Mean shift: A robust approach toward feature space analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence

Z. David, et. al., 2017, Preconditioned Spectral Clustering for Stochastic Block Partition Streaming Graph Challenge, 2017 IEEE High Performance Extreme Computing Conference

R.O. Duda, el. al., 1973, Pattern Classification and Scene Analysis, Wiley, New York

M. Ester, et. al., 1996, A density-based algorithm for discovering clusters in large spatial databases with noise, Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, 226–231

Gaussian, 2019, https://gaussian.com/

J. Jaworska, et al., 2002, Probabilistic assessment of biodegradability based on metabolic pathways: CATABOL System, SAR QSAR Environ. Res. 13, 2, 307-323

O. Kettani, et. al., 2014, Agglomerative clustering via maximum incremental path integral, Pattern Recognition, 46, 11, 3056–65

S. J. Kim, et. al., 2007, An Interior-Point Method for Large-Scale L1-Regularized Least Squares, in IEEE Journal of Selected Topics in Signal Processing

Laboratory of Mathematical Chemistry, 2021. http://oasis-lmc.org/products/software/catalogic.aspx.

NITE, 2020, https://www.nite.go.jp/data/000111311.pdf

OECD, 1992, http://www.oecd.org/chemicalsafety/risk-assessment/1948209.pdf

E. Schubert, et. al., 2017, DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN, ACM Trans. Database Syst., 42, 3, 19:1–19:21

D. Sculley, 2010, Web Scale K-Means clustering, Proceedings of the 19th international conference on World wide web

A. K. Singh, et al., 2021, Trends in predictive biodegradation for sustainable mitigation of environmental pollutants: Recent progress and future outlook, Science of The Total Environment, 770, 144561

SMILES, 2016, http://opensmiles.org/opensmiles.html

J. Tunkel, et al., 2000, Predicting ready biodegradability in the Japanese ministry of international trade and industry test, Environ. Toxicol. Chem. 19, 10, 2478 -2485

J. H.Ward, 1963, Hierarchical Grouping to Optimize an Objective Function, Journal of the American Statistical Association, 58, 301, 236–244

T. Zhang, et al., 1996, BIRCH: An efficient data clustering method for large databases. https://www.cs.sfu.ca/CourseCentral/459/han/papers/zhang96.pdf

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50289-X

Surrogate Classification based on Accuracy and Complexity

Maaz Ahmad[a], Iftekhar A. Karimi[a,*]

[a]*Department of Chemical & Biomolecular Engineering, 4 Engineering Drive 4, National University of Singapore, Singapore-117585, Singapore*

cheiak@nus.edu.sg

Abstract

The prevalence of various machine-learning modeling techniques and numerous possible model configurations generates a long list of unique surrogate forms. Exhaustive enumeration and search for the best surrogate form from a large pool of candidate forms is a non-trivial task. In this work, we aim to assess similarities in modeling capabilities among many different surrogate forms. We examine modeling capabilities for noisy and non-noisy data based on two different surrogate performance metrics. We use a similarity metric to identify similar pairs of surrogate forms, and then group mutually similar forms into distinct families. The similarities among various forms vary depending on the quality of data set and choice of performance metric. This work enables us to exploit families of similar forms to create a reduced search set of contrasting surrogate forms, and facilitate surrogate form selection.

Keywords: machine-learning, surrogate, surrogate selection.

1. Introduction

With growing digitalization in industrial process operations, the use of digital twins for decision making has been on the rise. While high-fidelity models provide detailed understanding and analysis, they are limited by computational expense. Computationally cheaper, data-driven surrogate models offer a viable alternative. They simply learn the correlations between the input-output data generated by the system under study. The development of a surrogate model starts with the identification of an appropriate modeling technique or a learning algorithm. Artificial Neural Network (ANN), Radial Basis Function (RBF), Support Vector Regression (SVR) are examples of modeling techniques. The next step is to create a surrogate form by choosing its configurations and analytical functions. For instance, an ANN technique using a sigmoid activation function and having one hidden layer with ten nodes defines an ANN form. Similarly, an RBF technique with a linear basis function creates an RBF surrogate form. The last step involves constructing the final surrogate model with known parameters by training the form on a data set. Various possible modeling techniques and associated model configurations can generate innumerable distinct surrogate forms. Determining the best form is demanding not only due to the expansive search required over many candidate forms, but also due to the importance of searching across various diverse forms. While one would expect unique forms to behave differently from the other, recent works (Ahmad and Karimi, 2021; Bhosekar and Ierapetritou, 2018; Garud et al., 2018) have indicated similarities in modeling capabilities among different surrogate forms. Clearly, it would be more rational to search across a set of contrasting surrogate forms to expedite

the selection process. To this end, in this work, we aim to assess and analyse the similarities in performance across many surrogate forms. We group forms with similar performance into distinct families such that all forms within a family are mutually similar. In the subsequent sections, we discuss our scope of work, detail our numerical methodology to construct families of similar surrogate forms, and highlight some key observations. Finally, we conclude in section 5 with the scope of future work.

2. Data sets, Surrogate Forms, and Performance Metrics

For the sake of consistency, let us denote the k^{th} sample point of any data set with N-dimensional inputs as $x_n^{(k)}$ $(n = 1, 2, \dots, N)$. In this work, we only consider single response at any sample point. Let $y^{(k)}$ denote the response at the k^{th} sample point. We train any surrogate form on a given data set to obtain the final surrogate model $S(\boldsymbol{x})$.

2.1. Data sets

We gather diverse data sets from various sources. The presence of noise in response affects the predictive performance of different surrogate models. Hence, we aim to study similarities in surrogate performances for non-noisy and noisy data sets separately. We gather various data sets from 93 analytical test functions (Surjavonic and Bingham, 2013), 20 simulation runs (Coimbatore Meenakshi Sundaram and Karimi, 2021), and plant observations. In total, we have 1508 non-noisy data sets and 1591 noisy data sets, identical to those used in our previous work as well (Ahmad and Karimi, 2021). The maximum input dimension over all non-noisy and noisy data sets was 20.

2.2. Surrogate Forms

We considered 49 surrogate forms from eight modeling techniques: Polynomial Response Surface Model (PRSM), Kriging (KRG), RBF, SVR, Multivariate Adaptive Regression Spline (MARS), ANN, Gaussian Kernel Regression (GKR), and Power Law (PL). Each technique offers various options for functional forms and model configurations to generate many possible distinct surrogate forms. The 49 surrogate forms and their notations are listed in Table 1 in the next page.

2.3. Performance Metric

We use one accuracy-based performance metric and one hybrid accuracy-complexity-based performance metric (PM) to evaluate each surrogate's performance. We consider R^2 or coefficient of determination as the accuracy-based PM, and SQS or Surrogate Quality Score (Ahmad and Karimi, 2021) as the hybrid PM. SQS balances accuracy and complexity by favouring an accurate model based on R^2 while penalizing a complex model based on extent of freedom (eof) or number of independent model parameters.

$$R^2 = 1 - \sum_{k=1}^{K'} \left(y^{(k)} - S\left(\boldsymbol{x}^{(k)}\right)\right)^2 \Bigg/ \sum_{k=1}^{K'} \left(y^{(k)} - \bar{y}\right)^2 \tag{1}$$

$$SQS = (1 - R^2) \times \left\{\frac{\ln(1 + eof)}{\ln(1 + K)}\right\}^{0.5} \tag{2}$$

$\bar{y}$ in Eq. 1 is the mean response, while K' is the number of sample points over which PM is computed. SQS can be used to compare across qualitatively different modeling techniques efficiently.

Table 1: Modeling Techniques, Surrogate Forms, and their Notations used in this work.

Modeling Techniques	**Surrogate Forms (shorthand notation in parenthesis)**	**Notations for Forms**
PRSM	1st order and 2nd order PRSM	PRSM1, PRSM2
KRG	Regression Functions: constant (0), PRSM1 (1), PRSM2 (2) Correlation Functions: Exponential (e), Gaussian (g), Linear (l), Spherical (s), Cubic (c)	K0e, K1e, K2e, K0g, K1g, K2g, K0l, K1l, K2l, K0s, K1s, K2s, K0c, K1c, K2c
RBF	Basis Functions: Bi-harmonic (BH), Multi-quadratic (MQ), Inverse Multi-quadratic (IMQ), Thin Plate Spline (TPS), Gaussian (G) Tail Functions: constant (0), PRSM1 (1), PRSM2 (2)	RBH0, RBH1, RBH2, RMQ0, RMQ1, RMQ2, RIMQ0, RIMQ1, RIMQ2, RTPS0, RTPS1, RTPS2, RG0, RG1, RG2
SVR	Kernel functions: Linear (lin), 3rd order Polynomial (poly), Gaussian (gauss)	SVRlin, SVRpoly, SVRgauss
MARS	{Max interactions, Max basis functions}: $\{2,5N\}$ (1) and $\{3,10N\}$ (2)	MARS1, MARS2
ANN	Activation functions: tansig (T), logsig (L), radial basis (R) {Number of hidden layers, Number of nodes in each layer}: $\{1,N\}$, $\{1,2N\}$, $\{2,N\}$	ANN1TN, ANN1T2N, ANN2TN, ANN1LN, ANN1L2N, ANN2LN, ANN1RN, ANN1R2N, ANN2RN
GKR	Projected dimension $R = 2N$	GKR
PL	Sum and Product of $x_n, n = 1, 2, \dots, N$ terms	APL, MPL

3. Similarity Assessment and Families Identification

Since the performance of any surrogate model would depend on the quality of data (noisy vs non-noisy) and performance metric (R^2 vs SQS), we aim to assess similarities among various forms for the four cases separately. We denote them as NNR2, NNSQS, NR2, and NSQS, where NN and N denote data quality (non-noisy and noisy respectively) while R2 and SQS denote PM (R^2 and SQS respectively). Furthermore, for noisy data, KRG and RBF are not the ideal techniques since they always fit noise. Hence, we consider 49 forms for non-noisy data and 19 forms (excluding 15 KRG and 15 RBF forms) for noisy data.

The first step to initiate our assessment and analysis involves constructing and evaluating all surrogate models for all data sets. A surrogate is trained on a few input-output points which constitute the train set, while it is evaluated over a few additional new sample points which constitute the test set. To this end, for synthetic and simulated data, we generate additional K sample points for each data set to make up a test set with $2K$ sample points (train set + K new points). For real-world data, we randomly select a few points ($ceil(K/2)$) as the train set, while evaluate at all K points. This gives us a performance vector (PV) for each surrogate. A surrogate's PV consists of its PM values for all data sets. The next step involves identification of similar pairs of surrogates. For this, we used

the concept of concordance correlation coefficient (ρ_C) (Lin, 1989) that measures similarity between two vectors of data based on the expected value of squared deviation between them. ρ_C considers the direction of Pearson's correlation and is a signed metric. However, in our case, we simply need to check for resemblance between the performance vectors of any two surrogates. Hence, we develop a metric Similarity Index (SI) by modifying the normalization factor used by ρ_C such that $SI \in [0,1]$. $SI = 1$ for identical performance vectors.

$$SI\big(S_1(\boldsymbol{x}), S_2(\boldsymbol{x})\big) = 1 - \frac{E\big(PV_{S_1(\boldsymbol{x})} - PV_{S_2(\boldsymbol{x})}\big)^2}{E\left[\big(PV_{S_1(\boldsymbol{x})} - PV_{S_2(\boldsymbol{x})}\big)^2 | \rho = -1\right]} \tag{3a}$$

$$SI\big(S_1(\boldsymbol{x}), S_2(\boldsymbol{x})\big) = \frac{2(1+\rho)\sigma_{S_1(\boldsymbol{x})}\sigma_{S_2(\boldsymbol{x})}}{\big(\mu_{S_1(\boldsymbol{x})} - \mu_{S_2(\boldsymbol{x})}\big)^2 + \big(\sigma_{S_1(\boldsymbol{x})} - \sigma_{S_2(\boldsymbol{x})}\big)^2} \tag{3b}$$

ρ, μ, and σ denote Pearson's correlation coefficient, mean, and standard deviation for the two PVs for $S_1(\boldsymbol{x})$ and $S_2(\boldsymbol{x})$. We compute SI between all pairs of surrogates for each of NNR2, NNSQS, NR2, and NSQS. The final step involves grouping mutually similar forms into distinct families. We use $SI \geq 0.90$ to consider two forms to be similar. Then, we create families of similar forms by solving an MILP formulation. Considering I forms, let us consider a similarity matrix with $a_{ij} = 1$ if $SI(i,j) \geq 0.9$, otherwise $a_{ij} = 0$, for $i = 1, 2, \dots, I; j > i$. We use a binary variable f_i to indicate whether form i belongs to a family ($f_i = 1$) or not ($f_i = 0$). First, we find the family with maximum size, or most number of mutually similar forms. This entails solving $\max \sum_{i=1}^{I} f_i$ subject to the following similarity constraint:

$$f_i + f_j \leq 1 + a_{ij}, \quad j > i \tag{4}$$

Eq. 4 ensures that if forms i and j belong to a family, then they must be similar. Solving this problem gives us the largest family with size S. Then, starting with $s = S$, we identify all distinct families of size $s = S, S-1, S-2, \dots, 1$ one at a time. This essentially reduces to a feasibility problem to identify a family of a given size, if it exists. This can be achieved as follows.

$$\min obj = 1 \tag{5a}$$

$$\sum_{i=1}^{I} f_i = s \tag{5b}$$

Solving Eq. 5a subject to constraints Eq. 4 and Eq. 5b gives us one family of size s. Since, we get one family on each run and multiple families of size s may exist, we re-solve the MILP after adding the previously obtained solution (z_i^*) as a new cut-constraint (Eq. 5c).

$$\sum_{i=1}^{I} z_i^* \leq s - 1 \tag{5c}$$

This would eliminate the problem to yield previously found families or their subsets in subsequent runs. In case the problem is infeasible, we search for the next smaller sized family with size $s-1$. This procedure is repeated for each $s = S, S-1, S-2, \dots, 1$ to get a complete set of families.

4. Families for NNR2, NNSQS, NR2, and NSQS

Based on the detailed methodology discussed in the previous section, we obtain different sets of families for NNR2, NNSQS, NR2, and NSQS (Tables 2 – 5).

Table 2: Families of surrogate forms for NNR2

	Surrogate Forms		**Surrogate Forms**
F1	K0l, K1l, K0s, K1s, K0c, K1c	F6	RIMQ1, RIMQ2
F2	K0e, K1e, K0l, K1l, K0s, K1s	F7	RTPS1, RTPS2, RG0
F3	K0g, K1g	F8	MARS1, MARS2, ANN1T2N
F4	RG1, RG2	F9	ANN1TN, ANN1T2N, ANN2TN, ANN1LN, ANN2LN, ANN2RN
F5	K2e, K2l, K2s, K2c, RBH0, RBH1, RBH2, RMQ0, RMQ1, RMQ2, RIMQ0, RTPS0	F10	PRSM2, SVRpoly, SVRgauss, ANN1TN, ANN1T2N, ANN2TN, ANN1LN, ANN1L2N, ANN2LN, ANN1RN

Table 3: Families of surrogate forms for NNSQS

	Surrogate Forms		**Surrogate Forms**
F1	K0l, K1l, K0s, K1s, K0c, K1c	F8	MARS1, MARS2
F2	K0e, K1e, K0l, K1l, K0s, K1s	F9	ANN2TN, ANN2LN, ANN2RN
F3	K0g, K1g	F10	ANN1TN, ANN1T2N, ANN2TN, ANN1LN, ANN1L2N, ANN2LN, ANN1RN
F4	RG1, RG2	F11	PRSM2, SVRgauss, ANN1TN, ANN1T2N, ANN1LN, ANN1L2N
F5	K2e, K2l, K2s, K2c, RBH0, RBH1, RBH2, RMQ0, RMQ1, RMQ2, RIMQ0, RTPS0	F12	PRSM2, ANN1TN, ANN1T2N, ANN1LN, ANN1L2N, ANN2LN, ANN1RN
F6	RIMQ1, RIMQ2	F13	PRSM2, K2l, K2s, K2c, RBH0, RBH1, RBH2
F7	RTPS1, RTPS2, RG0		

Table 4: Families of surrogate forms for NR2

	Surrogate Forms		**Surrogate Forms**
F1	SVRgauss, SVRpoly, ANN1TN, ANN1L2N	F6	SVRgauss, ANN1T2N, ANN1LN, ANN2LN
F2	SVRgauss, MARS1, ANN1TN, ANN1L2N	F7	SVRgauss, ANN1T2N, ANN1L2N, ANN2LN
F3	SVRgauss, MARS1, ANN1T2N, ANN1L2N	F8	ANN2TN, ANN2LN
F4	SVRgauss, ANN1TN, ANN1LN, ANN2LN	F9	PRSM2, SVRpoly
F5	SVRgauss, ANN1TN, ANN1L2N, ANN2LN	F10	SVRgauss, MARS1, MARS2

Table 5: Families of surrogate forms for NSQS

	Surrogate Forms		**Surrogate Forms**
F1	MARS1, MARS2	F3	ANN1TN, ANN1T2N, ANN1LN, ANN1L2N, ANN2LN
F2	ANN2TN, ANN2LN		

The derived sets of families reveal some important observations. For NNR2 and NNSQS, KRG forms with a constant or linear regression function show similarities irrespective of the correlation function, with the exception of Gaussian correlation. Similarly, K2e, K2l, K2s, and K2c show similarities, but K2g is not similar to any of them. This indicates a significantly contrasting behaviour of KRG using Gaussian correlation functions than others. Many RBFs show similarities with KRG using PRSM2 regression function for NNR2 and NNSQS. Both KRG and RBF techniques precisely learn the responses, and

hence can potentially make similar predictions. Since KRG and RBF have identical and maximum complexity ($eof = K$), SQS penalizes both models identically. Hence, we observe the same sets of families F1-F7 for NNR2 and NNSQS. For all cases, we see that MARS1 and MARS2 are similar as they only differ in their hyperparameter settings. Also, ANNs using *tansig* and *logsig* activations readily show similarities irrespective of the network depth and width. This may be attributed to similar analytical forms of the two activation functions, and simple, shallow networks considered in our work. However, ANNs using radial basis functions differ based on their network configurations. We also observe that SVRgauss shows similarities with many ANN forms for NNR2, NNSQS, and NR2. Both SVRs and ANNs offer sufficient fidelity to model complex data sets effectively. One final observation from all families is that certain forms such as PRSM1, SVRlin, GKR, APL, and MPL do not belong to any family. In other words, such forms do not show similarities with any other form for either non-noisy or noisy data based on either PM.

The key application of families of similar forms is to facilitate and expedite surrogate form selection. In our future work, we aim to extract one representative surrogate form from each family to obtain a reduced search space of contrasting forms. We then search among the surrogate forms in this reduced search space. Based on the performance of the true best surrogate form and the form identified from the reduced space, we can validate the efficacy of our proposed surrogate form selection procedure.

5. Conclusions

This work aims to compare performances of various surrogate forms for modeling non-noisy and noisy data sets based on two performance metrics. Based on our numerical evaluation of similarity, we group mutually similar surrogate forms into families. Certain forms having different modeling techniques show similar performance, while certain others do not resemble with any other form. This work may act as an important prelude to our future work on expediting the surrogate form selection process by identifying and searching across contrasting surrogate forms.

References

Ahmad, M., Karimi, I.A., 2021. Revised learning based evolutionary assistive paradigm for surrogate selection (LEAPS2v2). Computers & Chemical Engineering 152, 107385.
Bhosekar, A., Ierapetritou, M., 2018. Advances in surrogate based modeling, feasibility analysis, and optimization: A review. Computers & Chemical Engineering 108, 250–267.
Coimbatore Meenakshi Sundaram, A., Karimi, I.A., 2021. State transients in storage systems for energy fluids. Computers & Chemical Engineering 144, 107128.
Garud, S.S., Karimi, I.A., Kraft, M., 2018. LEAPS2: Learning based Evolutionary Assistive Paradigm for Surrogate Selection. Computers & Chemical Engineering 119, 352–370.
Lin, L.I.-K., 1989. A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometrics 45, 255.
Surjavonic, S., Bingham, D., 2013. Virtual Library of Simulation Experiments: Test Functions and Datasets. [WWW Document].

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50290-6

Training Stiff Dynamic Process Models via Neural Differential Equations

William Bradley[a], Gabriel S. Gusmão[a], Andrew J. Medford[a], Fani Boukouvala[a]*

[a]*Georgia Institute of Technology, School of Chemical & Biomolecular Engineering, 311 Ferst Dr., Atlanta, GA, 30332, USA*
fani.boukouvala@chbe.gatech.edu

Abstract

A common step in developing generalizable, dynamic mechanistic models is to fit unmeasured parameters to measured data. Fitting differential equation-based models can be computationally expensive due to the presence of nonlinearity and stiffness. This work proposes a two-stage indirect approach where Neural ODEs approximate state derivatives, which are used to estimate the parameters of a differential model. In addition to its computational efficiency, the proposed method demonstrates the ability to work in concert with direct methods to accurately estimate parameters, even in the case of stiff systems. The method is shown here for the training of a microkinetic model.

Keywords: Neural Networks, Parameter Estimation, Stiff ODEs, Neural ODEs.

1. Introduction

The task of finding parameter values of a differential equation (DE) model to explain available experimental data is ubiquitous throughout engineering. The physical meaning of these DE models (also referred to here as a mechanistic model) permit the modeler to predict a system's behavior in unexplored experimental spaces, assuming the parameters have been estimated correctly. However, due to the complexity of DE systems, methods that automate their parameter estimation must often balance efficiency and accuracy. Gradient-based 'direct' methods either rely on repeated integration of the ODEs being regressed, or formulating a constrained nonlinear program discretizing the system of ODEs to solve for the parameter values (Li et al. 2005, Hamilton 2011). Both 'direct' methods face computational tractability issues, which become more severe when the initial parameter estimates are far from the true values, or the ODEs are nonlinear with respect to their parameters. Another problem, common to reaction systems, is the presence of rate terms which vary over large orders of magnitude, resulting in a system with fast and slow dynamics (i.e., at different timescales). Ultimately, to make these regression problems tractable for direct methods, a modeler may need to apply model reduction strategies, ranging from setting tight bounds on parameters to fixing insensitive parameters. Such strategies require domain expertise, which may not be available, as well as user-intervention, preventing automation of the parameter estimation process.

As an alternative to the direct approach, an indirect parameter estimation approach has been proposed, which avoids discretizing the mechanistic model (Swartz and Bremermann 1975, Brunel 2008). In this 2-stage approach, the experimental data is interpolated by a data-driven model, which is differentiated to obtain system derivative estimates. Those derivative estimates combined with state estimates of the interpolating model can be used to estimate the parameters of the mechanistic DEs via nonlinear programming (NLP). The indirect 2-stage approach is so named since it breaks up a single regression problem into two regression problems whose combined computational

cost is generally less than that of the direct approaches. Yet despite having the advantage of being computationally cheap, this method is often limited in accuracy due to the difficulty in accurately estimating a system's derivative information.

Recently, we proposed using Neural ODEs (NODEs) as the data-driven surrogate to interpolate measurement data (Figure 1) and for estimating system derivatives (Bradley and Boukouvala 2021). In that work, NODEs compared favorably with other methods for automated parameter estimation of a nonlinear mechanistic DE. However, one class of DEs not covered in that work were those with 'stiff' dynamics. This class of problems can be particularly challenging for parameter estimation methods. One reason for this is the need for numerical methods that balance the number of functional evaluations (i.e., computation) and stability (i.e.. accuracy). Recent work has evaluated numerical techniques for fitting Neural ODEs to stiff system data, and for parameter estimation of stiff systems (Kim et al. 2021), however further work is needed to develop methods that are both general and accurate.

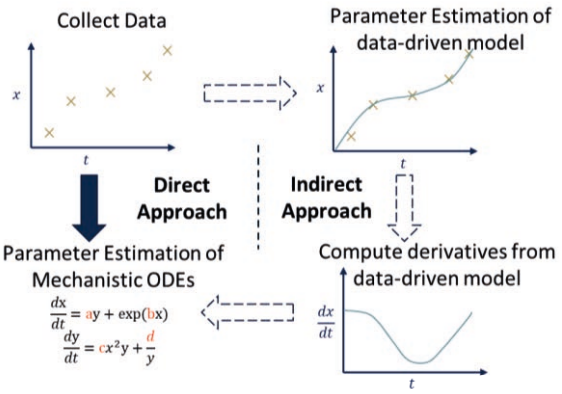


Figure 1. Depiction of the direct vs indirect approach to parameter estimation

2. Methods

In this work, several approaches, and potential combinations thereof, are compared for the parameter estimation of stiff DEs. To start, direct approaches find the parameters p to a mechanistic model $f(x,p)$ by minimizing the following discrepancy function:

$$\min \sum (x_{k,j,meas} - x_{k,j,pred})^2 \tag{1}$$

$$s.t.\ \frac{dx_{k,MM}}{dt} = f(x_k, p) \tag{2}$$

Here, K state variables x_k, where $k = 1, \dots, K$, are measured and predicted at time points j, where $j = 1, \dots, J$, by integrating the mechanistic model (MM) with respect to independent variable t. Though statistically robust, this method can be computationally intensive. For such cases, a 2-stage indirect approach can be attractive.

As illustrated in Figure 1, the 2-stage indirect approach fits the parameters of the mechanistic model by solving 2 separate regression problems. In the first stage, the parameters of the data-driven model are fitted using the original measurement data. In the second stage, the parameters of the mechanistic ODE are found using the state and derivative estimates of the data-driven model. The data-driven model used in our work is a NODE model. This is done by first solving Eq.(1) subject to Eq.(3):

$$s.t.\ \frac{dx_{k,NODE}}{dt} = NN(x_k, w) \tag{3}$$

Neural Network parameters w are fitted to minimize an objective function equal to the sum of squared errors between the model prediction and measured state data. Once the

NODE is trained, derivative estimates are obtained by evaluating the trained NODE at times where measured data was collected using the same process conditions of the measured data. Following the procedure of (Bradley and Boukouvala 2021), we exclude derivative estimates at time t=0, which tend to be less reliable, to improve parameter estimates of the mechanistic DE. For stage two, an NLP is formulated as in Eq.(4) and (5) to find the parameters of the mechanistic DE without integrating the mechanistic DE.

$$\min \sum \left(\frac{dx_{j,k,NODE}}{dt} - \frac{dx_{j,k,MM}}{dt} \right)^2 \tag{4}$$

$$s.t. \quad \frac{dx_{j,k,MM}}{dt} = f(x_{j,k,NODE}, p) \tag{5}$$

Depending on the required accuracy, the indirect 2-stage approach may be sufficient for the needs of the model-building problem at hand. However, if increased accuracy is required, we hypothesized a more robust fit would require including the mechanistic model constraints when fitting the measured state data. A tempting option would be a simultaneous approach, which combines the objective functions of the 2-stage approach into a single hybrid objective function:

$$\min \sum (x_{k,j,meas} - x_{k,j,pred})^2 + \lambda \sum \left(\frac{dx_{j,k,NODE}}{dt} - \frac{dx_{j,k,MM}}{dt} \right)^2 \tag{6}$$

Like the indirect approach, Eq.(6) uses the data-driven Neural ODE of Eq.(2) to fit the state data and provide derivative estimates. However, in the hybrid objective both the Neural ODE fit and mismatch between NODE and mechanistic DE are minimized simultaneously, their relative weights controlled by the hyperparameter lambda, λ.

A final alternative to increasing model fidelity is to fit the mechanistic DE directly (i.e., minimize Eq.(1) subject to Eq.(2)). However, as mentioned earlier this incurs an increased compute overhead. In the case of stiff systems, the increased compute cost comes from the finer discretization required to stably integrate the mechanistic model. To reduce compute costs, the direct approach can use parameter estimates informed by the indirect approach. Specifically, the parameters estimated from the 2-stage fitting are used as an initial guess for the DE of Eq.(2). A single application of the indirect followed by the direct approach is herein referred to as the incremental approach.

Throughout this work, we use PyTorch's LBFGS solver and IPOPT within PYOMO as the nonlinear optimizers of the stage 1 and stage 2 regression problems, respectively. For the sake of consistency, the structure of the NODE is fixed to a single hidden layer with tanh activation function and 15 hidden nodes. Further, we assume minimal knowledge of the true parameters prior to model-fitting, and thus all parameters are initialized to the same order of magnitude, specifically a value of 2, for the direct and indirect approaches.

3. Results

To demonstrate the effectiveness of the 2-stage approach, we chose as an example a microkinetic model (MKM) for heterogeneous catalysis (Gusmão et al. 2020). MKMs represent a large class of coupled differential equations which exhibit stiffness due to the presence of both slow and fast rate terms caused by parameter values varying over large orders of magnitude. The MKM system of ODEs governed by a material balance and rate equations are outlined in Figure 2. Table 1 lists the true parameter values.

$$\frac{d[A]}{dt} = -r_1 \qquad \frac{d[C]}{dt} = -r_3 \qquad r_1 = k_3[A][*] - k_4[A*]$$

$$\frac{d[A*]}{dt} = r_1 - r_5 \qquad \frac{d[C*]}{dV} = r_3 + r_5 \qquad r_2 = k_5[B][*] - k_6[B*]$$

$$\frac{d[B]}{dt} = -r_2 \qquad \frac{d[D*]}{dt} = 2r_4 - r_5 \qquad r_3 = k_7[C][*] - k_8[C*]$$

$$\frac{d[B*]}{dt} = r_2 - r_4 \qquad \frac{d[*]}{dt} = -r_1 - r_2 - r_3 - r_4 + r_5 \qquad r_4 = k_{11}[B*][*] - k_{12}[D*]^2$$

$$r_5 = k_{13}[A*][D*] - k_{14}[C*][*]$$

Figure 2. Full MKM system of ODE equations

In this process, gaseous reactants A and B adsorb to a solid surface to form intermediate species before the final product C desorbs into the gas phase. Reactants bound to a catalyst surface site [*] are indicated by an asterisk '*'. All reactions are reversible.

Two datasets were used to represent possible fitting scenarios for the 2-stage approach, each comprising data simulated from two sets of initial conditions. In one dataset, state variables are sampled 15 times logarithmically for each run in the range t = [10e-3, 0.5], amounting to a sample size of 30 datapoints. The second dataset includes the same number of points sampled linearly from time t=0 to t=0.5. At first, the 2-stage approach was applied on the linear dataset. Specifically, the data was used to fit a NODE whose derivatives were then used to solve for the parameters of the mechanistic DE. The parameters found through this approach are compiled in Table 1, column labelled 'Linear Indirect'. Results show that some of the parameters found differ significantly from the true parameters. This is not surprising since data is available only sparsely at earlier times where state values change rapidly due to the stiffness of the system.

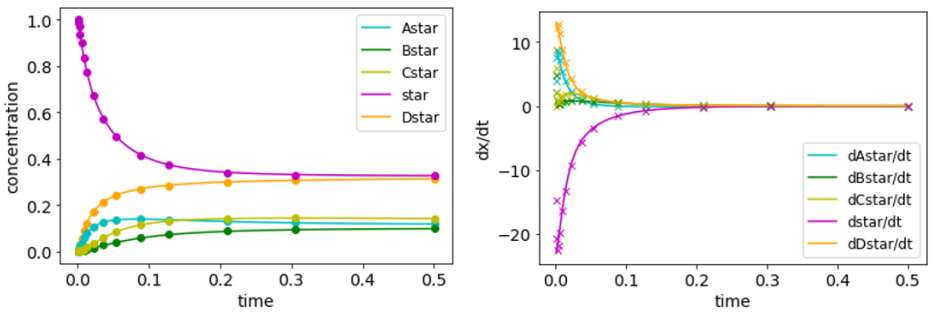


Figure 3. NODE fit (solid lines) to log-sampled data (dots). True derivative shown with x's.

The fitting procedure was repeated with data sampled logarithmically with respect to time. The fit of the Neural ODE to the log-sampled data is presented in Figure 3 for the adsorbed species. Noticeably, despite the NODE fitting the state outputs perfectly (effectively to machine precision), the data-driven model does not capture the exact profile of the derivatives. This result is believed to be due to the inherent flexibility of NODEs, which are not as constrained in outcomes as the simulating mechanistic model. The results of the 2-stage regression including the mean absolute error (MAE) of the fitted model on the log-sampled data are compiled in Table 1 ('Log Indirect' column).

Aiming to improve the accuracy of the fitted mechanistic model, the simultaneous approach was applied using various values for lambda. However, minimizing the hybrid objective function did not result in significantly improved parameter estimates vs

the indirect approach, notwithstanding its higher compute cost. This finding was again attributed to the flexibility of NODEs, their being able to interpolate state data despite estimating derivatives that may not exactly match the 'true' derivatives. Due to their low accuracy, results of the simultaneous approach were not included in Table 1.

Instead, the remaining columns in Table 1 include the computational cost and model accuracy from integrating the mechanistic DE during training, either using an uninformed initial guess (i.e., the direct approach) or initializing the mechanistic parameters with the parameters found by the 2-stage methods (i.e., the incremental approach). Figure 4 displays the mechanistic model fit to the linearly-sampled data via the indirect and direct approach. The direct and incremental approaches gave similar simulated trajectories so only the results of the direct method are plotted.

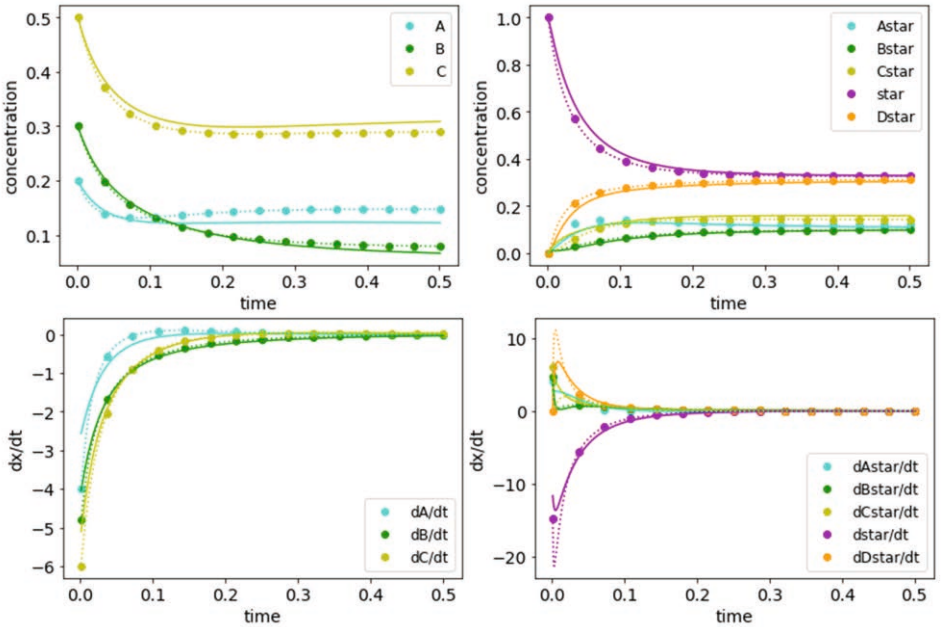


Figure 4. State and derivative estimates of the mechanistic model after parameter estimation via indirect (solid line) and direct approach (dotted line) on linearly-sampled data (solid dots).

A couple trends are worth noting. Firstly, the MAE of the final simulation is lower after using the direct approach, regardless of sampling strategy, indicating increased accuracy can be gained via the direct approach. What's more, applying the incremental approach offers compute savings over the direct approach with an uninformed initial guess, at least for the log-sampled case. However, when fitting the linearly-sampled data, the compute savings from incremental approach are negligible. At least two factors are believed to cause this discrepancy. First, the parameters found through the indirect approach on the linear data were further from the true parameters than for the log-sampled case, offering a poorer initial guess. Second, an increased number of Euler steps were required between datapoints for integrating mechanistic model on the log-sampled data (n=256 vs 56 in the linear case) to avoid divergence issues near the equilibrium region, exacerbating the computational load in the log-sampled case when uninformed initial estimates are used. Ultimately, this indicates that, given a sufficiently sampled experimental space, the incremental approach can merge the direct and indirect approaches in ways that balance both accuracy and efficiency.

Table 1. Table of compute times, parameters estimated, and model errors for different approaches

	True params	Log Indirect	Log Incr.	Log Direct	Linear Indirect	Linear Incr.	Linear Direct
Fit (s) Time	N/A	15.21	137.7	362.31	12.62	74.22	80.94
MAE	N/A	2.46E-3	6.05E-4	4.52E-4	1.45E-2	4.70E-4	2.69E-4
k_3	20	20.46	19.98	19.97	12.82	19.67	19.80
k_4	8	9.061	7.983	7.994	4.556	7.854	7.917
k_5	16	16.50	15.69	15.81	13.49	16.07	15.84
k_6	4	3.490	3.825	3.886	2.637	4.051	3.938
k_7	12	11.38	11.99	12.09	10.19	11.89	11.92
k_8	8	7.695	7.998	8.048	6.719	7.940	7.957
k_{11}	1200	2615	2607	1793	400.4	446.5	1809
k_{12}	400	849.3	871.6	600.8	138.7	147.6	604.0
k_{13}	2000	1672	1662	1117	38.28	2999	1745
k_{14}	1600	1320	1332	890.3	24.05	2401	1395

*Abbreviations: Incr. (Incremental Approach)

4. Conclusions

This work demonstrated a method for accelerating the regression of mechanistic ODEs for stiff systems and evaluated the ability of NODEs to estimate mechanistic ODE parameters with a large magnitude of variability in their true values using different sampling strategies. While the NODE-based incremental approach presents a promising step towards automated parameter estimation of stiff systems, several challenges remain. Neural Networks have limited ability to make predictions that vary over large orders of magnitude, common to stiff systems, which to overcome may require modifying the data-driven model structure to enable greater accuracy. In future, a comparison with incremental and simultaneous methods employing other surrogate models should be performed to assess the Neural ODE's suitability as a general-purpose DE estimator.

References

Bradley, W. and F. Boukouvala (2021). "Two-Stage Approach to Parameter Estimation of Differential Equations Using Neural ODEs." Industrial & Engineering Chemistry Research.

Brunel, N. J. B. (2008). "Parameter estimation of ODE's via nonparametric estimators." Electron. J. Statist. **2**: 1242-1267.

Gusmão, G. S., A. P. Retnanto, S. C. da Cunha and A. J. Medford (2020). "Kinetics-Informed Neural Networks." arXiv preprint arXiv:2011.14473.

Hamilton, F. (2011). "Parameter Estimation in Differential Equations: A Numerical Study of Shooting Methods." SIAM Undergraduate Research Online **Volume 4**.

Kim, S., W. Ji, S. Deng and C. Rackauckas (2021). "Stiff Neural Ordinary Differential Equations." Chaos **31 9**: 093122.

Li, Z., M. R. Osborne and T. Prvan (2005). "Parameter estimation of ordinary differential equations." IMA Journal of Numerical Analysis **25**(2): 264-285.

Swartz, J. and H. Bremermann (1975). "Discussion of parameter estimation in biological modelling: Algorithms for estimation and evaluation of the estimates." Journal of Mathematical Biology **1**(3): 241-257.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50291-8

Wiz 4.0: A Novel Data Visualisation and Analytics Dashboard for a Graphical Approach to Industry 4.0

Louis Allen,[a†] Jack Atkinson,[a†] Joan Cordiner,[a] Mohammad Zandi,[a] and Peyman Z. Moghadam[a,*]

[a]*Department of Chemical and Biological Engineering, The University of Sheffield, Sheffield S1 3DJ, UK*
p.moghadam@sheffield.ac.uk

[†] These authors contributed equally.

Abstract

Proliferation of data owing to the onset of Industry 4.0 (I4.0) has led to many traditional data analysis approaches becoming redundant. Novel and innovative solutions are required to facilitate the new era of data-driven manufacturing characteristic of I4.0. This work demonstrates one such approach in the formation of a bespoke web-based visualisation and machine learning analytics platform, designed to bridge the gap between the old ways and new. Our unique I4.0 data analytics platform, called Wiz 4.0, enables advanced big data analytics in conjunction with user-friendly features and multivariate data visualisations. This allows for both a holistic overview of manufacturing processes as well as detailed analysis of data. Wiz 4.0 lays the foundations of an industry defining software to grant deep insight into the inner relationships between process variables to the everyday user. The software provides the ability to analyse data using a variety of machine learning algorithms and plot the data in high dimensional space through the innovative no-code platform hosted on the Siemens MindSphere. This approach is set to revolutionise the value creation of data in the new IoT and smart factory paradigms emerging from the transition towards I4.0.

Keywords: visualization, analytics, dashboard, IoT, Industry 4.0

1. Background

Widespread development of technologies such as cloud computing, Internet of Things (IoT), Automated Intelligence (AI), additive manufacturing (AM) and digital twins are driving the adoption of new working paradigms in manufacturing. These have the potential to revolutionise the sector when implemented effectively to create large autonomous systems and smart factories. At the root of this is an increased reliance on data. The transition is seen as a new age in manufacturing and wider industry; this fourth industrial revolution (Industry 4.0, I4.0) is associated with increased productivity, efficiency, and profit.

A joint 2019 study by Deloitte and the Manufacturers Alliance for Productivity and Innovation (MAPI) (Wellener et al. 2019) presented key findings about the benefits of smart factory initiatives. While a direct connection was found between these smart factory initiatives and business value via an increase in metrics such as manufacturing output, labour utilisation and labour productivity, only 3% of the sample indicated full scale smart factory adoption. These conflicting findings indicate a barrier preventing the manufacturing industries shift to I4.0. An aging workforce (Oppert & O'Keeffe, 2019) and lack of I4.0 expertise (Raj et al., 2020) is likely to decrease the agility of small and medium enterprise (SME) manufacturers to changes needed and will hinder the adoption of technology associated with I4.0 and smart factory initiatives. In comparison, larger firms who because of their size have an increased budget for retraining and upskilling of work, as well as a higher turnover and intake of younger staff, are more adaptable to change. Organisations struggling to adopt I4.0 require a stepping-stone approach toward digitalization, with a shallow learning curve and low barrier- for-entry.

Data visualization techniques offer an attractive way of jump-starting this shift. Effective visualizations offer universal understanding of complex data and relationships. If this can be combined with manufacturing specific analytics, utilising machine learning (ML) for applications such as predictive maintenance or autonomous control, far-reaching business value can be realized at a fraction of the cost of full-scale digitalisation. Furthermore, visualization of data will aid the operator in the role of abstract critical thinking and diagnoses, something that will certainly become more prominent as increasing levels of automation are used for ever more complex tasks.

Here, we have demonstrated the shift to digital manufacturing at the University of Sheffield's multi-million-pound Diamond Pilot Plant (DiPP), initially installed in 2017. The continuous powder processing plant equipment, shown in Figure 1, can produce 20

Figure 1. The DiPP's ConSigma tabletting line equipment. Includes: (a) twin screw granulator, (b) 6-chamber fluidised bed dryer, (c) mill and (d) continuous tableting press.

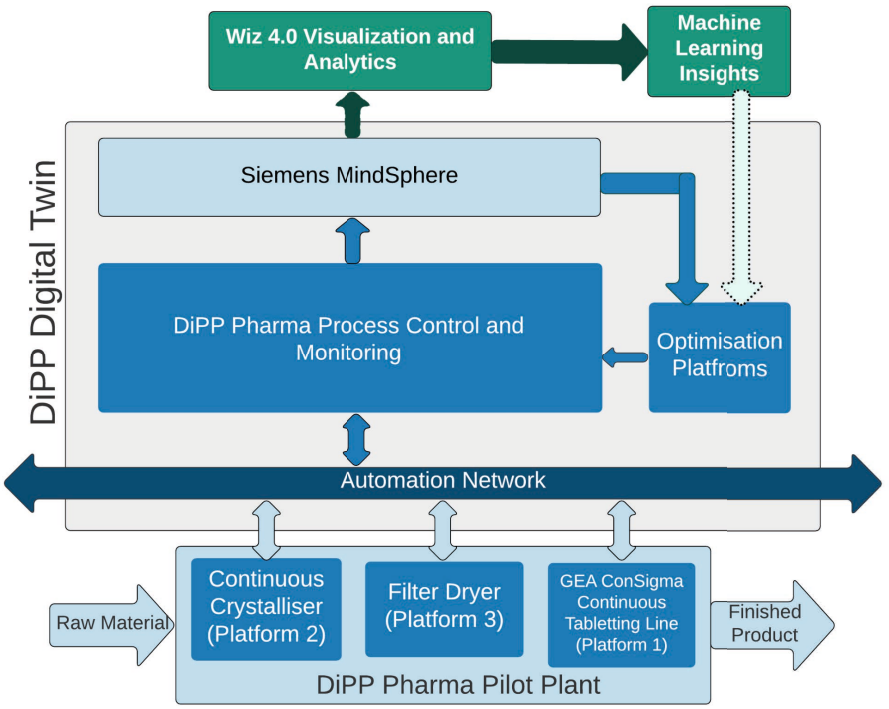


Figure 2. A schematic showing its digitization architecture. Wiz 4.0 can utilise Siemens MindSphere for data collection to create intuitive visual and analytical insight. The software has been made modular to allow the integration of future analytics such as process optimisation, fault detection, and predictive maintenance. This insight can affect the overall process decision making and is indicated by the dashed arrow. Figure adapted from Zandi, 2019.

kg/h of tablets and includes an AWL filter dryer, Nitech continuous crystalliser and GEA ConSigma tableting line. These were retrospectively digitized through the combined efforts of The University of Sheffield, Siemens, and Perceptive Engineering, resulting in an IoT enabled world-leading industrial demonstrator. Figure 2 shows how this digitalisation has allowed a data-driven approach to advanced process control and automation due to the plant's ability to collect, align and centrally store data in Siemens MindSphere: Siemens' IoT platform. In this industrial microcosm, these developments presented an exciting and unique opportunity to engineer Wiz 4.0.

2. Methods

When developing the software, careful consideration was given to the user interface to directly address the barrier to entry and expertise challenges identified in I4.0. In addition, an emphasis was placed on an effective integration of high-dimensional visualisations for the dashboard interface to maintain power in real-world manufacturing scenarios where analysis and exploration of results is routinely multidimensional. With this in mind, the objectives of the software we developed were i) to map and understand complex relationships between process variables in an easy and intuitive way and ii) to consider

users in the design stage i.e. the software must be accessible to those from a range of backgrounds across manufacturing shop floors, including those typically unfamiliar with data science tools.

Figure 3 demonstrates the interaction of the different packages and frameworks utilised in the formation of the software. In general terms, we can bind the analytical capabilities to the backend processing. This includes cloud data storage using AWS, and regression, classification and outlier detection algorithms from SciKitLearn (Pedregosa et al., 2011). Similarly, the visualisation capabilities are expressed through the frontend. Plotly.js utilises Web Graphics Library (WebGL) enabling the high fidelity visualisation of large datasets. Use of React enables dynamic dashboard functionality, enabling the user to configure their own individual space to analyse and observe their data. Figure 3 also details the link of the software to Siemens MindSphere via the MindSphere API.

3. Results and Discussion

Visualisation: Data taken from programmable logic controllers (PLCs) in the DiPP is uploaded to the Siemens MindSphere cloud foundry, where the application can access it through the Siemens MindSphere API. The front end makes a call to the API every two seconds, refreshing user-constructed visualisations in real time allowing for an understanding of conditions within the process equipment at any given point. The extended range of visualisations the software is capable of grants versatility, allowing for both a holistic approach as well as enabling detailed insight. Figure 4 demonstrates this with both a

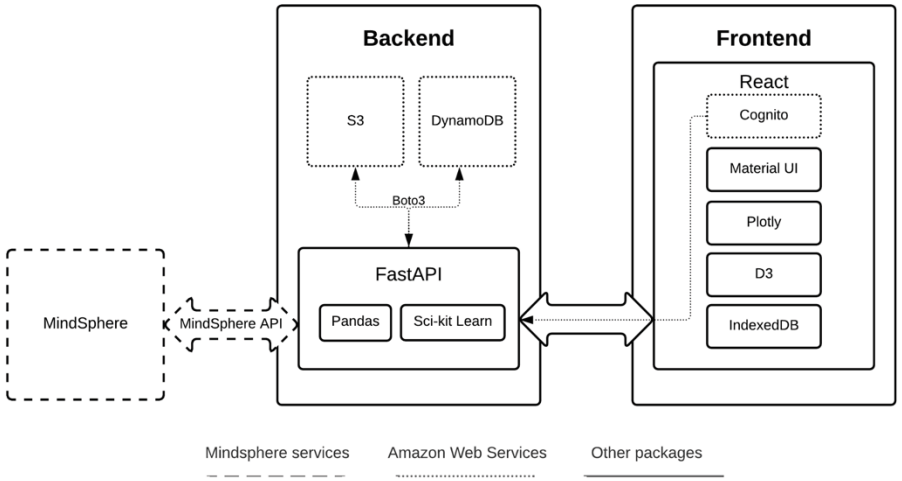


Figure 3. Schematic detailing interaction of fundamental packages and frameworks used in the construction of the software including front end construction in React JavaScript utilising Plotly.js charting library, and back-end construction utilising various AWS packages and analytics modules, constructed in Python, using SciKitLearn ML packages.

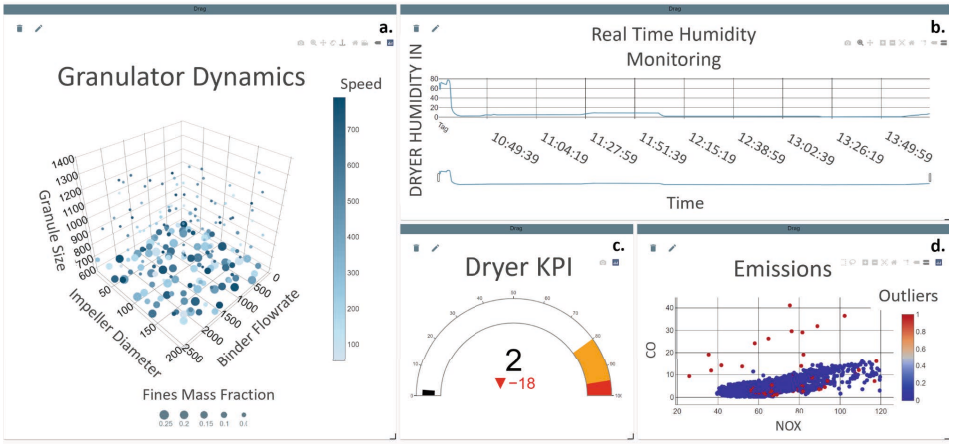


Figure 4: Wiz dashboard interface developed in Siemens MindSphere: (a) 5-dimensional plot showing conditions within the twin screw granulator, (b) real-time monitoring of the humidity in the fluidised bed dryer, (c) gauge plot showing dryer temperature as a key performance indicator (KPI) showing how the data is trending, and (d) scatter plot visualisation of local outlier factor (LOF) unsupervised machine learning algorithm showing outlier data detection.

detailed 5-dimensional plot of twin-screw granulator dynamics (Figure 4a) and a gauge showing purely KPIs for the spray dryer (Figure 4c).

Analytics: The software also acts as a no-code platform for low-level machine learning. This means users can construct, train, and store their own machine learning algorithms with a view to applying these to live data. So far, unsupervised (Principal Component Analysis (PCA) and outlier detection) and supervised (classification) algorithms have been used successfully to analyse uploaded datasets. As well as storing the models, the insights from the models created can be displayed intuitively across the dashboard environment allowing for fast and effective decision making from machine learning - all without ever having to write a line of code. This distillation of complex data analysis methods into a familiar graphical interface is where this software holds its value in industry.

4. Conclusion

Here, we present Wiz 4.0, a data analytics/visualisation platform with a rich variety of features including real-time high-dimensional scatter plots, conducive to displaying relationships in multivariate data common in manufacturing. In addition, analysis tools including configurable ML algorithms have been added to give users access to additional insight and importantly these have been unified with the dashboard and plotting features.

Wiz 4.0 forms part of a more general trend in industry towards data-dashboards as a way to effectively communicate the larger amounts of data that are being used at all decision-making levels in real Industry 4.0 settings. Due to the software's modular dashboard interface, information pertinent to all levels of an organization can be collected, for example for the creation of a management level dashboard. This could provide less detailed information about the equipment but could still leverage important data collected by Siemens MindSphere to inform managers of important statistics such as operational

run time, energy consumption, efficiency, and product output. In this way, Wiz 4.0 retains its power for a multitude of scenarios and users. More importantly, the implementation of ML algorithms enables key intelligent decision-making technologies such as fault detection and predictive maintenance: crucial capabilities for a truly successful smart manufacturing.

Acknowledgements:

We thank Wellcome Leap and the University of Sheffield for funding the project. We also thank Perceptive Engineering and Siemens for useful discussions.

References

C. Balzer, R. Oktavian, M. Zandi, D. Fairen-Jiminez, P. Moghadam, 2020, Wiz: A Web-Based Tool for Interactive Visualization of Big Data, Patterns, 1, 8, 100-107

M. L. Oppert, V. O'Keeffe, 2019, The future of the ageing workforce in engineering: relics or resources?, Australian Journal of Multidisciplinary Engineering, 15, 1, 100-111

F. Pedregosa, G. Varoquax, A. Gramfort, V. Michel, B. Thirion, M. Blondel, 2011, SciKit-learn Machine learning in Python, the Journal of machine Learning research, 12, 2825-2830

A.Raj, G. Dwivedi, A. Sharma, A. B. Lopes de Sousa Jabbour, S. Rajak, 2020, Barriers to the adoption of industry 4.0 technologies in the manufacturing sector: An inter-country comparative perspective, International Journal of Production Economics, 224

P. Wellener, S. Shepley, B. Dollar, S. Lapper, H. Manolian, 2019, 2019 Deloitte and MAPI Smart Factory Study, Research Centre for Energy & Industrials Group, Deloitte.

M. Zandi, 2019, Preparing students for 21st Century Chemical Industry, Perspectives from academia and industry for dealing with a digital world, University College London, UK

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50292-X

About data reduction techniques and the role of outliers for complex energy systems

Luise Middelhauve[a], François Maréchal[a]

[a] *École polytechnique fédérale de Lausanne, Rue de l'Industrie 17, 1951 Sion, Switzerland, luise.middelhauve@epfl.ch*

Abstract

Optimal design and scheduling of energy systems with a high share of renewables is a complex and computationally demanding task. The mismatch of supply and demand of energy requires the consideration of timeseries with a granularity of a few minutes, which is in contrast to the lifetime of the system of multiple decades. This paper proposes an algorithm for systematically reducing the input data and computational effort in mixed integer linear programming (MILP) of energy systems. Unlike the state- of- the-art, the influence of different numbers of typical periods is not examined on the on the quality of the clustering algorithm but on the objective function and the integer decisions. The issue is addressed by exploiting the two-stage nature of the optimal design and planning of the system by sequentially performing k-medoids clustering. The demonstration of the proposed algorithm shows that very few typical periods are sufficient to achieve near optimal decisions. The proposed approach is outperforming algorithms for time series aggregation (TSA) in this field by reducing CPU time by more than 40 %. The inclusion of the integer decision in the algorithm allows the application to multi objective optimization (MOO). The case study demonstrates that the runtime of the MOO can be reduced by approximately 90 %, while diverting less than 2 % on Pareto optimal solutions. Outliers have no impact on the techno-economic analysis but may lead to significant electricity peaks in energy systems with a high share of renewables.

Keywords: Energy system design, Renewable energies, Mixed integer linear programming, Data reduction, k medoids clustering, outliers

1. Introduction

Daily and seasonal cycles lead to reoccurring patterns in the supply and demand of energy. Hence, it is popular to aggregate yearly time series to typical periods, in order to reduce computational effort for optimization problems of energy systems. Hoffman et al. (2020) have reviewed Time Series Aggregation (TSA) methods for modelling energy systems applied in 130 different publications. The authors conclude that for the same computational time, the more intuitive aggregation of seasons or months result in insignificantly larger errors than machine learning techniques. TSA methods of latter category performed similarly well, although k-medoids were most reliable for approximating costs.

In current TSA methods, the identification of the optimal number of typical periods is the first challenge. State-of-the-art approaches in modelling complex energy systems almost exclusively base their decision on Key Performance Indicators (KPI) of the clustering algorithm itself. The performance of the intended application to the energy system optimization is not considered in the selection process. Schütz et al. (2018) have demonstrated that the KPI *sum of square error*, typically applied when evaluating the

clustering is not suitable for the application in energy system optimization problems. Therefore, Bahl (2018) developed a systematic method, which evaluates the error in the objective function of the energy system optimization for different TSA aggregation lengths. The impact of the unit decisions taken in the optimization problem is not considered in their approach.
The previous paragraph allows the conclusion that machine learning algorithms for data reduction techniques are not very well integrated in the optimization technique of complex energy systems. Thus, this work presents an iterative method for the TSA for mixed integer linear programming (MILP) of energy systems.

2. Method

An overview of the usual steps involved in TSA is provided in Figure 1. The steps, which are modified in comparison to the current state-of-the-art are highlighted.

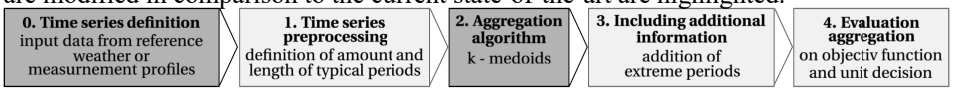


Figure 1 Procedure of time series aggregation (TSA) with relevant pictures in light grey.

Schütz et al. (2018) as well as Hoffman et al. (2020) have identified that the k- medoids algorithm is most reliable for techno- economic evaluation of energy systems. Hence, k-medoids algorithm is chosen as aggregation method in step 2.
Another challenge, which is generally overlooked in TSA for complex energy systems is the role of extreme periods. These extreme periods serve as protection or guarantee that the energy system can still provide required services, even in these extreme situations (Kotzur et al. (2018)). Extreme periods are often given by national regulations, are added in a postprocessing step (Stadler et al. (2018)) and are not further analysed. This work additionally investigates the role of outliers, which serve as extreme situations.
In the following, the aggregated problem (AP) is the energy system optimization based on typical periods, whereas the operating problem (OP) is the optimization of only the operation, with fixed unit decisions, on a full timeseries.

2.1. Time Series Aggregation

The first part of the TSA (Figure 2) is an iterative process, testing two convergence criteria on the objective function. After the first convergence criterion $|\varepsilon_a| \geq \frac{AP_n - AP_{n-1}}{AP_n}$ is valid, the OP is solved. The second convergence criterion $|\varepsilon_b| \geq \frac{OP_n - AP_n}{AP_n}$ is evaluated. In case both criteria are true, the sizing variables of the unit design are added to the solution space.

The second part is initiated with the first element, characterized by n clusters in the solution space. Additionally, the AP is executed with n+m steps. This way, different energy system designs with similar value of the objective function are identified and added to the solution space. If the unit decisions remain unchanged, the TSA is terminated with n typical periods.

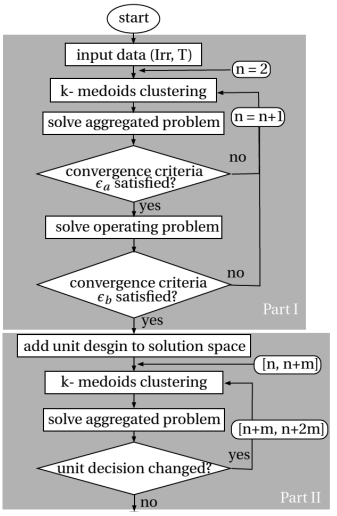


Figure 2 Overview of the iterative timeseries aggregation algorithm.

2.2. Outlier detection

Similar to the method developed by Liu et al. (2021), outliers are detected during the process of clustering. The procedure is as follows:

1. Clustering data set
2. Calculation of Euclidian distance from all points to their centroid
3. Removal of *o* outlier periods with the largest distance from data set
4. Repetition of steps 1-3 until centroids do not change anymore.

The detected *o* outliers are added as individually occurring typical period to the optimization problem. In contrast, demand peaks remain as extreme period. As current standard in optimization of energy systems, one cold weather and one hot weather period is added, both consisting of one single timestep (Stadler et al., (2018)).

3. Case study

The data reduction technique is demonstrated on optimizing the energy system of a typical residential building located in the climatic zone of Geneva, Switzerland. The building is a single- family home with 2 floors and in total 250 m^2 heated surface from around 1980. The considered energy demands are: electricity demand, thermal demand for space heating and hot water. The optimization is formulated as a MILP problem with the aim to find the optimal sizing and operation among nine energy conversion and storage technologies. For further insights on the modelling approach of the building energy system, the reader is kindly referred to Middelhauve et. al (2021). On a full timeseries of one year, this case study leads to over 840 thousand constrains and 790 thousand variables, among which are almost 10'000 binaries. The tuning parameter of the algorithm are set to be $\varepsilon_a = 5\%,\ \varepsilon_b = 5\%$ and $m = 3$.
The k-medoids clustering with the R package *wcKMedoids* is performed for aggregating one typical year (DOE, (2020)). The problem is formulated in AMPL Version 20191001 and solved with CPLEX 12.9.0.0 on a local machine with following processor details: Intel(R) Core (TM) i7-8559U CPU @ 2.70GHz. The relative tolerance between relaxed linear problem and best integer solution is set to *mipgap*=5e-7. The remaining CPLEX settings are equal to the default settings reported at (CPLEX, (2020)).

4. Results and Discussion

In a pre-processing step, global irradiation and the ambient temperature are clustered to increasing number of k-medoids, each typical period is chosen to be the length of one full day. The state-of-the-art procedure, which bases the selection of optimal number of periods at KPI slope thresholds, would lead to around 10 typical periods. Thus, the result of the proposed TSA is compared to 10 typical periods in the following.

4.1. Time Series Aggregation

First, the AP with an increasing number of clusters is solved. The objective function is total expenses, which is equally weighting two conflicting objectives, the capital expenses (CAPEX) and operational expenses (OPEX). Part one of the proposed TSA algorithm is demonstrated in Figure 3. After already three clusters, the first convergence is approximately 5%. The unit decisions are fixed and the OP is solved. The second convergence criterion is also below 5 % and therefore three typical periods are chosen as result of Part 1. Both problems the AP and the OP are solved for up to 12 typical days for demonstration purposes. The difference of the objective functions, which is the second convergence criterion, remains below 1% after 7 typical periods.

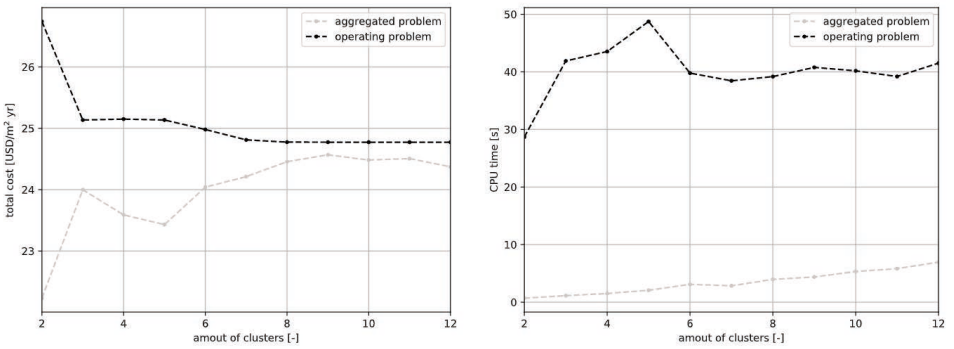


Figure 3 Demonstration the Part 1 of the proposed TSA algorithm. Impact of different numbers of clusters on a) the objective function, b) the runtime.

The aim of Part 2 of the algorithm is to confirm that the unit decisions are sufficiently taken with the detected number of typical periods (compare Figure 2). For two clusters, an air-water heat pump in combination with electrical heater and thermal storage tanks is detected as most economical decision. For three clusters, photovoltaic modules are chosen additionally. This configuration remains unchanged for not only n+m = 6 clusters, but for all 12 clusters examined. Therefore, the three typical periods can be confirmed and further used during Multi objective optimization (MOO).

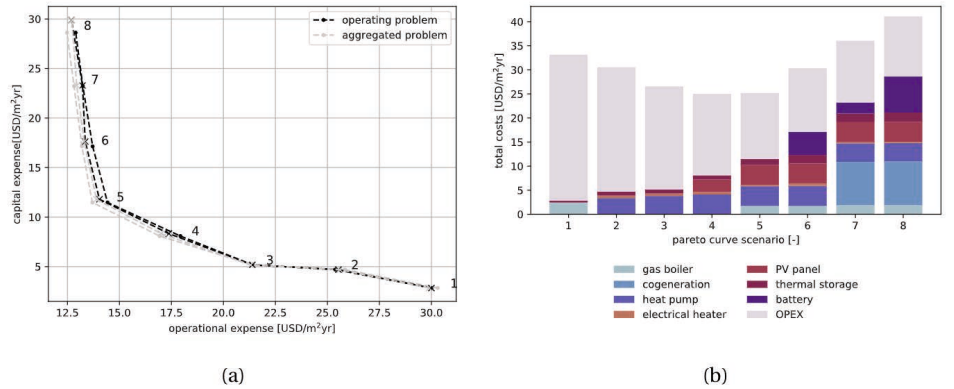


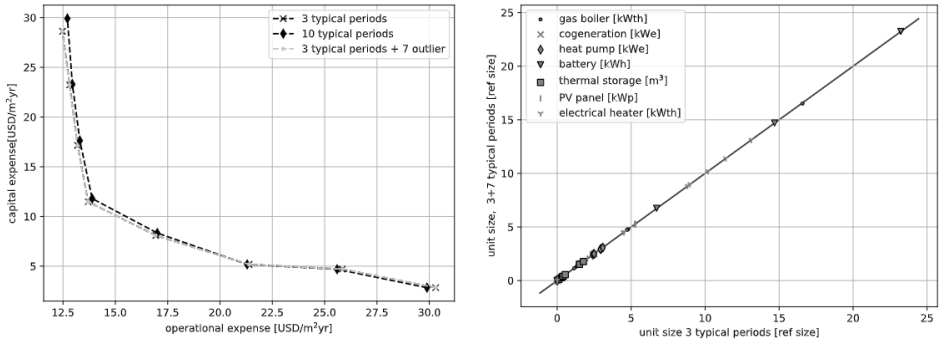
Figure 4 Results of multi objective operation a) Pareto frontier with 10 (x) and 3 (•) typical periods. b) unit decisions along Pareto frontier.

One requirement of appropriate TSA in energy system optimization is that the aggregation is valid also in MOO problems. Therefore, Figure 4 compares the Pareto curves of the AP to the associated OP for the previously resulting 3 typical periods and compares them to the state-of the-art choice of 10 typical periods. The unit decision of Scenario 1 is based on a natural gas boiler and small thermal storage tanks. In this case, both APs as well as their linked OPs result in almost identical points on the Pareto frontier. The energy system is more diverse in higher investment scenarios. It includes renewable energy sources such as PV panels, an air-water heat pump and thermal and electrical storage in Scenario 8. The performance of this energy system is more depending on the weather data or, in terms of the storage, the optimal scheduling within one typical period.

Nevertheless, comparing the OPEX of the AP to the related problem on the fulltime series, the OP shows a deviation by only 3%. The difference among the AP with 3 typical periods and with 10 typical periods is less than 2%, maximum occurring in higher investment scenarios.

4.2. Outlier detection

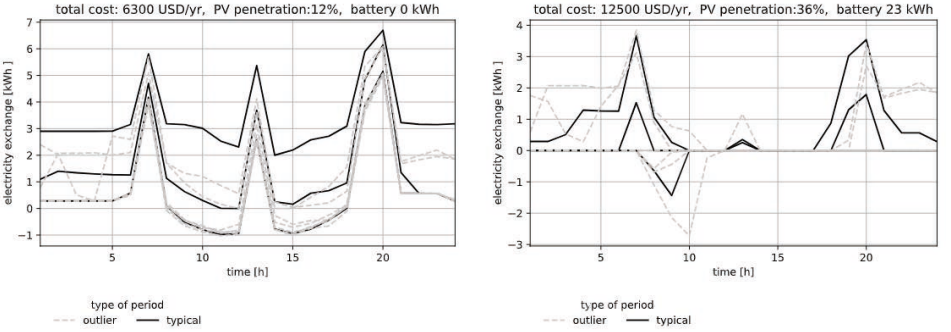
K-medoids clustering is a method, which is itself robust to outliers. Therefore, the centroid is not changing during the detection of outliers. Nevertheless, seven outliers have been identified and added to the typical periods. Figure 5 demonstrates the impact of outliers on the MOO of the energy system. Along the pareto curve, identical unit decisions are taken and the pareto curve with and without 7 outliers are identical.



(a) (b)

Figure 5 Comparison of multi objective optimization with and without detected outliers. a) Pareto curve b) Parity plot of all unit decisions along the pareto curve.

In contrast to the detected scenarios along the pareto curve, Figure 6 shows an important impact of outliers.



(a) (b)

Figure 6 Electricity exchange for energy system designs with a) a lower investment and b) higher investment and greater share of renewable energy integration

Outliers may not be relevant for the thermal energy side or the unit decisions. But in future energy systems with a high share renewables, outliers reveal electricity peaks of the grid exchange, which are higher than during typical periods.

5. Conclusion

This work proposes a novel method for TSA of complex energy systems. Compared to state-of-the-art approaches in this field the CPU time was reduced by 40% in the presented case study. This is achieved by using two convergence criteria, which avoids the computationally intensive computation of the OP at each iteration step. Additional innovation is to take unit decisions into account, which allows the application of the TSA method to MOO problems.
In contrast to comparable work in MOO of energy systems, the selection of the appropriate number of typical days is not based on KPIs evaluating the underlying machine learning algorithm. Hence, presented TSA method allows to significantly reduce the runtime by more than 90%, while diverting less than 2% on optimal solutions. The impact of the TSA is greater, the more renewable energies are included in the system. For systems with a high share of renewables, outliers reveal electrical peaks, which are greater than during typical periods. Outliers are however neither impacting the thermal energy side nor economic evaluation of presented MOO problems. One possible extension of this work is to analyse the impact of the tuning parameter of the proposed TSA method. Additionally, the usage of one typical year to represent a project horizon can be questioned. This includes the challenge to predict changing weather data subject to climate change.

References

T. Schütz, M.H. Schraven, M. Fuchs, P. Remmen, and D. Müller, 2018, "Comparison of clustering algo-rithms for the selection of typical demand days for energy system synthesis", Renewable Energy vol.129, pp. 570–582.

P. Stadler, L. Girardin, A. Ashouri, and F. Maréchal, 2018, "Contribution of Model Predictive Controlin the Integration of Renewable Energy Sources within the Built Environment", Frontiers in Energy Research, vol. 6

M. Hoffmann, L. Kotzur, D. Stolten, and M Robinius, 2020, "A Review on Time Series Aggregation Methods for Energy System Models", :Energies,vol. 13.3, p. 641

L. Middelhauve, L. Girardin, F. Baldi, and F. Maréchal. 2021 "Potential of Photovoltaic Panels on Building Envelopes for Decentralized District Energy Systems", Frontiers in Energy Research p. 689-781

L. Kotzur, P Markewitz, M. Robinius, and D. Stolten. 2018, "Impact of different time series aggregation methods on optimal energy system design", Renewable Energy, vol. 117, pp. 474–487

H. Liu, J. Li, Y. Wu, and Y.Fu, 2021 "Clustering with Outlier Removal", IEEE Transactions on Knowledge and Data Engineering, vol. 33

DOE. Department of Energy, BTO. Building Technologies Office, and NREL. National Renewable Energy Laboratory, Weather Data, EnergyPlus, https://energyplus.net/weather accessed Mar. 2020

CPLEX Options for AMPL. https://ampl.com/products/solvers/solvers-we-sell/cplex/options/, accessed Oct. 2021.

B. Bahl, 2018, "Optimization-Based Synthesis of Large-Scale Energy Systems by Time-Series Aggregation" ; RWTH Aachen University. PhD thesis, 148 pages

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50293-1

DeepGSA: Plant Data-Driven Global Sensitivity Analysis using Deep Learning

Adem R.N. Aouichaoui[a], Resul Al[b], Gürkan Sin[a*]

[a]*Process and Systems Engineering Center (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Kgs.Lyngby, DK-2800, Denmark*
[b]*Novo Nordisk A/S, Bagsvaerd, DK-2880, Denmark*
gsi@kt.dtu.dk

Abstract

Data-driven modeling provides a viable alternative for process modeling especially in applications where mechanistic modeling falls short of explaining the underlying phenomena. The increasing amount of plant data collected through various sensors and lab tests lays the foundation for various data-driven modeling approaches such as Deep Neural Networks (DNN). In this work, we present a new software tool, named deepGSA, incorporating well-established variance-decomposition and derivative-based global sensitivity analysis (GSA) methods, such as Sobol sensitivity indices, with the plant data-driven deep learning modeling techniques. The deepGSA aims at enabling non-specialist practitioners to leverage DL-based models for GSA application purposes. The tool is successfully applied on a benchmark case study as well as the case of modeling liquid nitrous oxide concentration in a wastewater treatment plant to highlight its capabilities. The deepGSA toolbox, documentation, installation guide, and several examples are freely available on GitHub through the link: https://github.com/gsi-lab/deepGSA.

Keywords: Deep-Learning, Big Data, Sensitivity Analysis, Wastewater Treatment

1. Introduction

Digitalization has become an increasingly hot topic in various fields across academia and the chemical industry issuing a new industrial revolution under the guise of "industry 4.0". At the core of this digital transformation, we find Digital Twins, which are virtual mock models capable of simulating/describing real-life physical processes (de Beer and Depew, 2021). These models often rely on first principle mechanistic models particularly in cases where the phenomena occurring are well understood e.g. distillation and extraction processes. However, this process understanding is not as mature in many life science applications especially if it combines a wide range of different processes e.g. chemical, biological, and mechanical processes. A prime example of this is the modeling wastewater treatment plants (WWTP) (Sin and Al, 2021). Mathematical models have been applied to model the microbial activity in WWTP in the form of Activated Sludge Models (ASMs)(Sin and Al, 2021). The original ASM models have been modified and expanded in the past decades to include additional microbial conversion processes such as single-pathway and two-pathway models for N_2O production (Sin and Al, 2021). However, these efforts resulted in very complex mathematical models that are difficult to apply in practice due to the high number of model parameters that need to be calibrated (Chen *et al.*, 2019). Data-driven modeling of industrial processes provides an alternative approach to construct digital twins by leveraging the vast amount of plant data collected

through sensors to some extent considered "Big data". Data-driven models such as Deep Neural Networks (DNNs) although providing good predictive performance across many fields and applications (Hwangbo *et al.*, 2021), lack the aspect of transparency and interpretability due to their black-box nature. Global Sensitivity Analysis (GSA) could potentially help elucidate the workings of these models. In this work, we present a Matlab-based toolbox, deepGSA, with the aim of streamlining data-driven modelling to non-specialist practitioners and enabling them to leverage their plant data to gain process understanding through various GSA methods implemented in the toolbox.

2. Methods

Global sensitivity analysis aims at quantifying the effect of an independent variable while all other variables are also varied. In doing so, the GSA takes into account interactions among the variables (in contrast to One-At-a-Time methods) and does not depend on the choice of a nominal point. Two GSA techniques are integrated into the tool: the Sobol method and the derivative-based global sensitivity method (DGSM).

2.1. Sobol Method

Sobol's method is based on variance decomposition of Monte-Carlo simulations. The method provides both first order (S_i) and total order indices (S_{Ti}) for the input parameter x_i. S_i measures the individual contribution of input x_i to the total output variance, while S_{Ti} measures the total contribution to the total output variance including those resulting from interactions with other inputs (Saltelli *et al.*, 2007). The expressions for S_i and S_{Ti} are shown in Eq.(1), where $V[E(y|x_i)]$, $V[E(y|x_{-i})]$ and $V(y)$ are the conditional variance, the conditional variance derived from all variables but x_i varied, and the unconditional total output variance respectively, all of which are numerically estimated through Monte-Carlo simulations. The difference between the two indices (S_{Ti}-S_i) is a direct indicator of the strength of interactions variable x_i is involved in.

$$S_i = \frac{V[E(y|x_i)]}{V(y)} \qquad S_{T_i} = 1 - \frac{V[E(y|x_{-i})]}{V(y)} \tag{1}$$

2.2. Derivative-Based Global Sensitivity Method (DGSM)

DGSM uses the second moment of the model derivatives to measure the importance of the input parameter x_i as formulated in Eq.(2) (Kucherenko *et al.*, 2009). The aim of this is to avoid canceling off negative and positive impacts of the local sensitivity (E_i) over the entire input domain H_n.

$$v_i = E\left[\left(\frac{\partial f(X)}{\partial x_i}\right)^2\right] = \int_{H_n} \left(\frac{\partial f}{\partial x_i}\right)^2 dx \text{ with } \mathrm{E_{i,n}} \frac{\partial f}{\partial x_i} \tag{2}$$

A mean measure for the global sensitivity can then be obtained by Eq.(3).

$$\mu_i = \sqrt{v_i} \tag{3}$$

The mean measure is often used to screen unimportant factors in the input domain for which μ_i is very small or negligible. Additionally, the method gives insights into the directional pull of the input factor (positive or negative) by capturing the local sensitivity.

Although empirical, there is a link between the Sobol S_{Ti} and v_i as expressed in Eq.(4), where D is the total variance. The relation can be interpreted as the upper bound of S_{Ti}.

$$S_{T_i} \leq \frac{v_i}{\pi^2 D} = S_{Ti}^{DGSM} \tag{4}$$

3. DeepGSA: Toolbox for plant data-driven and DL-assisted GSA

Data-driven modeling

Given a large dataset X representing inputs and the corresponding response variable vector y, deepGSA is able to develop a DNN model to predict y from the supplied information X. The DNN model hyperparameters are optimized using a grid-search tuning policy for which a user can supply a pool of architectures, a set of training functions, and a set of activation functions. Each possible combination is trained on 80% of the supplied data, while the remaining data are equally split between validation and test purposes to avoid data leakage and overfitting. The toolbox also leverages Matlab's parallel computation using high-performance GPUs and computer clusters (if present) to speed up model development and GSA calculations. The optimal model can be selected based on various criteria evaluated on the validation set. The user can choose between: the coefficient of determination (R^2), the mean squared error (MSE), the root mean squared error (RMSE), the Bayesian/Akaike information criterion (BIC/AIC).

Inference of input distributions

Sampling from the input data is an important step to perform data-driven GSA. In some cases, the distributions of the input factors are known a priori, thus can be readily supplied to the toolbox. However, in many engineering applications, the underlying distribution of X is unknown. By supplying X, the toolbox can also infer the distributions of variables by fitting a kernel distribution to the data, which then allows sampling from a distribution similar to the input factor distributions to preserve the characteristics of the system.

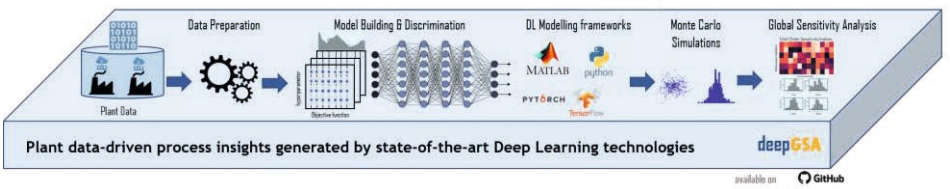


Figure 1: deepGSA Framework

4. Case Study

4.1. Benchmark Case Study: Sobol g-function

The g-function of Sobol, as given in Eq.(2), is often used as a benchmark function for numerical experiments in sensitivity analysis literature. The reason being the theoretical first-order Sobol indices are analytically available (Marrel *et al.*, 2009).

$$f(x) = \prod_{i=1}^{d} \frac{|4x_i - 2| + a_i}{1 + a_i}, \text{with } a_i = \frac{i-2}{2} \text{ and } x_i \sim U(0,1) \text{ for all } i = 1,..,d \quad (2)$$

Through Monte-Carlo simulations, a sample of size N = 5,000 is generated and used to evaluate the corresponding f(x) values following Eq.(2) and perform GSA. Figure 2 shows the sensitivities indices obtained: analytically, by evaluating the g-function, and by constructing DNN. All methods produce the same sensitivity indices which validate not only the implementation of the GSA techniques but also that DNN models retain the dynamics of the process it models. The obtained DNN was trained with the Bayesian regularization backpropagation algorithm in MATLAB and had the layers [32,16,8,4] with hyperbolic tangent sigmoid transfer function.

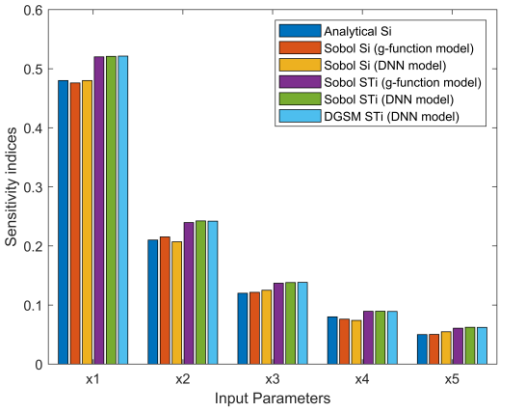


Figure 2: Sensitivity indices for the Sobol g-function estimated through various methods

4.2. Plant Data Case Study: Modelling liquid N_2O concentration in WWTP

DeepGSA is applied on a second case study using various sensor data collected from Avedøre WWTP located in Copenhagen Denmark to predict the liquid Nitrous oxide concentration (N_2O). The data and the plant are described in detail by (Hwangbo *et al.*, 2020, 2021). In this case study, a sample of the data spanning over 90 days starting on March 26^{th,} 2018, and ending on June 24^{th,} 2019 was selected. The input factors chosen are the dissolved oxygen (DO), ammonium concentration (NH_4), nitrate concentration (NO_3), temperature (T), the airflow for aeration (Q_{air}), and the influent flow (Q_{inf}). A total of 25,500 data points were used and all inputs were smoothed using the moving median method with a moving window of 12 timesteps and scaled between 0 and 1. The grid search included the following architectures: [128,64,32,16,8,4], [64,32,16,8,4], [32,16,8,4] with two possible training functions: '*trainlm*' and '*trainbr*' and two possible transfer functions: the log-sigmoid and positive linear functions. The selection criteria chosen is the R^2, the distributions of the input factors are inferred by the tool and a sample of size N=100,000 was used to conduct GSA. The sampling number was set very high since the computational solution of DNN is very affordable. The output layer activation function adapts to the characteristics of the target output e.i. constrained between [0, inf] through the positive linear (poslin) activation function. The selection criteria in this case study were chosen to be the R^2 metric. The best DNN achieved R^2 of 0.81 with a hidden configuration of [12,8,64,32,16,8,4] trained with 'trainbr' with poslin activation function.

Table 1: Inference statistics and sensitivity measures related to the DGSM and Sobol methods

	DGSM			Sobol	
Variable	μ (E)	σ (E)	S_{Ti}^{DGSM}	S_i	S_{Ti}
Q_{inf}	00.2	63.2	2.121	0.0000	0.3502
DO	00.1	46.9	1.169	0.0083	0.3238
NH_4	15.3	96.1	5.018	0.0317	0.4553
NO_3	11.4	103.2	5.718	0.1087	0.5848
T	01.1	109.2	6.332	0.0329	0.6579
Q_{air}	04.4	79.9	3.398	0.0529	0.5821

Table 1 provides various inferred statistics related to the DGSM and Sobol. The Sobol method identifies T, the NO_3, NH_4 concentrations as well as the influent and air flowrates as the most important factors. This is in line with the results obtained by (Hwangbo *et al.*, 2021). There is a large difference between S_i and S_{Ti}, with the Si values being very close to zero and the sum of S_{Ti} exceeding unity. These are all evidence of the presence of strong interaction effects of the input factors on the response variable (N_2O liquid concentration). These observations are consistent with the process insight of the N_2O dynamics in a WWTP and are explained in detail by (Hwangbo *et al.*, 2021). The S_{Ti}^{DGSM} is qualitatively valid for all input factors. Furthermore, the ranking is almost identical to the one obtained by Sobol indices except for two factors for which the ranking is swapped (Q_{air} and NH_4). This is in line with observations by (Sobol and Kucherenko, 2009) that for highly non-linear systems the ranking might not hold, and instead, the Sobol indices are more reliable. The statistics of the local sensitivity reported in Table 1 and the visualization of the empirical cumulative distribution function for selected input factors in Figure 3 suggest that the factors influence the output both positively and negatively. This is in fact due to data coming from a closed-loop (controlled) plant and it shows the effects of the inputs are nonlinear and therefore the effects depend on relative values of other inputs (input space has both regions where a factor produces a negative impact and positive impacts). Figure 3 further shows that the distribution exhibits some heavy-tailed distribution known as "fat tails", making reporting the mean and standard deviation misleading.

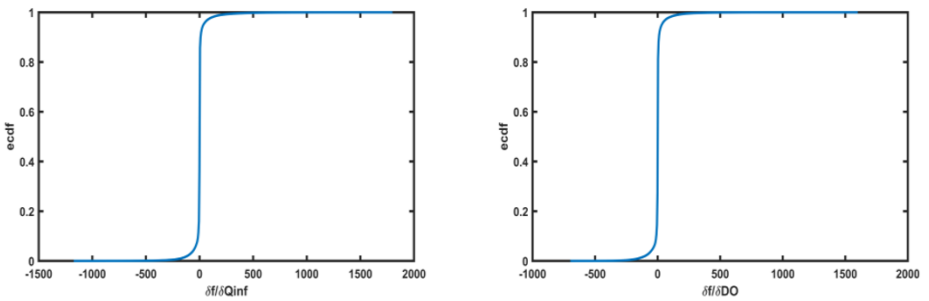


Figure 3: Empirical Cumulative distribution function of the local sensitivities of T and DO revealing a heavy-tailed distribution. In the tails, one observes a small change input has a disproportionately high effect on the output which is N_2O emissions.

5. Conclusions & Future work

DeepGSA provides an easy-to-use, plug and play tool to perform data-driven modeling using deep neural networks and perform global sensitivity analysis using both variance decomposition-based and derivative-based sensitivity analysis methods. The tool constructs a search space to tune the hyperparameters of the neural network and provide various metrics to select the optimal network. The GSA performed on the model is then used to provide process insights that conform with domain knowledge for two cases studies presented in this work. Future improvements to the toolbox include implementing more GSA techniques (Morris screening, standard regression coefficient (SRC), and GSA for dependent inputs) and the addition of new features such as an internal plotting tool, data cleaning, and scaling options. Finally, a wrapper function will be developed to allow users to plug in DNN models developed in open source DL frameworks such as TensorFlow and Pytorch.

References

de Beer, J. and Depew, C. (2021) 'The role of process engineering in the digital transformation', *Computers and Chemical Engineering*, 154, p. 107423.

Chen, X., Mielczarek, A.T., Habicht, K., Andersen, M.H., Thornberg, D. and Sin, G. (2019) 'Assessment of Full-Scale N 2 O Emission Characteristics and Testing of Control Concepts in an Activated Sludge Wastewater Treatment Plant with Alternating Aerobic and Anoxic Phases', *Environmental Science & Technology*, 53(21), pp. 12485–12494.

Hwangbo, S., Al, R., Chen, X. and Sin, G. (2021) 'Integrated Model for Understanding N2O Emissions from Wastewater Treatment Plants: A Deep Learning Approach', *Environmental Science and Technology*, 55(3), pp. 2143–2151.

Hwangbo, S., Al, R. and Sin, G. (2020) 'An integrated framework for plant data-driven process modeling using deep-learning with Monte-Carlo simulations', *Computers & Chemical Engineering*, 143, p. 107071.

Kucherenko, S., Rodriguez-Fernandez, M., Pantelides, C. and Shah, N. (2009) 'Monte Carlo evaluation of derivative-based global sensitivity measures', *Reliability Engineering and System Safety*, 94(7), pp. 1135–1148.

Marrel, A., Iooss, B., Laurent, B. and Roustant, O. (2009) 'Calculations of Sobol indices for the Gaussian process metamodel', *Reliability Engineering and System Safety*, 94(3), pp. 742–751.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. and Tarantola, S. (2007) *Global Sensitivity Analysis. The Primer*. Chichester, UK: John Wiley & Sons, Ltd.

Sin, G. and Al, R. (2021) 'Activated sludge models at the crossroad of artificial intelligence—A perspective on advancing process modeling', *npj Clean Water*, 4(1), p. 16.

Sobol, I.M. and Kucherenko, S. (2009) 'Derivative based global sensitivity measures and their link with global sensitivity indices', *Mathematics and Computers in Simulation*, 79(10), pp. 3009–3017.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50294-3

Analysing Different Dynamically Modelled Data Structures and Machine Learning Algorithms to Predict PM2.5 Concentration in China

Danny Hartanto Djarum[a], Nur Hidanah Anuar[a], Zainal Ahmad[a*], and Jie Zhang[b]

[a] *School of Chemical Engineering, Universiti Sains Malaysia, Engineering Campus, 14300, Nibong Tebal, Penang, Malaysia.*
[b] *School of Engineering, Merz Court, Newcastle University, Newcastle upon Tyne NE1 7RU, United Kingdom.*
chzahmad@usm.my

Abstract

Harmful air pollutant such as $PM_{2.5}$ is still a major concern in many countries. At high concentrations, it could lead to adverse health effect on human, escalating the risk of cardiovascular and respiratory diseases. In order to mitigate this issue, continuous air quality monitoring systems have been deployed to alert the general public of high $PM_{2.5}$ level. However, such monitoring system requires substantial budget and resources to construct, thus may not be accessible in some regions especially developing countries. Therefore, it is important to develop a high performance $PM_{2.5}$ prediction model that only employs easily attainable input parameters as a more cost-effective alternative. In this study, common meteorological data from five different cities in China were utilized for the $PM_{2.5}$ prediction model. Dynamic model such as the nonlinear autoregressive network with exogenous inputs (NARX) with different input/output time lag were applied to transform training dataset into different data structures. Additionally, machine learning algorithms were analysed and evaluated to predict $PM_{2.5}$, namely: multi-linear regression (MLR), and feed-forward artificial neural network (FANN). The results shows that FANN model with 10 hidden neurons using NARX-2 data structure is the best model combination with an R^2 values of up to 0.973.

Keywords: $PM_{2.5}$, NARX, MLR, FANN.

1. Introduction

The world population has grown significantly in the past decades, as a result, this causes the rise in rapid industrialization and urbanization to cope with the high demand of food and energy resources (Mobaseri et al., 2021). The excessive rate of industrialization consequently translates to generation of pollutants leading to extreme deterioration in global air quality. $PM_{2.5}$ which stands for particulate matter 2.5, is considered as one of the major and most dangerous atmospheric pollutants in the air. Any airborne particle that has an aerodynamic diameter of less than 2.5 μm is categorized as $PM_{2.5}$. Various researchers have shown in their study that $PM_{2.5}$ is associated with serious health diseases such as strokes, chronic heart disease, lung cancer, respiratory infections, and premature death (Gakidou et al., 2017). In a 2019 study, Chen et al. (2019) discovered that an increase of 10 $\mu g/m^3$ in atmospheric $PM_{2.5}$ levels would lead to a reduction in adult life expectancy by approximately 0.8 years.

With the growing concern over the dangerous health hazard poses by $PM_{2.5}$, continuous $PM_{2.5}$ monitoring stations have been proposed by many to alert the general public. However due to the high initial investment required, such system is only available in specified locations. In fact, the world health organization (WHO) reported that approximately 9 out of 10 people in the world still resides in area with $PM_{2.5}$ level exceeding the WHO air quality guidelines (WHO, 2021). In search for a more cost-effective alternative, numerous studies have been carried out to develop prediction models that can accurately reflect $PM_{2.5}$ concentrations. A feed forward artificial neural network (FANN) model was developed by (Perez et al., 2020) to predict $PM_{2.5}$ concentrations in Coyhaique, Chile. Their results shows that the model was able to capture the instance when $PM_{2.5}$ falls under critical region 85 % of the time. (Biancofiore et al., 2017) compared three different models to predict the levels of $PM_{2.5}$ for multiple days ahead (recursive neural network, feed forward neural network, and multiple linear regression). They concluded that recursive neural network outperforms the other models with correlations coefficients up to 0.89.

In recent years, artificial neural network model seems to be of the up most interest for $PM_{2.5}$ forecasting. However, the number of research that analysed the effect of time series data structure on the ANN model seems pale in comparison. Therefore, in this study, a dynamic time series model with different input/output time lag were utilized to transform the data structure, namely nonlinear autoregressive network with exogenous inputs (NARX). Additionally, a feed forward artificial neural network (FANN) model utilizing different form of input data structure were analysed to predict $PM_{2.5}$ concentration. Furthermore, the resulting performance were compared against widely popular multi linear regression (MLR) model that utilized untransformed input data.

2. Methodology

2.1. Dataset

The data used for this study is an hourly air quality and meteorological measurements recorded for six years between 2010 and 2015 in five Chinese cities, which are Beijing, Chengdu, Guangzhou, Shanghai, and Shenyang. These data were obtained from UCL Machine Learning Repository published by (Liang et al., 2016). There are 52,584 rows of data and with a total of 10 parameters this translates to about 525,840 individual data points for each city. The list of the input and output parameters for the $PM_{2.5}$ prediction model could be found in Table 1.

Table 1: List of input and output features for $PM_{2.5}$ prediction model.

List of input parameters	Seasons, temperature, pressure, dew point, humidity, combined wind direction, cumulated wind speed, hourly precipitation, cumulated precipitation
List of output parameters	$PM_{2.5}$ concentration

2.2. Data Pre-processing

Multiple pre-pre-processing techniques were performed on the original dataset before transformation on the data structure and development of the prediction model. In the missing data analysis, it was found that more than half of the dataset comprised of missing values. Due to the high degree of missing data, median value substitution could not be carried out as it will generate unwanted bias in the training data. Therefore, the missing data were instead discarded from the dataset, with approximately 20,000 rows of data remaining for each city. Normalization was then performed on the data, since the dataset has parameters with highly varying magnitude which similarly could lead to bias when training the model. Lastly, label encoding was performed to convert the parameter with categorical value such as wind direction into numeric value as most machine learning algorithms prefer to work with numerical attributes.

2.3. Data Structures

A dynamic time series model called nonlinear autoregressive network with exogenous inputs (NARX) was used to transform the training data set into different data structure. In this study, three type of data structures were analyzed, namely: The base structure without any transformation, NARX-1 (NARX with time lag of 1), and NARX-2 (NARX with time lag of 2). The complete equations for the different transformation on the data structure could be observed from Eq.(1) for base structure, Eq.(2) for NARX-1 and Eq.(3) for NARX-2.

$$y(t) = Fn(x_t) \tag{1}$$

$$y(t) = Fn(y_{t-1}, x_{t-1}) \tag{2}$$

$$y(t) = Fn(y_{t-1}, y_{t-2}, x_{t-1}, x_{t-2}) \tag{3}$$

Where $y(t)$ represent the output and $x(t)$ is the input of the model at time t. 1 and 2 is the maximum input and output time lag. Fn is the nonlinear function.

NARX is one of the more robust time series models that has exogeneous inputs, it has the ability to generate forecast of future output value based on the past input and output values with varying degree of time lag. In theory, by transforming the training data structure with NARX model, the learning rate of the resulting trained neural network model should be more effective and generate better prediction performance. Additionally, the neural networks model should converge to the desired solution much faster, leading to an overall better training efficiency and generalization capability (Lin et al., 1996). A two-step combined algorithm based on NARX neural network were compared by (Buevich et al., 2021) against other models to predict various greenhouse gases (CH_4 and H_2O) concentrations. They found that the algorithms coupled with NARX model consistently outperform other models in predicting the greenhouse gaseous concentrations with correlation coefficient up to 0.87 when validated against the test dataset. In another study, Mohebbi et al. (2019) analyzed NARX modelled artificial neural network model against static neural network model to predict carbon monoxide concentration in Shiraz city. Their results, shows that the NARX-ANN model significantly improve the model performance with an R^2 values of up to 0.72 compared

to that of static neural network model that has an R^2 value of only 0.31. In our study the feed forward artificial neural network (FANN) model will utilized the NARX transformed data as the input to the model with maximum input and output lag time of 1 and 2.

2.4. Prediction Model

Multiple linear regression (MLR) and Feed forward artificial neural network (FANN) model were analyzed and evaluated to predict the $PM_{2.5}$ concentration. MLR is one of the most popular machine learning regression technique due to its simplicity, fundamentally, it explains a linear relationship between multiple input features to predict a single output feature. Eq.(4) below describe the universal formula for multiple linear regression model:

$$y = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n + \varepsilon \tag{4}$$

Where y is the independent variable, β_i is the coefficient, X_i is the independent variable, and ε is the residual error between the real value and forecasted value. When performing MLR modelling it is assumed that the data is normally distributed, achieved linearity, without extreme values, and input parameters must be independent of each other.

FANN were then analyzed which theoretically has better capability to capture the dynamic nonlinear relationship between the input parameters. Being one of the earliest classes of artificial neural network, FANN has been widely popular amongst the scientific community especially since it is generally superior compared to traditional statistical technique such as multi regression and ensemble decision tree technique (Grivas and Chaloulakou, 2006). Generally, the FANN algorithm has multiple neurons grouped within three types of layers: Input layers, hidden layers, and output layers. The input layer consists of various neurons that receive information from the dataset, each neuron is connected to every single neuron in next layer (hidden layers) with its own associated weight. The weight determines the strength between the two connected neurons. Each neuron sums all of the information received and pass it on to the output layer where the output value will be computed based on predefined activation function or transfer function (Kim et al., 2010).

Levenberg-Marquardt algorithms was selected to train the network, as it converges faster to the desired solution compared to the more popular error backpropagation algorithm (EBP). Early stopping mechanism was also utilized, where training will be stopped immediately once the changes in validation errors no longer improve, thus avoiding overfitting and poor generalization capability. From early analysis, it was discovered that the best training parameter for the FANN model is 1 hidden layer and 10 neurons. Therefore, there are three type of FANN model that will analyzed, the base FANN model, FANN-NARX(1), and FANN-NARX(2)

The training dataset is randomly split into 70% training set, 15% test set, and 15% validation set. The performance of the model is compared and evaluated based on the coefficient of determination (R^2) values and root mean squared error (RMSE).

3. Results and Discussion

Meteorological data from five different cities in china were utilized to predict $PM_{2.5}$ concentration, where the performance of four different machine learning models were analysed, namely: MLR, FANN, FANN-NARX(1), and FANN-NARX(2). Initially, MLR and FANN model were trained using the original untransformed data structure. The performance of these base models in predicting the $PM_{2.5}$ concentration for the test data set could be observed from Table 2. Due to the slight difference in scale between dataset from different cities, the RMSE values may slightly varies.

Table 2: $PM_{2.5}$ prediction performance of base MLR and FANN models for test dataset.

Model	Beijing		Chengdu		Guangzhou		Shenyang		Shanghai	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
MLR	0.29	0.873	0.27	0.840	0.20	0.899	0.26	0.881	0.22	0.858
FANN	0.69	0.528	0.64	0.586	0.44	0.737	0.54	0.695	0.53	0.688

The results from Table 2 shows that the FANN model consistently outperform the MLR model with much higher R^2 values and lower RMSE values throughout all five cities. It could be observed that even the best performing MLR model (Beijing dataset) could only score an R^2 values of 0.29 which is about 58% lower than that of the best performing FANN model that scored an R^2 values of 0.69. The RMSE is also lower at only 0.528 compared to MLR model at 0.873. This shows that despite of only using the base untransformed data structure, FANN model is far superior to MLR model in predicting the $PM_{2.5}$ concentration. However, as described in earlier section, the performance of the FANN model could be further improved by training it with NARX transformed data. Table 3 below depicts the performance of FANN model coupled with NARX with time lag of 1 and 2 on test data.

Table 3: $PM_{2.5}$ prediction performance of FANN-NARX(1) and FANN-NARX(2)

Model	Beijing		Chengdu		Guangzhou		Shenyang		Shanghai	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
FANN-NARX(1)	0.96	0.084	0.97	0.047	0.93	0.148	0.96	0.069	0.97	0.406
FANN-NARX(2)	0.96	0.080	0.97	0.044	0.94	0.136	0.97	0.063	0.97	0.360

By utilizing the NARX transformed data structure on the FANN model, significant improvement in prediction performance could be observed across all five datasets. In term of time lag, the result shows that FANN-NARX(2) has an overall slight edges in performance compared to FANN-NARX(1) model, this is especially evident through the lower RMSE scores. The largest increase in performance could be observe from the Guangzhou dataset, where the FANN-NARX(2) model scored an R^2 values of 0.94 which is 53% higher than the base FANN model and a staggering 79% higher than the base MLR model.

4. Conclusion

In this study, nonlinear autoregressive network with exogenous inputs (NARX) were analysed in transforming the training data structure for $PM_{2.5}$ predictions. Two variations were examined NARX(1) with time lag of 1 and NARX(2) with time lag of 2. The models utilised for the prediction models are multi linear regression (MLR) and feed forward artificial neural network (FANN). When only trained with the base structure, the results shows than the FANN model is vastly better with an R^2 values of up to 0.69 (Beijing dataset) compared to that of MLR at 0.29. However, when coupled with NARX transformed data structure, the performance of the FANN model significantly improved. The best performing FANN-NARX(2) model achieved an R^2 values of up to 0.97 (Chengdu dataset) which is about 34% higher than base FANN model and 72% higher than MLR model. In future study, the generalization capability of the model could be further improved by training the model with a combined dataset instead of individual dataset and testing the model with real world external data.

Acknowledgement

This work was supported by Universiti Sains Malaysia (USM), special gratitude to Department of Environmental (DOE) Malaysia for providing the air quality data for this study and Kementerian Pendidikan Malaysia (KPM) through Fundamental Research Grant Scheme (FRGS) grant number FRGS/1/2018/TK02/USM/02/10.

References

Mobaseri, M., Mousavi, S.N., Mousavi Haghighi, M.H., 2021. Causal effects of population growth on energy utilization and environmental pollution: A system dynamics approach. Caspian Journal of Environmental Sciences 19, 601–618.

Gakidou, E., Afshin, A., Abajobir, A. A., Abate, K. H., Abbafati, C., et al., 2017. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet 390, 1345–1422.

Chen, C.-C., Chen, P.-S., Yang, C.-Y., 2019. Relationship between fine particulate air pollution exposure and human adult life expectancy in Taiwan. Journal of Toxicology and Environmental Health, Part A 82, 826–832.

World Health Organization (WHO). 2021, New WHO global air quality guidelines aim to save millions of lives from air pollution. Available at: https://www.who.int/news/item/22-09-2021-new-who-global-air-quality-guidelines-aim-to-save-millions-of-lives-from-air-pollution (Accesed: 30/10/2021)

Perez, P., Menares, C., Ramírez, C., 2020. PM2.5 forecasting in Coyhaique, the most polluted city in the Americas. Urban Climate 32, 100608.

Biancofiore, F., Busilacchio, M., Verdecchia, M., Tomassetti, B., Aruffo, E., Bianco, S., di Tommaso, S., Colangeli, C., Rosatelli, G., di Carlo, P., 2017. Recursive neural network model for analysis and forecast of PM10 and PM2.5. Atmospheric Pollution Research 8, 652–659.

Liang, X., Li, S., Zhang, S., Huang, H. & Xi Chen, S. 2016. PM2.5 data reliability, consistency and air quality assessment in five chinese cities. Journal of Geophysical Research: Atmospheres, 121, 10,220-10,236.

https://ieeexplore.ieee.org/document/548162

Buevich, A., Sergeev, A., Shichkin, A., Baglaeva, E., 2021. A two-step combined algorithm based on NARX neural network and the subsequent prediction of the residues improves prediction accuracy of the greenhouse gases concentrations. Neural Computing and Applications 33, 1547–1557.

Mohebbi, M.R., Karimi Jashni, A., Dehghani, M., Hadad, K., 2019. Short-Term Prediction of Carbon Monoxide Concentration Using Artificial Neural Network (NARX) Without Traffic Data: Case Study: Shiraz City. Iranian Journal of Science and Technology, Transactions of Civil Engineering 43, 533–540.

Grivas, G. & Chaloulakou, A. 2006. Artificial neural network models for prediction of PM10 hourly concentrations, in the greater area of Athens, Greece. Atmospheric Environment, 40, 1216-1229.

Kim, M.H., Kim, Y.S., Lim, J., Kim, J.T., Sung, S.W., Yoo, C., 2010. Data-driven prediction model of indoor air quality in an underground space. Korean Journal of Chemical Engineering 27, 1675–1680.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50295-5

A multi-output machine learning approach for generation of surrogate models in process engineering

Jimena Ferreira[a,b], Martín Pedemonte[b], Ana Inés Torres[a*]

[a] *Instituto de Ingeniería Química, IIQ, Facultad de Ingeniería, Universidad de la República, Montevideo, 11600, Uruguay*
[b] *Instituto de Computación, INCO, Facultad de Ingeniería, Universidad de la República, Montevideo, 11600, Uruguay*
aitorres@fing.edu.uy

Abstract

This work discusses a multi-output strategy to generate surrogate models in the chemical engineering context. The rationale behind our approach is that in a chemical process many outputs, such as concentrations of different species, are highly related by the underlying physicochemical phenomena. Hence, the expressions that model them should have terms that are common to all. In here, we extend our previous work on surrogate modelling to develop an algorithm that, in the same execution, learns models for several related output variables. We evaluate the algorithm by taking different combinations of CSTR arrangements as case studies, that represent structures with different levels of sharing. We conclude that the multi-output strategy is successful in building models that share common functionalities with adequate errors.

Keywords: Machine Learning; Surrogate model; Multi-output regression; Kaizen Programming; Evolutionary Algorithms.

1. Introduction

The use of surrogate models is related to the need of simple but accurate representations of the variation of output variables with several different input variables. Common techniques to build surrogate models including Gaussian Process (Kriging, Krige, 1951), Support Vector Machine (Smola and Scholkop, 2004) or ALAMO (Cozad et al., 2014), make use of a set of predefined functional bases to generate single-output models.

However, chemical processes are rarely single-output; many outputs such as concentrations of different species are highly related by the underlying physicochemical phenomena. For this reason, it is desirable to have techniques that can construct multi-output surrogate models. We hypothesize that in these multi-output models, outputs that we know have the same underlying physical-chemical phenomena, need to have models with terms that are common.

In this work, we extend the algorithm in Ferreira et al. (2019) to allow for the generation of several surrogate models concurrently for the different outputs of a system in one execution of the algorithm.

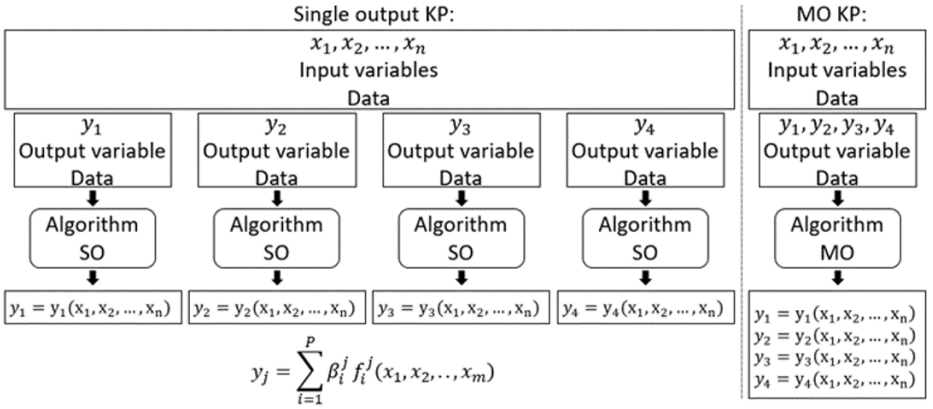


Figure 1 – Generation of surrogate models for several outputs with the same input data. Left: the single-output strategy repeatedly executes the algorithm to learn one model at a time. Right: the multi-output strategy learns all the models in the same run.

As before, the models are learned based on our implementation of Kaizen programming (KP, Ferreira et. al, 2021a). KP is an iterative algorithm for solving symbolic regression problems as a linear combination of nonlinear bases, with no a priori assumption of the functional bases. Thus, in the proposed multi-output extension of KP, models that share functional bases are favoured. A first approach to multi-output KP applied to the benchmark functions in Veloso et al. (2018) was included in Ferreira et al. (2021b).

The paper is organized as follows: section 2 presents the multi-output strategy, section 3 discusses the case studies, section 4 presents the numerical results.

2. Multi-output machine learning strategy

Figure 1 schematizes the general problem we are solving and the difference with the current approach. The general problem is finding surrogate models for several related output variables. Currently, most machine learning methods solve this problem by repeatedly executing the same algorithm using the same inlet data (represented by several vectors x) and the corresponding outlet data (one of the vectors y). We refer to this approach as the single-output (SO) strategy. By doing this, we obtain a set of outlet surrogate models for y, where each y_j is computed as a combination of the nonlinear basis f_i^j In KP these f_i^j are learned from the data, thus not necessarily the same set of f_i^j is obtained if a different execution is run for each output y_j. We will refer to the single-output strategy using Kaizen Programming as SO-KP.

If, the output variables y_j are related to each other, as is the case with many chemical engineering applications, it is desirable to learn the surrogate models for all the output variables in the same execution, as this allows for using the same functional bases for all the models that are being learned. We will refer to this multi-output strategy using Kaizen Programming as MO-KP.

Recalling that that there are two main steps in SO-KP: one where the functional bases (f_i^j) are created and modified using genetic programming, and another one where the

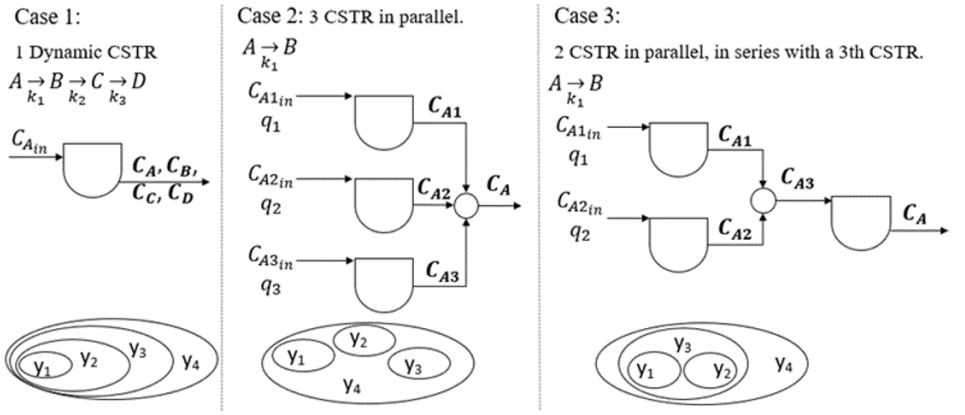


Figure 2 – Case studies and the scheme of shared functional bases. Output variables are in bold.

coefficients (β_i^j) are calculated by ordinary least squares (OLS), the modifications required for the MO-KP algorithm were at this second step and are as follows:

(i) We substituted the SO-OLS step, by a multi-output linear regression based on OLS (MO-OLS). At the MO-OLS step we first specified that the same set of functional bases f_i^j was considered for all outputs y_j; then, we added a step in which the functional bases were ranked according to the *p-value* (this parameter is used to measure the relevance of a particular functional basis in each final expression for y_j).

(ii) We used a different fitness function as a criterion for selection of the model: in SO-KP an adjusted R^2 of the output variable model was used, whereas in MO-KP, the minimum of all the adjusted R^2 of the output variables was used. In both cases this adjusted R^2 refers to training data.

It is important to note that these changes do not force the final expressions of y_j to all have the same functional bases but enhances the probability of that happening. More details of the algorithms can be found in Ferreira et al. (2019) and Ferreira et al. (2021).

3. Case studies

The proposed MO-KP algorithm was evaluated by considering three levels of overlap in the terms of the output variable models y_j, that may arise in chemical engineering settings (see Figure 2):

<u>*Case study 1-Large overlap (CS1)*</u>: considers the dynamic model of a CSTR with three first order reactions in series (A→B→C→ D). By deriving the analytical solution for this system (for a constant volume and inlet flowrate of reactant A) it can be shown that the functional bases of y_1 (in this case C_A (t, C_{Ain})) are included in the functional bases of y_2, (in this case C_B (t, C_{Ain})), those of y_2 in y_3 (in this case C_C (t, C_{Ain})), and so on.
<u>*Case study 2-Least overlap (CS2)*</u>: considers three steady state CSTRs in parallel, with a mixing point to combine the outlets. A first order reaction (A→B) occurs in each reactor. This could model, for example, three municipal wastewater treatment plants that receive different loads and are combined before final disposal. In the analytical solution of this system, y_1 (in this case $C_{A1}(q_1)$), y_2 (in this case $C_{A2}(q_2)$) and y_3 (in this

case $C_{A3}(q_3)$) do not share any functional bases, and y_4 (in this case $C_A(q_1, q_2, q_3)$) shares functional bases with the other three output variables.

Case study 3 (CS3)-Medium overlap: considers two steady state CSTRs in parallel and a third one in series with the other two; first order reactions (A→B) are considered for the three of them. This could represent also waste water treatment plants that receive different loads and are sent to a final treatment plant before disposal. In the analytical solutions for this case, y_1 ($C_{A1}(q_1)$) and y_3 ($C_{A3}(q_1, q_2)$) share some functional bases, y_2 ($C_{A2}(q_2)$) and y_3 share some other functional bases, but y_1 and y_2, do not share any functional bases; y_4 ($C_A(q_1, q_2, q_3)$) shares functional bases with the other three.

4. Numerical results and discussion

Input-output datasets were obtained by considering the analytic solutions of each output variable for each case study in Fig. 2. The following values were used: V=1 L, k_1=3.7 min^{-1}, k_2=4 min^{-1}, k_3=3 min^{-1}, q=2 Lmin^{-1}, $C_A(0)=C_B(0)=C_C(0)=C_D(0)=0$, C_{A1in}=1 molL^{-1}, C_{A2in} =2 molL^{-1}, and C_{Ain} =3 molL^{-1}. 100 points were randomly generated using a uniform distribution with the following ranges: $C_{Ain} \in$ [0.5-3] molL^{-1}, $t \in$ [0-3] min, $q \in$ [0.5-4] Lmin^{-1}. Each case study was executed 100 times, with a different training set, the parameters of the algorithm were as in Ferreira et al. (2021), except the number of iterations, which is set in 3000. The expressions with the best adjusted R^2 of MO algorithm over the learning set are as follows:

Case study 1:

$$y_1 = 41.38f_1 + 0.026f_2 + 0.075f_3 + 0.11f_4 + 0.31f_5 + 15.28f_6 + 4.01f_7 + 40.91f_8$$
$$y_2 = 28.70f_1 + 0.012f_2 + 0.69f_3 + 0.40f_4 - 0.66f_5 + 8.86f_6 + 4.15f_7 - 28.58f_8$$
$$y_3 = -33.67f_1 + 0.074f_2 - 1.98f_3 - 0.53f_4 + 1.51f_5 - 11.50f_6 - 4.36f_7 + 34.34f_8$$
$$y_4 = -64.59f_1 - 0.32f_2 + 4.87f_3 - 0.10f_4 - 0.26f_5 - 24.70f_6 - 5.33f_7 + 63.07f_8$$

With $f_1 = \frac{C_{A_{in}} t}{(t + 0.46)}$, $f_2 = \frac{C_{A_{in}} t}{(0.087t^2 + 1.03)}$, $f_3 = \frac{C_{A_{in}} t}{(1.73\, t + 0.087t^2 + 0.66)}$, $f_4 = \frac{C_{A_{in}}}{(0.28/_t + 0.50\, t)}$

$f_5 = \frac{C_{A_{in}}}{(0.63/_t + t)}$, $f_6 = \frac{C_{A_{in}}}{(t + 0.50)}$, $f_7 = \frac{C_{A_{in}}}{(t + 0.087t^2 + 0.38)}$, $f_8 = C_{A_{in}}$

Case study 2:

$$y_1 = 1.49f_1 + 0.034\, f_2 + 1.59\, f_3 - 0.0097f_4 + 0.034f_5 + 0.86f_6 + 0.14f_7 - 0.92 + 0.031q_1 + 0.031q_2$$
$$y_2 = 0.32f_1 + 0.049\, f_2 + 3.04\, f_3 - 0.049f_5 + 0.18f_6 + 0.13f_7 - 0.59 + 0.073q_1 + 0.073q_2 - 0.0039\, q_3$$
$$y_3 = -0.0058\, f_1 - 0.0058\, f_2 + 1.32\, f_4 + 0.010\, f_7 - 0.34 + 0.20\, q_3$$
$$y_4 = -2.57\, f_1 + 0.66\, f_2 - 2.37f_3 - 0.36f_4 + 0.66f_5 - 1.48\, f_6 - 0.76\, f_7 + 2.09 + 0.068\, q_1 + 0.068\, q_2 + 0.081\, q_3$$

With $f_1 = \frac{q_1}{(q_1 + q_2 + 1.59)}$, $f_2 = \frac{q_1}{(q_1 + 2\, q_2 + 1.38)}$, $f_3 = \frac{q_2}{(q_1 + 2\, q_2 + 1.40)}$, $f_4 = \frac{q_3}{(q_3 + 0.65)}$,

$f_5 = \frac{q_3}{(q_1 + 2q_2 + 1.38)}$, $f_6 = \frac{1}{(q_1 + q_2 + 1.59)}$, $f_7 = \frac{1}{(q_1\, q_2 + 0.58\, q_1 + q_2 q_3)}$

Case Study 3:

$$y_1 = -0.0015f_1 - 0.00078f_2 - 0.25f_3 - 0.34f_4 + 0.0070\, f_5 - 1.51f_6 + 0.63 + 0.18q_1 - 0.071q_2$$
$$y_2 = -0.0028\, f_1 - 0.0014\, f_2 - 0.39\, f_3 + 1.04\, f_4 - 0.54 + 0.28\, q_1$$
$$y_3 = 0.0062\, f_1 + 0.0031\, f_2 + 1.017\, f_3 - 1.18\, f_4 - 0.023\, f_5 - 3.83\, f_6 + 1.73 - 0.66\, q_1 + 0.26\, q_2$$

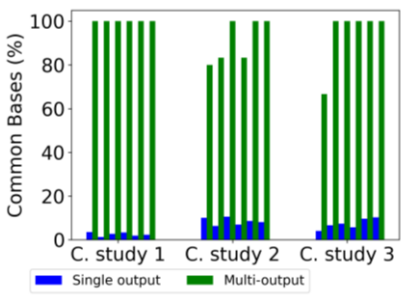


Figure 3 - Pairwise comparison of shared terms for each outlet expression, ordered as: y_1-y_2, y_1-y_3, y_1-y_4, y_2-y_3, y_2-y_4, y_3-y_4.

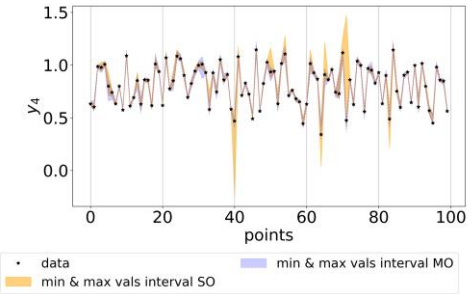


Figure 4 – Validation data points vs predictions for MO-KP and SO-KP for 100 executions. Results correspond to CS2.

$$y_4 = 0.00042\, f_1 + 0.00021\, f_2 + 0.43 f_3 - 0.70\, f_4 - 0.0069\, f_5 - 1.40\, f_6 + 0.60 - 0.24\, q_1 + 0.17\, q_2$$

With $f_1 = q_1 q_2$, $f_2 = q_1\, q_2^2$, $f_3 = \frac{q_1^2}{(1.56\, q_1 + q_2 + 1.26)}$, $f_4 = \frac{q_2}{(0.67 q_1 q_2 + 1.02\, q_2 + 2.17)}$, $f_5 = q_2^2$, $f_6 = \frac{1}{(0.85\, q_1 + q_2 + 2.17)}$

A graphical comparison of the number of terms shared by the different models in each case study, together with a comparison with the results from the SO-KP algorithm, is depicted in Figure 3. From here we conclude that MO-KP successfully finds all the overlaps that should be found (e.g., y_4 with all the other y_i in every case study, y_3 with y_1 and y_2 in CS3). However, it is also noticeable that expressions with overlapping terms for cases where there should not be any, are also returned. Examples of the latter are y_1-y_2-y_3 in CS2, and y_1-y_2 in CS3. This is a side effect of tuning the algorithm to favor common bases in all cases, regardless of the expected degree of sharing. On the other hand, SO-KP rarely finds a common term between expressions that do have them.

A comparison of the average and median RMSE (over validation sets) for the 100 executions is shown in Table 1. As seen both MO-KP and SO-KP provide accurate models. As expected, as every function is learned independently, SO-KP generally performs better in terms of the median of the error. However, it is interesting to note that in CS1, where the overlap of the functions is the largest, the error of MO-KP is similar or even better than that of SO-KP. Figure 4 shows a typical comparison between the output data points and the distribution of responses for both algorithms. As seen, despite lower overall errors, the SO-KP produces a few but significant outliers, which suggests that the model overfitted the learning data, thus did not generalize well to validation data. This was not seen in MO-KP, hence suggesting that MO-KP models should be preferred for applications as there is a lower risk of a highly erroneous prediction. However, preventing overfitting was not the main objective of this work, thus further studies on this direction may be needed.

5. Conclusions and future directions

In this work we analyse a multi-output strategy for learning surrogate models for several related output variables. In the analysis we consider three arrangements that are commonly found in Chemical Engineering settings and have different structures in terms of expected share of functional bases. We conclude that in terms of error in validation data the multi-output strategy builds models that are competitive with those

Table 1 - Average and median of RMSE distribution for 100 executions over validation sets

Output	Alg.	Case study 1		Case study 2		Case study 3	
		Average	Median	Average	Median	Average	Median
y_1	SO	3.16E+00	1.59E-03	9.02E-05	1.93E-05	2.83E-05	4.31E-06
	MO	8.01E-02	4.41E-04	8.34E-03	8.27E-03	1.73E-03	1.89E-03
y_2	SO	4,64E-01	1,74E-03	6,41E-02	2,26E-05	4,02E-05	6,50E-06
	MO	4,04E-01	1,18E-03	1,61E-02	1,55E-02	3,59E-03	3,72E-03
y_3	SO	1,22E+03	1,90E-03	2,46E-04	5,24E-05	1,83E-03	1,39E-03
	MO	2,89E-01	1,13E-03	2,48E-02	2,50E-02	2,73E-03	2,69E-03
y_4	SO	6,02E-01	9,58E-04	1,52E-02	1,16E-02	6,32E-04	5,39E-04
	MO	4,73E-01	1,66E-03	1,95E-02	1,95E-02	1,52E-03	1,48E-03

obtained with single-output strategy, and even better in the case study whose structure has a large overlap of terms. In addition, the multi-output strategy performs better than the single-output in terms of the presence of outliers. However, the strategy tends to find shared terms even in those cases where the structure prevents so. Thus, future versions of the algorithm should include steps that allow to find a balance in term-sharing based on the structure of the problem. Another future direction includes learning expressions that resemble the underlying physicochemical phenomena. In this sense, the expressions learned for CS2 and CS3 already include terms that have similarities with the analytical ones. This feature should be further exploited so that the algorithm returns models that are closer to the theoretically expected expression.

Acknowledgements

J. Ferreira thanks the Agencia Nacional de Investigación e Innovación of Uruguay for financial support for her graduate studies (Award No. POS NAC 2018 1 152185).

References

A. Cozad, N. V. Sahinidis and D.C. Miller, 2014, Learning surrogate models for simulation-based optimization. AIChE Journal. 60, 2211–2227.

V. V. de Melo and W. Banzhaf, 2018, Automatic feature engineering for regression models with machine learning: An evolutionary computation and statistics hybrid, Information Sciences, 430, 287-313.

J. Ferreira, M. Pedemonte and A. I. Torres, 2021a, Development of a Machine Learning-based Soft Sensor for an Oil Refinery's Distillation Column. Submmited.

J. Ferreira, A. I. Torres and M. Pedemonte, 2021b, Towards a Multi-Output Kaizen Programming Algorithm, 2021 IEEE Latin American Conference on Computational Intelligence (LA-CCI), Temuco, Chile. Accepted.

J. Ferreira, A. I. Torres and M. Pedemonte, 2019, A Comparative Study on the Numerical Performance of Kaizen Programming and Genetic Programming for Symbolic Regression Problems, 2019 IEEE Latin American Conference on Computational Intelligence (LA-CCI), Guayaquil, Ecuador, 1-6.

D.G. Krige, 1951, A statistical approach to some mine valuations and allied problems at the Witwatersrand. Master's thesis of the University of Witwatersrand.

A. Smola and B. Scholkop, 2004, A tutorial on support vector regression, Statistics and Computing, 14, 199–222.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50296-7

Practical Human Interface System for Transition Guidance in Chemical Plants using Reinforcement Learning

Shumpei Kubosawa[a,b*], Takashi Onishi[a,b], Yoshimasa Tsuruoka[a,c], Yasuo Fujisawa[d], Masanori Endo[d], Atsushi Uchimura[d], Masahiko Tatsumi[d], Norio Esaki[d], Gentaro Fukano[e], Tsutomu Kimura[e], Akihiko Imagawa[e], Takayasu Ikeda[e]

[a]*NEC-AIST AI Cooperative Research Laboratory, National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, 135-0064 Japan*
[b]*Data Science Research Laboratories, NEC Corporation, Kanagawa, 216-8666, Japan*
[c]*Department of Information and Communication Engineering, The University of Tokyo, Tokyo, 113-0033, Japan*
[d]*Production and Technology Center, Mitsui Chemicals, Inc., Tokyo, 105-7122, Japan*
[e]*Omega Simulation Co., Ltd., Tokyo, 169-0051, Japan*
kubosawa@nec.com

Abstract

In chemical plants, transition operations, such as changing the production load from 100% to 80%, are commonly performed to satisfy production needs. As plant models used in conventional automatic control methods (e.g. step response models) cannot predict non-steady states, these transition operations warrant manual control. Previously, we proposed an automatic optimal control method using dynamic simulators and reinforcement learning, a machine learning method in artificial intelligence (AI), for transition operations. We implemented this existing AI system in an actual industrial plant and determined that further improvements were required in the interaction between the system and human operators for reliable and acceptable guidance. In this paper, we propose a human interface system for realising optimal transition operations by enabling AI to cooperate with human operators. To validate and authorise the AI-proposed manipulations performed by human operators, the interface system presented the entire procedure and sensors influencing the AI decision for online disturbance rejection prior to actual manipulations. The interface, coupled with the control method, was evaluated experimentally in an actual chemical plant. The proposed system demonstrated optimised transition operations for producing purity changes under abrupt heavy rain disturbance in terms of guidance.

Keywords: chemical plant, reinforcement learning, optimisation, explainable AI.

1. Introduction

Modern chemical plants are commonly equipped with advanced controllers, such as model predictive control (MPC), to maintain steady states and stable production. In addition to steady operations, transition operations, such as changing production loads, are frequently performed in chemical plants to satisfy production needs. Although several automation methods, including sequence control, have been introduced to aid these

transition operations, limited methods exist for their optimal control. Previously, we proposed an automatic optimal control system for transition operations that leverages dynamic simulation and reinforcement learning (RL) (Kubosawa et al., 2021a and 2021b). However, in addition to control methods, the human interface is crucial for realising acceptable and reliable operations. This particularly applies to chemical plants, wherein any unforeseen incident can significantly impact society. In this paper, we propose a human interface system to guide transition operations in chemical plants. The interface system coupled with the control method was evaluated experimentally in an actual plant, which demonstrated that the optimised operations were efficiently performed with guidance from the system.

2. Related work

2.1. Simulation-based optimal control

Kubosawa et al. (2021a) proposed a method to optimise and control the transition operation of an actual plant by leveraging dynamic simulation, tracking simulation, RL and domain randomisation. The proposed method utilised dynamic simulators for training human plant operators as the plant model. They further improved the state identification performance using RL and proposed an RL-based disturbance rejection method (Kubosawa et al., 2021b). Both improvements were aimed at reducing the gaps between simulated and real states. These gaps can be triggered by multiple factors, including modelling errors, incorrect identification of states and external disturbances, such as changes in weather. If the difference in behaviours can be reduced by adjusting the parameter values of the model with time, then RL can potentially improve the simulation behaviour and enhance the state identification accuracy of the tracking simulation. As the simulation-to-reality gaps can also be caused by disturbances, human operators are required to handle disturbances by leveraging domain knowledge and plant dynamics based on their experiences. This defines the quantitative relationship between manipulation and response. Previously, we focused on this aspect and designed the RL task to minimise the gaps between simulated states and real states online. These methods were evaluated in an actual chemical plant, and the results verified the optimisation of transition operations under abrupt heavy rain disturbances.

Notably, the objectives of the proposed framework and nonlinear MPCs can be identical, that is, optimal control; however, their approaches are significantly different. MPCs commonly adopt online (on-site) optimisation methods, whereas RL is an offline optimisation method (i.e. prior optimisation to actual control). In addition, RL is a variant of machine learning methods, which consist of a field of artificial intelligence (AI). In this light, we distinguish the two approaches and refer the proposed method as an AI.

2.2. Explainable AI

As machine learning and AI applications include mission-critical systems, performance improvement alone is insufficient to determine whether AI can be adopted in actual operations. Therefore, the Defense Advanced Research Projects Agency initiated the Explainable AI (XAI) program in 2017. This program aimed to develop AI systems capable of explaining their rationale, strengths, weaknesses and future behaviour. As the explainability of a policy is essential in RL, one concept of the program explored the

development of explainable policy functions, whereas the other focused on developing an analysis method for existing black-box policies (Gunning and Aha, 2019).

2.3. Sensitivity analysis

Sensitivity analysis involves investigating the behaviours of mathematical models considering the changes in variables, including parameters and inputs (Pannell, 1997).

$$g(\boldsymbol{x}, i) = \left| \frac{\partial f(x_1, x_2)}{\partial x_i} \right| \tag{1}$$

Eq. (1) is an example index $g(\boldsymbol{x}, i)$ for the analysis of a model $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ using a partial derivative of the i-th input variable x_i. This index describes the effect of a minor change in x_i on the output. In other words, it indicates the importance of each input in the model prediction (i.e. weight of the nonlinear model at the input). Sakahara and Kubosawa (2021) proposed the directions of applying these analysis methods to RL models.

3. Proposed method

3.1. Operation flow

The transition operations of chemical plants involve non-steady states and unfamiliar situations. Therefore, human operators conduct these procedures cautiously to handle unforeseen situations during operation. To support human operators anticipating various situations, AI-based procedures should be validated by human operators before acceptance or rejection. To develop validation methods, we leveraged the two concepts of XAI. One concept presents the future behaviour of the AI, whereas the other analyses the AI's perspective by determining which sensor affects AI decisions.

Figure 1 depicts the workflow of the human operator, overall architecture of the AI controller and proposed interface system. The thick and thin arrows indicate the process and data flows, respectively. The AI system comprises three RL agents: First, the

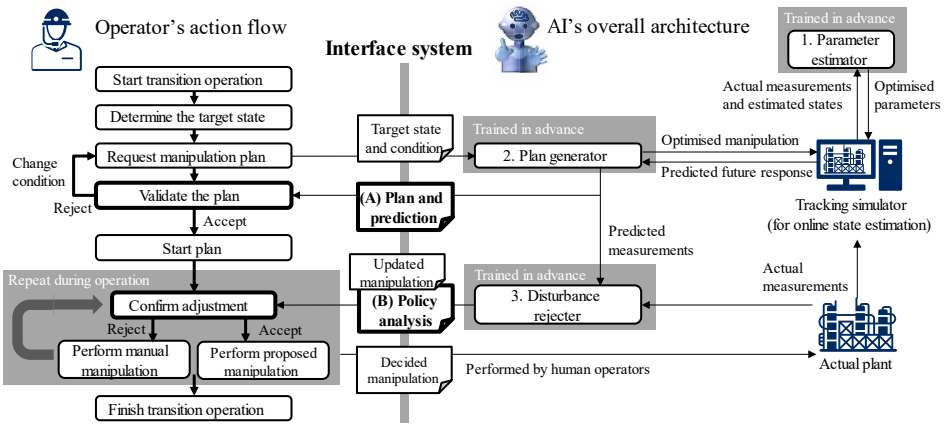


Figure 1. Operation flow and AI architecture with the proposed interface system.

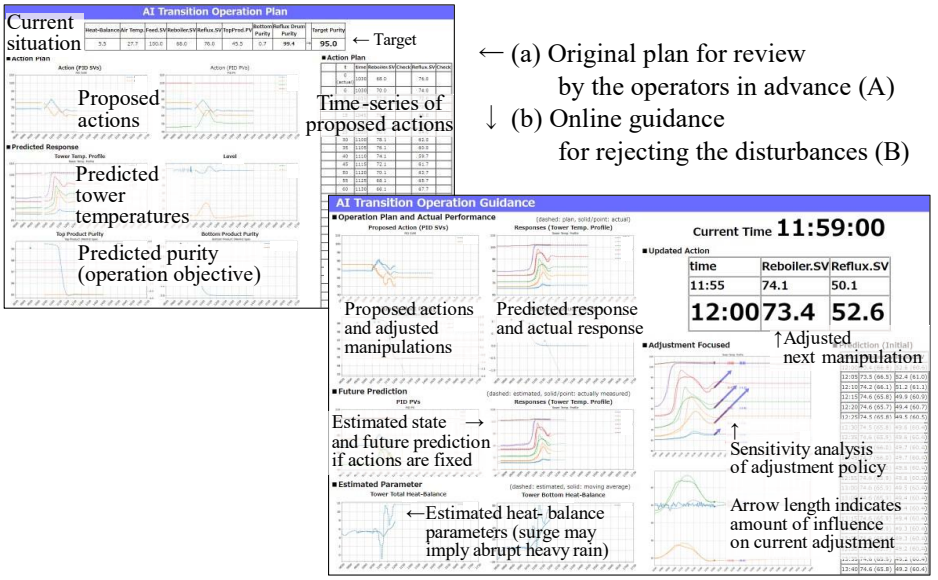


Figure 2. Interface system screenshots.

parameter estimator provides optimised parameter values to the tracking simulator, which receives real-time manipulations from the actual plant and estimates the plant state online. Second, the plan generator presents an optimised procedure that begins with the current plant state and finally achieves the required target state. Third, the disturbance rejecter adjusts the planned manipulation values and presents optimal values if the gap between the actual and predicted states is unexpectedly widened (Kubosawa et al., 2021b).

As indicated in Figure 1, two major interaction points exist between human operators and AI in this interface. The first point (A) presents the time-series of the optimised procedure and predicted future response of the total operation to a specified target state and condition, such as target production load of 80%, or the priority of the procedure, including energy efficiency and robustness. The second point (B) presents the adjusted manipulations to reduce simulation-to-reality gaps.

3.2. Assessment of the offline procedure

To begin transition operations, human operators determine and input the target situation and desired conditions to the interface system. The system presents (A) the operation plan comprising the time series of the optimised procedure, which includes the set-point values of the proportional–integral–differential controllers, and the predicted future response. The future behaviour of the AI is presented in its entirety prior to the actual manipulations. Figure 2(a) depicts a screenshot of the plan presented during the experiment. The qualitative direction of the manipulation over time, such as increasing or decreasing, is a major aspect that verifies the procedure. Human operators review the entire plan and determine its acceptance or rejection. If the plan is accepted, then it is initiated, and the system monitors the gaps between the predicted and actual responses.

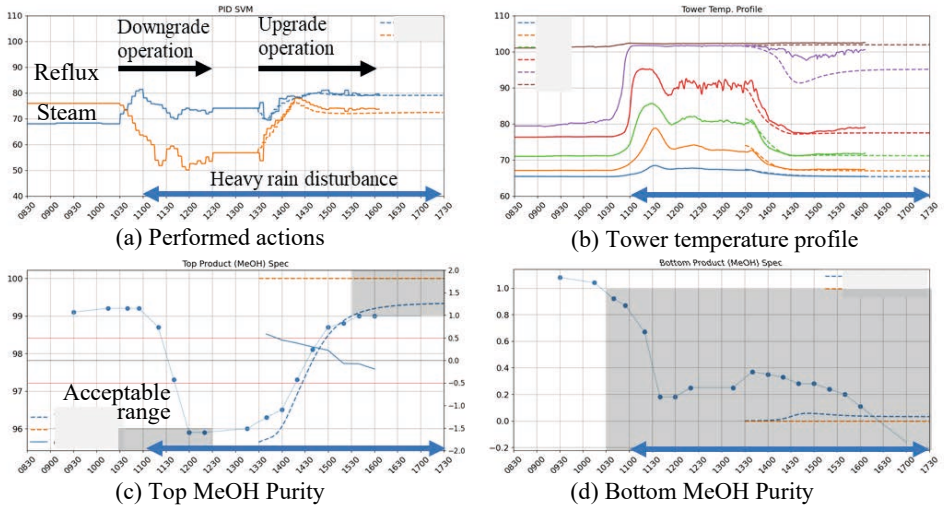
3.3. Assessment of the online action

During the procedure, gaps may emerge between the predicted and actual responses owing to a mismatch in the dynamics, such as temporal changes in model parameters, and external disturbances. To handle such abrupt changes in actual situations, the disturbance rejection agent works online. The agent monitors the gaps and proposes the adjusted and updated manipulation values to cancel the gap periodically. We used a control interval of 5 min in the experiment. To explain the adjustment and review the manipulations before performing them, the interface system analyses the behaviour and presents the result (B) to the operators 3 min prior to the subsequent manipulation, during which the operators review the manipulation and determine its acceptance or rejection. Figure 2(b) illustrates a screenshot of the guidance for rejection during the experiment. The system presents the importance of the sensors in the adjustment using the sensitivity analysis of the adjusting policy regarding the current state, which is a partial derivative of the policy function of states or each sensor. If the AI-focused sensors, which are typically mismatched, are considered inappropriate in the current situation, then the operators can reject the proposal.

4. Experiment

We evaluated our system at an actual methanol distillation plant for training human operators, as described in previous studies (Kubosawa et al., 2021a and 2021b). The plant separates methanol and water from its liquid mixture. The experimental task involves the transition operation of producing purity changes (downgrade and upgrade) under abrupt disturbances caused by heavy rains. A detailed description of the plant and its model used in this experiment is presented in Kubosawa et al. (2021b).

In the experiment presented in Figure 3, all proposals were accepted by the human operators. Figure 3 illustrates the responses obtained during the downgrade and upgrade operations. The dashed lines indicate the originally predicted actions or responses. In Figure 3(a), the gaps between the solid (adjusted) and dashed (originally planned) lines



(a) Performed actions (b) Tower temperature profile

(c) Top MeOH Purity (d) Bottom MeOH Purity

Figure 3. Experimental results of the purity transition operation using the proposed

indicate the adjustment of actions. Figure 3(b) depicts the oscillation phenomena observed in the response, which were suppressed owing to the adjustments. Figure 3(c) illustrates the time series of the top purity, and the points indicate the actual sampled purity. The shaded region in the figure indicates the acceptable purity. Both operations achieved acceptable states when counteracting the heavy rain disturbance, which began shortly after the first downgrade operation.

In the other experiment of a downgrade operation, the transmission of the sensor readings to the system was abruptly suspended (i.e. the sensor readings were fixed since then), and the excessive action values has been proposed. This is because the temperatures of the actual plant were changing, whereas, as for the system, the response of the proposed corrected action seemed to be stayed; thus, the system significantly increased the action values. In this case, the human operators rejected the proposal during 3 min for the judgement and selected the originally planned values instead.

5. Conclusions

As transition operations in chemical plants involve non-steady states, unforeseen situations may be triggered. Therefore, explainable interfaces for AI guidance systems are required. In this study, we proposed a practical explanatory interface system using an existing automatic control system. The interface system proposes optimal procedures that can be performed by human operators with reliability and acceptability. We evaluated the entire system considering the transition operation for producing purity changes in an actual chemical plant and demonstrated the optimisation of the operations. In the future, we intend to investigate the application and installation of the proposed system to industrial chemical plants to optimise the actual production.

Acknowledgements

The authors thank Kazumi Yamamoto, Toshikazu Tanaka, Kiyoshi Sugawara and Katsushige Sumida (Mitsui Chemicals Inc.) for evaluating the system as expert operators and providing general useful comments on chemical plant operations.

References

Gunning, D., and Aha, D. (2019). DARPA's explainable artificial intelligence (XAI) program. AI Magazine, Vol. 40, No. 2, pp. 44–58.

Kubosawa, S., Onishi, T., and Tsuruoka, Y. (2021a). Computing operation procedures for chemical plants using whole-plant simulation models. Control Engineering Practice, Vol. 114, p. 104878.

Kubosawa, S., Onishi, T., and Tsuruoka, Y. (2021b). Non-steady state control under disturbances: navigating plant operation via simulation-based reinforcement learning. In Proceedings of the 60th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), pp. 796–803.

Pannell, D.J. (1997). Sensitivity analysis: strategies, methods, concepts, examples. Agricultural Economics, No. 16, pp. 139–152.

Sakahara, M., and Kubosawa, S. (2021). Future directions in sensitivity analysis on deep reinforcement learning for process control via Fisher information matrix. In Proceedings of the 60th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), pp. 804–807.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50297-9

Surrogate modeling for nonlinear gasoline blending operations

Tasabeh H. M. Ali[a], Robert E. Franzoi[a], Brenno C. Menezes[a*]

[a]*Division of Engineering Management and Decision Sciences, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar*
bmenezes@hbku.edu.qa

Abstract

The application of surrogate modeling in engineering is surging recently for predicting the functional behavior of a system using analytical formulations as an alternative to complex models that often lead to non-convergence issues and not sufficiently accurate solutions in decision-making problems. The surrogate model building procedure addressed in this paper consists of four major steps to be applied in nonlinear blending of gasoline streams. The first is the input (x) dataset generation, performed using the Latin Hypercube Sampling (LHS) technique, which is coupled with a rescaling strategy, and used for evaluating the output (y) dataset. Secondly, the generated data are improved with a normalization procedure to mitigate numerical issues and to avoid biased surrogates. Thirdly, mixed integer quadratic programming (MIQP) formulation based on the least-squares regression is employed to build an optimizable surrogate function for each variable of interest. Fourthly, smaller and simpler surrogates are established and selected to be employed for gasoline blending operations by substituting the complex nonlinear and nonconvex rigorous formulation in an optimization case.

Keywords: surrogate modeling, blending operations, optimization, machine learning.

1. Introduction

Quality specifications, operational complexity, and environmental regulations are the most challenging obstacles in chemical processes, especially in crude-oil refineries, which affect the supply-chain profitability (Lotero et al., 2016). Most blend properties of final products are nonlinear and non-convex. Hence, estimating data-driven final product properties from computer-controlled in-line databases can be a quite complex process (Ounahasaree et al., 2016). The optimization addressed herein uses predicted formulas based on amounts of inlet streams to replace nonlinear complexities for optimizing the final products' amounts by matching blended properties with regulated specifications. Rigorous models of first principles, mechanistic, physics and engineering-based techniques can be reduced to black-box surrogate models to predict causation and correlation in blended properties of amounts of intermediate streams in scheduling, planning, multi-unit coordinating, and real-time optimization problems. There are previous works on surrogate-based optimization addressing MIQP (mixed-integer quadratic programming). Straus et al. (2018) applied MIQP on ammonia reaction processes to predict the optimal selection matrix that reduces the sampling-space in a self-optimizing variable surrogate modeling. Franzoi et al. (2020) used MIQP-based surrogate modeling to predict blending calculations of distillation unit outputs considering multiple feedstocks and operational variables. Franzoi et al. (2021) proposed an adaptive sampling MIQP-based surrogate modeling to predict reaction

system conversions. Moreover, surrogates can also be used as factor-flow balances in blenders for blend scheduling optimization (Kelly et al., 2018). In general, surrogate modeling is employed for representing algebraic formulas $Y = f(x)$ that formulates the causation relationships between the input of independent variable (x) and the output of dependent variable (Y) to model complex, unmodeled, or unknown systems.

2. Problem statement

Blending operations are continuous processes of feedstocks entering a blender to produce a final product with determined specification on qualities. In this paper, the gasoline blending operation is addressed. There are nine feedstocks (x) in the process presenting the most common refinery blends, which are shown in Table 1, with their respective nomenclatures for $j \in \{1..9\}$. Similarly, yields of $i \in \{1..7\}$ for properties (y) are considered, namely, Reid vapor pressure (RVP), aromatic content (ARO), olefin content (OLE), specific gravity (SG), sulfur content (SUL), research octane number (RON), and motor octane number (MON). These are the most important properties in operational and economical levels for enhancing the gasoline octane rating, which is especially important to increase fuel performance, suitability, and efficiency. These properties are also relevant for the fuel volatility, combustibility, and level of pollution.

Table 1: Gasoline blending feedstocks and their properties (Menezes et al., 2014).

Properties (y_i) / *Feedstocks (x_j)*	*RVP (kPa)*	*RON*	*MON*	*ARO (vol%)*	*OLE (vol%)*	*SG (g/cm^3)*	*SUL (ppmw)*
Hydrotreated Light Cracked Naphtha	55	93	82	25	10	0.729	0.005
Hydrocracked Naphtha	52	94.05	81.78	45	29	0.758	0.002
Hydrotreated Cocker Light Naphtha	56	83	76	2	1	0.718	0.005
Light Naphtha	90.24	69.1	67.1	0.001	0.00001	0.699	0.0059
Reformate	85	98	90	54	20	0.79	0.005
Ethanol	17	109	90	0.1	0.02	0.79	0
Isomerate	69	106	100	0.1	0.02	0.85	0.005
Alkalyte	85	96	92	54	20	0.85	0.005
Butane	51.52	93.8	90	0	0	0.601	0

3. Surrogate modeling methodology

1) Data Generation: the independent input (x) variables of compositions are generated using the Latin Hypercube Sampling (LHS) technique, a statistical method of near-random distribution sampling. Each point in the data set is constrained by Eq.(1) to sum to 1; each sampled x' of feedstock j in IV (independent variables) is constrained to produce a final x of j. This ensures the mass balance across the blending operations.

$$x_j = \frac{x'_j}{\sum_{j \in IV} x_j} \quad \forall \; j \in IV \tag{1}$$

The dependent output (y) variables of yields are generated using the formulas from Eq. (2) to Eq.(6) for each sampled point of the dataset (Ounahasaree et al., 2016), which are regarded as the nonlinear properties. The subscripts j refer to the feedstock of the stream IN resembled in a data point entering the blender, and G is the gasoline product. The terms Q_j, sul_j, and sg_j respectively resemble the volumetric flowrate, sulfur content, and specific gravity of feedstock j. The octane properties of gasoline (RON and MON) are calculated using Eq.(7) to Eq.(11). Index v is the volume-based property, ARO_{VQ} is the volume-based property for ARO^2, and coefficients a to g are experimentally estimated values from real data retrieved from Menezes et al. (2014).

$$RVP_G = \frac{\left(\sum_{j \in IN}(Q_j RVP_j^{1.25})\right)^{0.8}}{\sum_{j \in IN} Q_j} \tag{2}$$

$$\mathrm{ARO}_G = \frac{\sum_{j \in IN} Q_j \mathrm{ARO}_j}{\sum_{j \in IN} Q_j} \tag{3}$$

$$\mathrm{OLE}_s = \frac{\sum_{j \in IN} Q_j \mathrm{OLE}_j}{\sum_{\in IN} Q_j} \tag{4}$$

$$sg_s = \frac{\sum_{j \in IN} Q_j sg_j}{\sum_{j \in IN} Q_j} \tag{5}$$

$$sul_G = \frac{\sum_{j \in IN} Q_j sg_j sul_j}{\sum_{j \in IN} Q_j sg_j} \tag{6}$$

$$RON_{jv} = RON_j + a\left[\left(RON_j - RON_v\right)\left(J_j - J_v\right)\right] + b\left(ARO_j - ARO_v\right)^2 + c\left(OLE_j - OLE_v\right)^2 + d\left[\left(ARO_j - ARO_v\right)\left(OLE_j - OLE_v\right)\right] \quad \forall \; j \in IN \tag{7}$$

$$MON_{jv} = MON_j + e\left[\left(MON_j - MON_v\right)\left(J_j - J_v\right)\right] + f\left(ARO_j - ARO_v\right)^2 + g[2\left(OLE_j - OLE_v\right)^2\left(ARO_{VQ} - {ARO_v}^2\right) - \left(ARO_{VQ} - {ARO_v}^2\right)^2] \quad \forall \; j \in IN \tag{8}$$

$$J_{j,v} = RON_{j,v} - MON_{j,v} \quad \forall \; j \in IN \tag{9}$$

$$RON_G = \frac{\sum_{j \in IN} Q_j RON_{jv}}{\sum_{j \in IN} Q_j} \tag{10}$$

$$MON_G = \frac{\sum_{j \in IN} Q_j MON_{jv}}{\sum_{j \in IN} Q_j} \tag{11}$$

The data set points are partitioned into 50% training and 50% testing, the former for building the surrogate functions, and the latter to evaluate their performance, reliability, and robustness.

2) Data improvement: the normalization of the generated x and y variables (Var) using Eq.(12) is performed to prevent the data from being biased or running across numerical

issues when building the surrogates through MIQP optimizations. The terms of $min(VAR_{i,j}^{tr})$ and $max(VAR_{i,j}^{tr})$ refer to the minimum and maximum values of the training set array. Also, DV and IV refer to dependent and independent variables, respectively.

$$VAR_{i,j} = \frac{VAR_{i,j} - min(VAR_{i,j}^{tr})}{max(VAR_{i,j}^{tr}) - min(VAR_{i,j}^{tr})} \quad \forall\ j \in IV, i\ \in DV \tag{12}$$

3) *Surrogate Model Building:* surrogates are built following a bilinear functional form estimation as in Eq.(13) that correlates the normalized x and y variable sets, with coefficients of b_{ij} and c_{ij} to be predicted using this correlation. The intercepts are eliminated to avoid any multi-collinearity issues. The estimation uses MIQP optimizations, utilizing the objective function of Eq.(14) that minimizes the least-squares error (LSE) for linearly regressing the y into Y, subject to constraints Eq.(15) to Eq.(17), using $n_{training}$ points only, with z as a binary decision variable and M as a sufficiently large number (Franzoi et al., 2020).

$$Y_i = \sum_{j \in IV} b_{ij} X_j + \sum_{j \in IV} \sum_{k \in IV} c_{ijk} X_j X_k \quad \forall \{(j,k) \in IV, j\ \leq k\}, i \in DV \tag{13}$$

$$Minimize\ \sum_{p=1}^{n_{training}} (y_{ip} - Y_{ip})^2 \tag{14}$$

$$-Mz_j \leq b_{ij} \leq Mz_j \qquad \forall\ j \in IV, i\ \in DV \tag{15}$$

$$-Mz_{jk} \leq c_{ijk} \leq Mz_{jk} \qquad \forall \{(j,k) \in IV, j\ \leq k\}, i \in DV \tag{16}$$

$$\sum_{j \in IV} z_j\ + \sum_{j \in IV, k \in IV, j \leq k} z_{jk}\ \leq\ K \quad \forall\ z_j, z_{jk}\ \{0,1\} \tag{17}$$

4) *Performance Check*: surrogates are used to-recalculate the generic dependent variables of Y, which are regarded as bilinear qualities of surrogates. The model performance is investigated by calculating the prediction errors and carrying out statistical analysis to ensure good predictability and reliability of results.

4. Results and discussion

Two distinct case studies are established for constructing the surrogate models and investigating their performance. For simplicity, only blended properties of the Y2 and Y3 plots are shown in Figure 1, Figure 2, and Figure 3, as they represent the most complex and difficult to predict variables considered in the formulation.

Case I: The model is built using data sizes $N\ \in \{100, 1000\}$; to testify the effect of the data set size in building it by comparing the MAE, referred to as prediction errors. Figure 1 illustrates that increasing the set size enhances the model overall performance, with lower testing and training MAE errors. In contrary, Table 2 reveals comparable model performance for both N trials, that shows prediction errors not exceeding $MAE \approx 10^{-4}$ for both trials. In contrast, the percentual errors did not exceed $MAE\% \approx 10^{-3}$ and $MAE\% \approx 10^{-4}$, respective to the trials, proving that prediction errors are enhanced with the data size increment. The overall results account for satisfying model performance. However, this is not sufficient for evaluating the model robustness and reliability. Hence, statistical analysis of MAE average, variance, standard deviations, and confidence intervals are also carried out to complement the analysis. Thus, 5%

uncertainty is accounted for testing the poor prediction regions (reliability) and the extent of applicability (robustness). With using 95% confidence interval probability, Figure 2 illustrates the estimation of the statistical analysis arranged in ascending order. Consecutively, it resulted in base points plot, lying between upper and lower bounds curves of MAE added to and subtracted from the confidence intervals respectively, showing narrow convergence between all the curves. This illustrates that the variances and the standard deviations are too small, and all regions' MAE are approximate to the punctual values, accounting for satisfying model reliability and robustness.

Table 2: Mean absolute errors for the surrogate predictions.

	Y1 (RVP)	Y2 (RON)	Y3 (MON)	Y4 (ARO)	Y5 (OLE)	Y6 (SG)	Y7 (SUL)
MAE (N=100)	$6.10\ 10^{-4}$	$7.28\ 10^{-8}$	$8.98\ 10^{-5}$	$8.94\ 10^{-8}$	$1.09\ 10^{-7}$	$3.43\ 10^{-10}$	$1.16\ 10^{-8}$
MAE% (N=100)	$1.07\ 10^{-3}$	$7.81\ 10^{-8}$	$1.09\ 10^{-4}$	$3.97\ 10^{-7}$	$1.09\ 10^{-6}$	$4.62\ 10^{-8}$	$2.70\ 10^{-4}$
MAE (N=1000)	$3.19\ 10^{-4}$	$6.12\ 10^{-8}$	$8.08\ 10^{-5}$	$1.25\ 10^{-7}$	$5.44\ 10^{-8}$	$3.20\ 10^{-10}$	$9.36\ 10^{-9}$
MAE% (N=1000)	$5.68\ 10^{-4}$	$6.57\ 10^{-8}$	$9.75\ 10^{-5}$	$5.70\ 10^{-7}$	$5.48\ 10^{-7}$	$4.30\ 10^{-8}$	$2.21\ 10^{-4}$

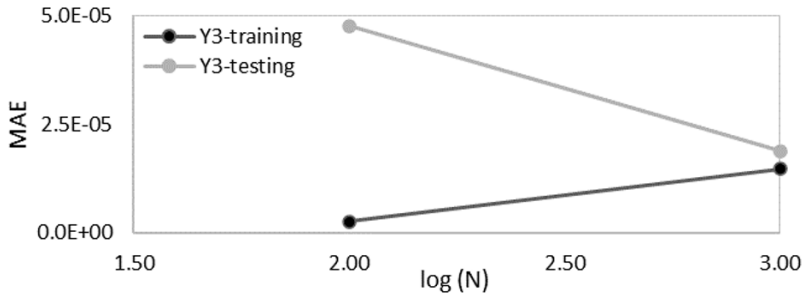


Figure 1: MAE of bilinear surrogates versus the number of sample points.

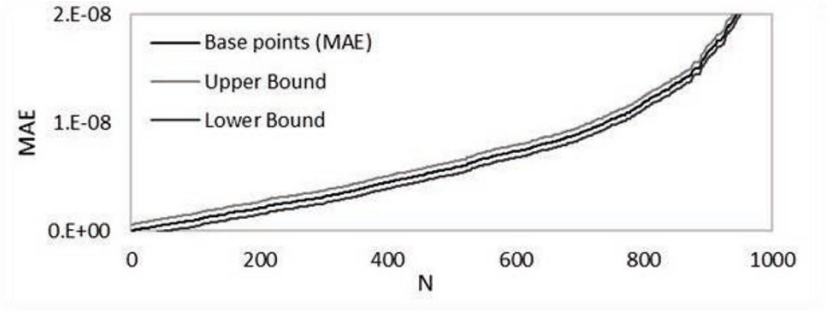


Figure 2: MAE with confidence interval range for the sample points of Y2.

Case II: Bilinear surrogates with simpler formulation and smaller size for $N \in 1000$ are identified by limiting the number of terms (K) in the surrogate (originally 54). Thus, the surrogates are tested by plotting the MAE against K as shown in Figure 3 to identify whether a lower number of coefficients with provides sufficient accuracy. As a result, the plot confirmed that using up to 45 simplified terms, the surrogates are predicted to generate the same original model accuracy. Using fewer coefficients compromises the quality of the surrogate. In general, this procedure does not give the exact optimal value of K, especially as it depends on different aspects, such as the problem type, dimensionality, functional form, data size, etc.

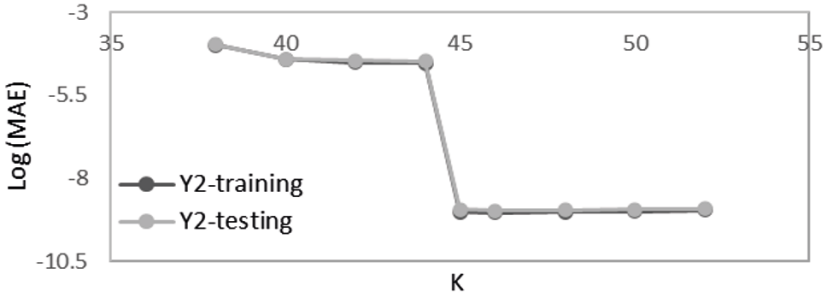


Figure 3: MAE versus different number of coefficients for Y2 surrogate.

5. Conclusion

Surrogate modeling is utilized to predict analytical formulations for complex systems. In this work, nonlinear equations used to predict gasoline blending properties are replaced by constructed bilinear surrogates. The surrogates for both values of N show percentual errors not exceeding $MAE\% \approx 10^{-3}$ and $MAE\% \approx 10^{-4}$ respectively, with noticeable error decrement as the data size increases. The deviation from MAE is also calculated using 5% uncertainty to account for eventual poor predictions, and to measure the extent of the model applicability. Hence, a 95% confidence interval is used to calculate the lower and upper bounds of the MAE. As a result, the surrogates show sufficient performance, robustness, and reliability in replacing the original equations. In the second case study, simpler and smaller size surrogates are achieved. Presenting Y2 results, decreasing the surrogate terms from fifty-four to forty-five shows prediction errors of high proximity between the two models. However, further methodologies need to be developed in future to properly adjust the surrogate model size within an automatic and systematic fashion.

References

R. E. Franzoi, J. D. Kelly, B. C. Menezes, C. L. E. Swartz, 2021, An adaptive sampling surrogate model building framework for the optimization of reaction system, Computers and Chemical Engineering, 152, 107371.

R. E. Franzoi, B. C. Menezes, J. D. Kelly, J. A. W. Gut, I. E. Grossmann, 2020, Cutpoint temperature surrogate modeling for distillation yields and properties, Industrial and Engineering Chemistry Research, 59 (41), 18616–18628.

J. D. Kelly, B. C. Menezes, I. E. Grossmann, 2018, Successive LP approximation for nonconvex blending in MILP scheduling optimization using factors for qualities in the process industry, Industrial and Engineering Chemistry Research, 57(32), 11076-11093.

I. Lotero, F. Trespalacios, I. E. Grossmann, D. J. Papageorgiou, M. S. Cheon, 2016, An MILP-MINLP decomposition method for the global optimization of a source based model of the multiperiod blending problem, Computers and Chemical Engineering, 87, 13–35.

B. C. Menezes, L. F. L. Moro, W. O. Lin, R. A. Medronho, F. L. P. Pessoa, 2014, Nonlinear production planning of oil-refinery units for the future fuel market in Brazil: Process design scenario-based model, Industrial and Engineering Chemistry Research, 53(11), 4352–4365.

Y. Ounahasaree, U. Suriyapraphadilok, M. Bagajewicz, 2016, global optimization of gasoline blending model using bound contraction technique, Computer Aided Chemical Engineering, 38, 1293–1298.

J. Straus, S. Skogestad, 2018, Surrogate model generation using self-optimizing variables, Computers and Chemical Engineering, 119, 143–148.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50298-0

Continuous Manufacturing Process Sequential Prediction using Temporal Convolutional Network

Haoran Li[a,b], Tong Qiu[a,b*]

[a] *Department of Chemical Engineering, Tsinghua University, Beijing 100084, China*
[b] *Beijing Key Laboratory of Industrial Big Data System and Application, Tsinghua University, Beijing 100084, China*
qiutong@tsinghua.edu.cn

Abstract

In the era of intelligent manufacturing, the continuous manufacturing industry will benefit from digitalization technologies such as digital twins. This paper proposes a temporal convolutional network sequence-to-sequence (TCN-StS) model as a data-driven simulation tool for the construction of digital twins. The proposed model captures time delay information through temporal convolution operation and thus better predicts the process state variations than recurrent neural networks on an actual industrial sintering dataset and shows good robustness over time. This study sheds new light on process sequence-to-sequence modelling through convolutional networks.

Keywords: sequence-to-sequence; temporal convolution; digital twin

1. Introduction

The continuous process is ubiquitous in steel, chemical, pharmaceutical, and other manufacturing industries. Prediction and operation decision-making in traditional continuous manufacturing processes rely on the knowledge reserve and cognitive level of operators, which severely restricts the safe and efficient operation of the production process. Over the decades, the soaring development of big data and artificial intelligence has brought transformational opportunities for the digitization of the process industry. The concept of digital twin (Glaessgen and Stargel, 2012; Gockel et al., 2012), initially proposed by the National Aeronautics and Space Administration (NASA), has recently been transplanted and deemed as the future solution to the manufacturing industry (Rosen et al., 2015).

Sequence-to-sequence modelling is the closest approach as a digital twin, as it uses the historical operation sequences to capture the dynamics of the process and predict the future evolution. Chou et al. first designed a sequence-to-sequence soft sensor model and performed excellently for product impurity predictions of an industrial distillation column (Chou et al., 2020). Kang et al also built a sequence-to-sequence model and achieved rolling predictions in the process of vapor-recompression C3 (Kang et al., 2021). Although canonical recurrent neural networks such as LSTMs and GRUs are considered by most deep learning practitioners synonymous with sequence modelling, Bai et al indicated that temporal convolutional networks (Lea et al., 2016) outperformed recurrent neural networks across a diverse range of tasks and datasets, while demonstrating longer effective memory (Bai et al., 2018).

These recent researches illustrate the importance of sequence-to-sequence modelling for the manufacturing industry and the potential of convolutional neural networks as a

positive option for sequential modelling. This paper proposes a temporal convolutional network sequence-to-sequence (TCN-StS) model to achieve continuous manufacturing process sequential prediction. The model is applied to the sintering process in the iron-making industry and tested in an actual industrial dataset.

2. Methodology

2.1. Process description

The sintering process is an important thermochemical process in the blast furnace ironmaking system. It involves the heating of fine iron ore with flux and coke fines or coal to produce a semi-molten mass that solidifies into porous pieces of sinter with the size and strength characteristics necessary for feeding into the blast furnace.
This process is a typical continuous manufacturing process (Figure 1). Firstly, iron ore, coke, limestone, and returning sinter are mixed and then fed in a moving trolley to form a uniform sintering bed. Next, the igniter ignites the surface of the bed, and the blower under the moving trolley generates negative pressure in the bellow below the bed through the exhaust. As the trolley gradually moves to the end of the sintering machine, the combustion continues to develop downward (Zhou et al., 2019). In the end, the raw ore powder will gradually form sintered ore with a certain particle size, which will enter the subsequent blast furnace ironmaking production as iron material.

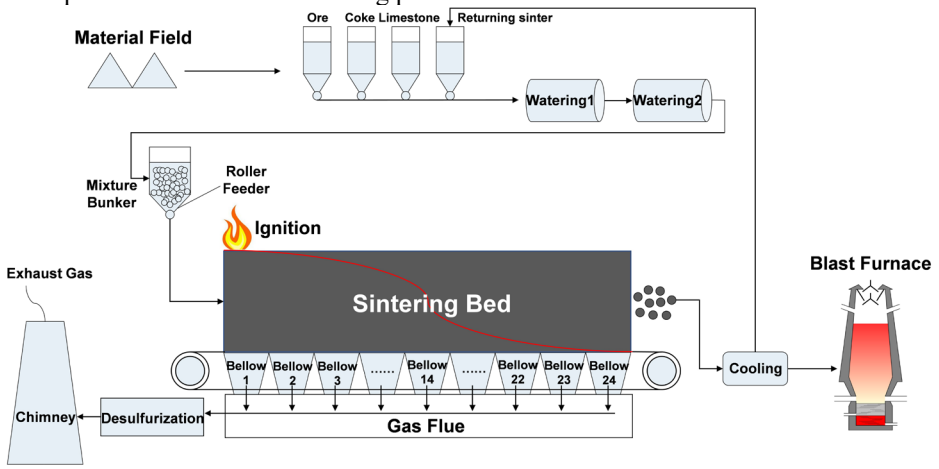


Figure 1. Sintering process schematic.

The sintering process owns the following two characteristics:

Time delay. There is a time interval between the change of process variables in the sintering system and its downstream variables, which is called mechanism time lag. In addition, for the same batch of raw materials, because different variables are measured at different times, there will be a technical time lag. The coexistence of the two types of time delays makes the sintering process exhibit strong time-delay characteristics.

Non-linearity. In the sintering process, a large number of chemical reactions such as coke combustion and limestone decomposition and a two-dimensional three-phase complex heat and mass transfer relationship exist at the same time, so the sintering system variables present obvious nonlinear relationship characteristics.

An industrial sintering process dataset of 27,000 samples was obtained and used in this study. The dataset was collected from 2019/12/05 to 2019/12/24 with a sampling frequency of 1 min. There are 15 key variables including 9 operating variables (OVs) and 6 state variables (SVs) in the process (Table 1).

Table 1. Key variables of the process.

Variable	Notation
Sintering bed thickness	OV1
Ignition intensity of row A	OV2
Ignition intensity of row B	OV3
Ignition temperature	OV4
Trolley speed	OV5
Round roller speed	OV6
Seven roller speed	OV7
Frequency of No.1 exhaust fan	OV8
Frequency of No.2 exhaust fan	OV9
Bellow 14 negative pressure	SV1
Bellow 22 negative pressure	SV2
Bellow 14 gas temperature	SV3
Bellow 22 gas temperature	SV4
End point of sintering	SV5
End point temperature of sintering	SV6

2.2. Temporal convolutional network sequence-to-sequence (TCN-StS) Model

The basic temporal convolutional network is a one-dimensional fully convolutional network with zero padding applied to make sure that the output sequence has the same length as the input sequence. To keep the convolution operation causal, which means for every i in {0, …, input_length − 1}, the i-th element of the output sequence only depends on the elements of the input sequence with indices {0, …, i}, zero-padding is applied only on the left side of the input tensor.

Nevertheless, it is very challenging to apply basic causal convolution directly to long-term sequence problems due to its ability to look back only in a linear order in the depth of the network. Dilated convolution, which enables an exponentially large receptive field, eliminates this problem. Formally, for a one-dimensional sequence input $x \in R^n$ and a filter $f: \{0, \ldots, k-1\} \to R$, the dilated convolution operation F on elements s of the sequence is defined as Eq. (1).

$$F(s) = (\mathbf{x} *_d f)(s) = \sum_{i=0}^{k-1} f(i) \cdot \mathbf{x}_{s-d \cdot i} \tag{1}$$

where d denotes the dilation factor, k is the filter size, and $s - d \cdot i$ accounts for the direction of the past.

Besides, to avoid the gradient exploding/vanishing problems in deep neural networks, residual blocks with skip connections initially designed in ResNet (He et al., 2015) are used in TCN. The skip connection is a branch leading out to a series of transformations $\boldsymbol{F}$, whose outputs are added to the input $\mathbf{x}$ of the block as Eq. (2).

$$o = \text{Activition}(\mathbf{x} + (\boldsymbol{F}(\mathbf{x})) \tag{2}$$

Within a residual block, two layers of dilated causal convolution, rectified linear unit (ReLU) activation, weight normalization, and spatial dropout are stacked. Figure 2(a) illustrates the TCN architecture.

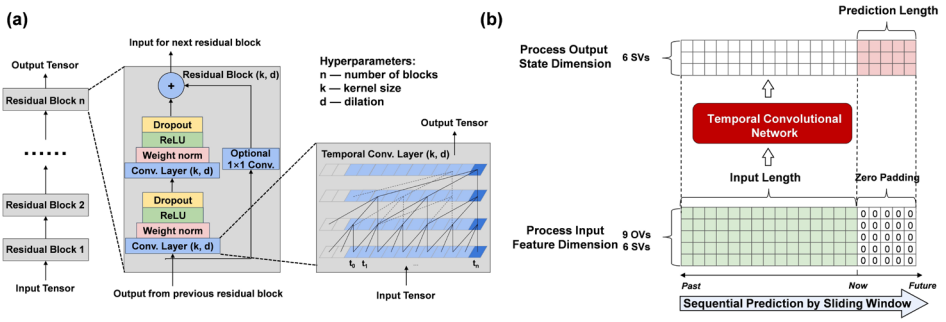


Figure 2. Proposed TCN-StS model. (a) Temporal convolutional network (TCN) architecture. (b) Sequence-to-sequence prediction demonstration.

A sequence-to-sequence prediction manner is proposed in Figure 2(b). At time T, a sequence of shape (input length, feature dimensions) are combined with zero values of shape (output length, feature dimensions). This sequence represents the history from time (T − input length) to T. The model output is a sequence of shape (output length, output dimension) predicting the process state from time $T+1$ to $(T+\text{output length})$. Formally, the TCN-StS model produces the mapping as Eq. (3).

$$\hat{\mathbf{y}}_{T+1}, \hat{\mathbf{y}}_{T+2}, \cdots, \hat{\mathbf{y}}_{T+\text{out\_len}} = \text{TCN\_StS}\left(\mathbf{x}_{T-\text{in\_len}}, \cdots, \mathbf{x}_{T-1}, \mathbf{x}_{T}\right) \tag{3}$$

3. Result and discussion

The dataset is split into 70 % training, 10 % validating, and 20 % testing. Each feature of the original dataset is standardized separately. The input feature dimension is 15 containing 9 OVs and 6 SVs while the output dimension is 6. An input length of 40 minutes is set according to the time delay of the sintering process and predictions are made by the TCN-StS model for the time length of 5, 10, 15, and 20 minutes.

Figure 3(a) presents a snapshot of sequential prediction for SV6. The prediction sequence shows good coincidence with the true sequence, especially at shorter prediction lengths such as 5 minutes and 10 minutes. The predictions shift away from the true values at longer prediction lengths. Two canonical recurrent neural networks, RNN and LSTM are adopted for comparison. The mean squared errors (MSEs) and mean absolute errors (MAEs) of the three models are given in Table 2. The TCN-StS model has lower MSEs and MAEs at all prediction lengths. The results indicate TCN-StS model outperforms recurrent neural networks at the sintering process sequential prediction.

Table 2. Sequential prediction accuracy of RNN, LSTM, and TCN-StS model.

Model	5 min		10 min		15 min		20 min	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
RNN	0.32	0.38	0.46	0.48	0.55	0.54	0.62	0.57
LSTM	0.30	0.37	0.38	0.44	0.50	0.52	0.54	0.55
TCN-StS	**0.26**	**0.33**	**0.36**	**0.40**	**0.41**	**0.44**	**0.46**	**0.47**

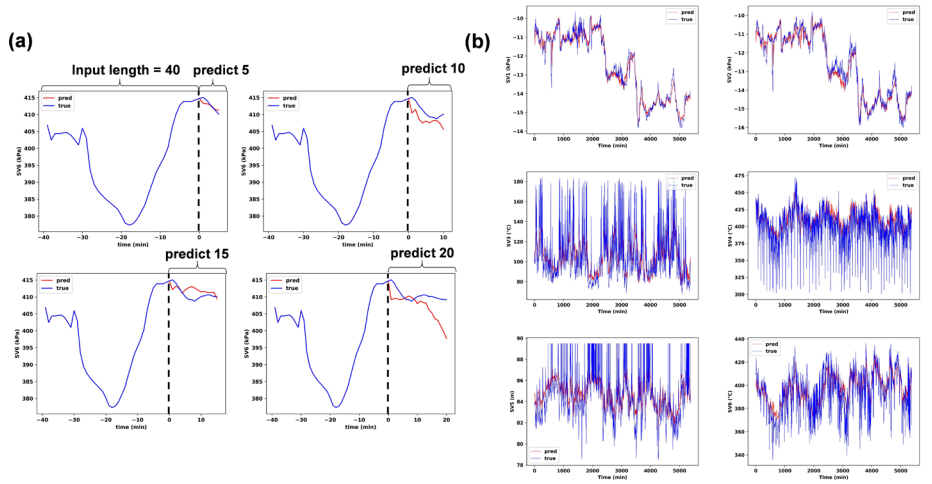


Figure 3. Prediction results. (a) A sequential prediction snapshot for SV6 at 5, 10, 15, and 20 minutes prediction length. (b) Prediction results at 10 minutes time points in the 10 minutes prediction sequence.

To better present the prediction performance over time, prediction results for each time point are extracted separately from the sequence. Figure 3(b) shows prediction results at 10 minutes time points in the 10 minutes prediction sequence. Prediction accuracies for each point in 10 minutes are shown in Figure 4. It can be found that as the prediction length increases, the prediction accuracy of all three models will become worse. For example, the TCN-StS model, with the average MSE of 0.36 and MAE of 0.40, shows the MSE and MAE of 0.24 and 0.33 at 1 min point, and 0.43 and 0.45 at 10 min point, respectively. The same growing trends are also found in RNN and LSTM models. Nevertheless, the TCN-StS model still has lower prediction errors than RNN and LSTM models at nearly all given time points, showing good robustness over time.

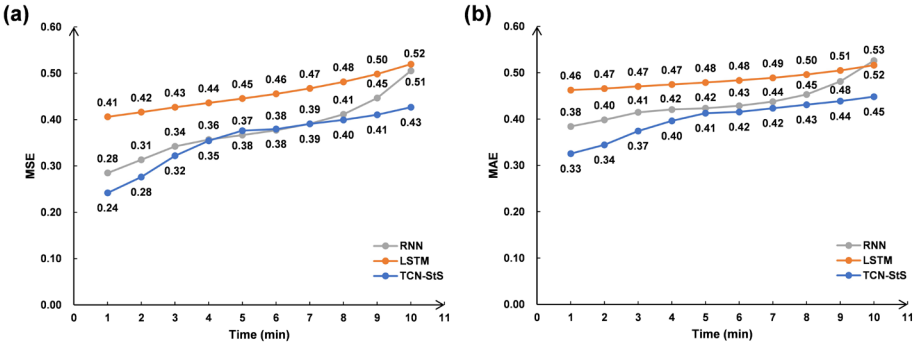


Figure 4. Prediction accuracy for each time point in the sequence. (a) Mean squared errors (MSEs). (b) Mean absolute errors (MAEs).

4. Conclusions

This paper designed a new convolutional-based sequence-to-sequence model architecture for continuous manufacturing process sequential prediction. Comparted to recurrent

neural networks, the proposed TCN-StS model demonstrates better prediction accuracy at all given time lengths on an actual industrial dataset as well as a robust prediction capability over time. This study addresses the effectiveness of convolutional networks for sequence modelling and gives insights into utilizing sequence-to-sequence modelling as an effective simulation tool for constructing digital twins in the continuous manufacturing process.

Acknowledgment

The authors gratefully acknowledge Technological Innovation 2030—"New Generation Artificial Intelligence" Major Project (2018AAA0101605) for its financial support.

References

Bai, S., Kolter, J.Z., Koltun, V., 2018. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. arXiv:1803.01271 [cs].

Chou, C.-H., Wu, H., Kang, J.-L., Wong, D.S.-H., Yao, Y., Chuang, Y.-C., Jang, S.-S., Ou, J.D.-Y., 2020. Physically Consistent Soft-Sensor Development Using Sequence-to-Sequence Neural Networks. IEEE Transactions on Industrial Informatics 16, 2829–2838. https://doi.org/10.1109/TII.2019.2952429

Glaessgen, E., Stargel, D., 2012. The Digital Twin Paradigm for Future NASA and U.S. Air Force Vehicles, in: 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference. American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/6.2012-1818

Gockel, B., Tudor, A., Brandyberry, M., Penmetsa, R., Tuegel, E., 2012. Challenges with Structural Life Forecasting Using Realistic Mission Profiles, in: 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference. American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/6.2012-1813

He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep Residual Learning for Image Recognition. arXiv:1512.03385 [cs].

Kang, J.-L., Mirzaei, S., Lee, Y.-C., Chuang, Y.-C., Frias, M., Chou, C.-H., Wang, S.-J., Wong, D.S.H., Jang, S.-S., 2021. Digital Twin Model Development for Chemical Plants Using Multiple Time-Steps Prediction Data-Driven Model and Rolling Training, in: Computer Aided Chemical Engineering. Elsevier, pp. 567–572. https://doi.org/10.1016/B978-0-323-88506-5.50090-5

Lea, C., Vidal, R., Reiter, A., Hager, G.D., 2016. Temporal Convolutional Networks: A Unified Approach to Action Segmentation. arXiv:1608.08242 [cs].

Rosen, R., von Wichert, G., Lo, G., Bettenhausen, K.D., 2015. About The Importance of Autonomy and Digital Twins for the Future of Manufacturing. IFAC-PapersOnLine 48, 567–572. https://doi.org/10.1016/j.ifacol.2015.06.141

Zhou, K., Chen, X., Wu, M., Cao, W., Hu, J., 2019. A new hybrid modelling and optimization algorithm for improving carbon efficiency based on different time scales in sintering process. Control Engineering Practice 91, 104104. https://doi.org/10.1016/j.conengprac.2019.104104

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50299-2

Surrogate modeling for mixed refrigerant streams in the refrigeration cycle of an LNG plant

Aisha A. Al-Hammadi[a,b], Robert E. Franzoi[a], Omar E. Ibrahim[a], Brenno C. Menezes[a*]

[a]*Division of Engineering Management and Decision Sciences, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, QATAR*
[b]*Operations Support Department, Qatargas, Doha, QATAR*
bmenezes@hbku.edu.qa

Abstract

Given the importance of liquefaction processes in the LNG value chain, it is necessary to model the complexity of such process. A key stage is the mixed refrigerant (MR) cycle used to liquify the natural gas in the liquefaction plant. The MR refrigeration cycle consists of compressors and heat exchangers in different compression stages that affect the MR properties in terms of temperature and pressure. In this work, the use of surrogate models is addressed for the compressor's power consumption and efficiency formulations along with the heat exchanger's performance in terms of heat duty after each compression stage. A training data set containing 500 points is used for building the surrogates, while a testing data set of 500 points verifies their accuracy. The surrogates built herein are shown to be sufficiently accurate to be further employed in decision-making industrial applications such as simulation, optimization, and control.

Keywords: Surrogate modeling, refrigeration cycle, machine learning, liquefaction.

1. Introduction

The natural gas liquefaction plant is the most essential and critical process in the liquefied natural gas (LNG) value chain. In this process, the natural gas is cooled and liquified to a cryogenic temperature of around -162 ◦C. To achieve good quality of liquified natural gas, a comprehensive understanding of the process is required for improved design and operations of liquefaction plants. A key component in such processes is the mixed refrigerant (MR) stream, which plays a significant role in precooling, subcooling, and liquifying the natural gas to produce LNG. The MR streams undergo certain processing in the refrigeration cycle that consists mainly of refrigerant compressors and cooler type heat exchangers. In this work, a MR network is considered, in which propane (C3) and MR (C1, C2, C3, and N2) are used to cool and liquify natural gas. The mixed refrigeration process controls the LNG production outputs, while the propane refrigeration system aims to provide cooling to the MR (Mokhatab et al., 2014). Since the MR is precooled and liquified using propane during the process, the criticality of this operation arises from maintaining optimal MR temperature in the refrigeration cycle. Very high temperature of MR results in lower LNG production. A main reason of operating with warm MR is due to the high-power consumption of the compressor and its lower efficiency in addition to the low heat duty of heat exchangers. Therefore, it is important to include details of rigorous liquefaction process to better optimize the power consumption, the compressor efficiency, and the heat duty to produce cooler MR, and hence, more efficient LNG production. However, given its thermodynamics complexity, it is intractable to introduce

such nonlinearities in optimizable and controllable environments. Therefore, additional methods are required for achieving simplified correlations that simultaneously provide good accuracy and can be further embedded in decision-making problems.

Recently, there has been an extensive use of surrogate modeling approaches for predicting the behavior of complex systems to provide improved computational tractability and avoid convergence issues in decision-making problems (Franzoi et al., 2021a). This work aims to build surrogate models from supervised simulated data of an MR refrigeration cycle unit that calculate the power consumption, efficiency of compressors, and heat duty of heat exchangers based on variable inputs. This method relies on three steps: 1) define base functions to be selected by a coefficient-setup approach that minimizes the variable outputs as the difference between $y_{predicted}$ and $y_{experimental}$; 2) the design of experiments including the sampling method and the data generation approach; and 3) the regression or identification method to determine the selected base functions and their respective coefficients (Tran and Georgakis, 2018; Hullen et al., 2020). To build the surrogate model, the data shall be accurate and can be determined in rigorous simulation software, collected from experiments in the plant, etc. Moreover, a proper balance between the quality of the model accuracy and the computational effort must be considered depending on the processing time and resources for a certain application (Mencarelli et al., 2020).

In this paper, a surrogate modeling approach is used to handle the compressor's power consumption and efficiency formulations as well as the heat exchanger performance after each compression stage in the liquefaction of natural gas, in which complex nonlinear equations are transformed into simpler bilinear and trilinear equations. The surrogate models correlate variations of independent X variables to dependent Y variables, which can be used to model different types of complex and unknown systems.

2. Problem Statement

The type of problem addressed herein concerns the cooling of mixed refrigerant, used to cool and liquefy the LNG in the main cryogenic heat exchanger. MR is widely employed within the natural gas liquefaction. The complexity of this process arises from the nonlinear terms that calculate the compressors efficiency and power and heat exchangers heat duty. The case study proposed in this work addresses the MR refrigeration cycle that undergoes three compression stages with intermediate cooling processes. Figure 1 illustrates the MR refrigeration network, in which the MR flow is recycled. It exits the main cryogenic heat exchanger (MCHE) top side in gaseous form and low pressure; throughout the network, the MR pressure is increased, liquefied in the C3 refrigeration cycle, and enters the MCHE again for cooling and liquefying the natural gas.

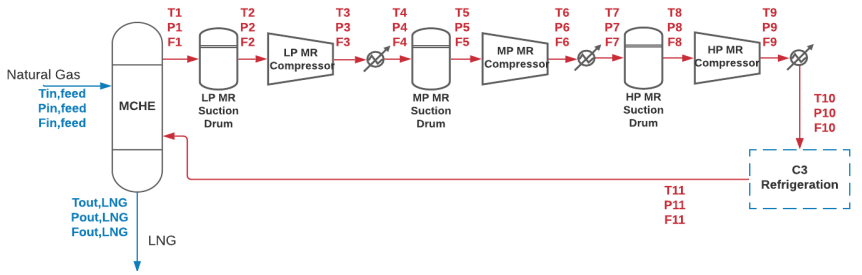


Figure 1: MR refrigeration cycle network.

This type of process is nonlinear and nonconvex because of the fluctuating temperatures and pressures after each stage and the requirement of calculating related variables such as the density and heat capacity for each temperature and pressure points. For each compression stage, Eqs.(1) and (2) are used to calculate the efficiency and power consumption for the compressors and the heat duty for the heat exchangers. The variables T_{i-1}, T_i, and T_i' are the inlet, outlet, and ideal discharge temperatures from the compressors, respectively, and η_{cj} is the compressor's efficiency. The indices i and j respectively belong to the stream flows (before and after each compressor (SFC)) and to the compression stage (CS). The compressor power consumption is calculated using Eq.(3), in which W_j is the power consumption for each compression stage and C_{pj} is the heat capacity for each compression stage.

$$T_i' = T_{i-1} \times (r_p^{(\frac{y-1}{y})}) \qquad \forall\ \mathrm{i} \in \mathrm{SFC} \tag{1}$$

$$\eta_{cj} = \frac{T_i' - T_{i-1}}{T_i - T_{i-1}} \qquad \forall\ \mathrm{j} \in \mathrm{CS}, \mathrm{i} \in \mathrm{SFC} \tag{2}$$

$$W_j = C_{pj}(T_i - T_{i-1}) \qquad \forall\ \mathrm{j} \in \mathrm{CS}, \mathrm{i} \in \mathrm{SF}C \tag{3}$$

The heat duty for each heat exchanger is calculated using Eqs.(4) and (5).

$$V_{actual} = \left[\frac{P_{normal} \times V_i}{T_{normal}}\right] \times \frac{T_{i-1}}{P_{i-1}} \qquad \forall\ \mathrm{i} \in \mathrm{SFH} \tag{4}$$

$$Q_j = \left[\rho_j \times V_{actual}\right] \times C_{pj} \times [T_i - T_{i-1}] \qquad \forall\ \mathrm{j} \in \mathrm{CS}, \mathrm{i} \in \mathrm{SFH} \tag{5}$$

In which Q_j is the heat duty for each compression stage j, ρ_j is the density for each compression stage, V_i is the flow at each stream flow i before and after each heat exchanger (SFH), while $P_{normal} = 101{,}325\ PA$ and $T_{normal} = 293\ K$.

3. Methodology

The main purpose of this paper is to build surrogate models that can effectively replace the complex thermodynamic equations. The surrogates are built to represent three dependent variables of interest Y, which are the properties of the compressors and heat exchangers shown in Figure 1 and represented by Eqs.(2), (3), and (5). Independent-dependent variations correlate as $Y_i = f(X_j)$. The input variables considered to form the surrogates represent the stream properties (i.e., temperature, pressure, and flow). When building surrogate models, it is important to consider some important aspects concerning the required data for training and testing, functional form of the surrogates and proper evaluation to verify their performance. A simplified framework methodology proposed by Franzoi et al.(2021b) is used in this work, as shown in Figure 2.

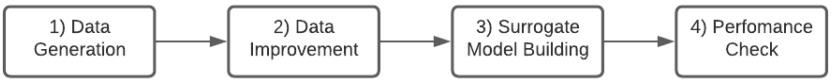


Figure 2: Framework for the surrogate model building strategy (adapted from Franzoi et al. 2021b).

1) *Data Generation*: Latin Hypercube Sampling (LHS) method is used to generate the required data for training the surrogates. This generates random sample points for the input (independent variables) dataset that includes temperatures, pressures, and flows. There are boundaries for each independent variable in each compression stage based on experiments. The independent dataset is used to calculate the output values (dependent variables) from compressors and heat exchangers related formulas in Eqs.(2), (3), and (5) for each sample point. Then, a complete X-Y input-output data set is constructed to estimate the behavior of the surrogates. The dataset is equally split into training and testing data. The former focuses on building the surrogates by identifying their bases and coefficients, while the latter is used to fairly evaluate the accuracy and performance of the surrogates.

2) *Data Improvement*: This step is performed to avoid any biased surrogates or numerical issues due to the different units and magnitudes used in the model, mostly because variables that are too small or too large can affect obtaining the accurate and reliable coefficients. Hence, data normalization is used to normalize all training input and output data sets for each variable using Eq.(6), which results in minimum and maximum values to be within 0 to 1 range.

$$x_{jp} = \frac{x_{jp} - min(x_j^{tr})}{max\left(x_j^{tr}\right) - min(x_j^{tr})} \qquad \forall\ p\ \in N_{training} \tag{6}$$

In which the minimum and maximum values of the training data x_j correspond to min (x_j^{tr}) and max$\left(x_j^{tr}\right)$.

3) *Surrogate Model Building*: To estimate the behavior of the MR refrigeration cycle processes for the compressors and heat exchangers, trilinear surrogate models are used, as shown in Equation (7) to calculate the dependent variable Y_i.

$$\begin{aligned} Y_i = I_i + \sum_{j\,\in\,DV} b_{ij}\,X_j + \sum_{j\,\in\,IV}\sum_{k\,\in\,IV} c_{ijk}\,X_jX_k \qquad & \forall\ i\ \in DV, \\ + \sum_{j\,\in\,IV}\sum_{k\,\in\,IV}\sum_{n\,\in\,IV} t_{ijkn}\,X_jX_kX_n \qquad & \forall\ j \le k, \\ & \forall\ k \le n \end{aligned} \tag{7}$$

In which I_i is the intercept of each point i within dependent variables (DV), and b_{ij}, c_{ijk}, and t_{ijkn} are the coefficients to be determined or estimated during the building process of the surrogate model by evaluating the accuracy of the input-output data to achieve the target of minimizing the prediction error for each independent variable point (IV). This optimization target is presented in Eq.(8) for minimizing the least-squares error (LSE).

$$Minimize \quad \sum_{p=1}^{n_{training}} (y_{ip} - Y_{ip})^2 \tag{8}$$

In which y_{ip} is the actual calculated value using thermodynamics equations for the variable i at each point p within the training data set $n_{training}$, whereas Y_{ip} represents the calculated values using the surrogate model estimated coefficients for each independent variable i at each point p.

4) Performance Check: The last step in the framework is to evaluate the surrogates by employing the identified coefficients from the surrogate model. This is done by using different set of points from the calculated ones previously, which are referred to as testing data set. Testing data sets are the points that were not used in training the surrogate, which can identify the accuracy of the predictions. The evaluation is done via calculating the mean absolute error using the testing data that indicate the surrogate accuracy.

4. Results

The surrogate models built are tested for three compression stages in the MR refrigeration cycle including compressors and heat exchangers, as shown in Figure 1. The size of the dataset considered in the surrogates is set to be $N \in \{1000\}$, which is equally split between training and testing data sets. Bilinear and trilinear surrogates are built for each compressor and heat exchanger. For the compressors, surrogates are built testing the accuracy of efficiency and power consumption equations, whereas heat duty is considered for the heat exchangers. The mean absolute error is calculated for each system to illustrate the difference between the calculated and predicted values, as shown in Table 1. Trilinear surrogates provide higher accuracy, mostly because the higher amount of coefficients and higher predictability power given their complexity. The methodology employed herein has proved its efficiency in terms of accuracy for nonlinear and complex systems, as those surrogates can successfully be employed in replacement of the thermodynamic equations and can be potentially used in further optimization and decision-making processes.

Table 1: Mean absolute error (MAE) for bilinear and trilinear surrogates.

Bilinear							
		LP- MAE		MP- MAE		HP- MAE	
System	SM	Training	Testing	Training	Testing	Training	Testing
Compressors	Y1- Efficiency	$2.83\ 10^{-3}$	$7.48\ 10^{-3}$	$1.73\ 10^{-2}$	$3.35\ 10^{-2}$	$2.11\ 10^{-2}$	$5.28\ 10^{-2}$
	Y2- Power	$1.22\ 10^{-5}$	$1.77\ 10^{-5}$	$2.00\ 10^{-3}$	$2.18\ 10^{-6}$	$2.63\ 10^{-6}$	$2.68\ 10^{-5}$
Heat Exchangers	Y3- Heat Duty	$9.45\ 10^{-4}$	$9.12\ 10^{-4}$	$8.14\ 10^{-4}$	$7.04\ 10^{-4}$	$8.14\ 10^{-4}$	$7.04\ 10^{-4}$
Trilinear							
		LP- MAE		MP- MAE		HP- MAE	
System	SM	Training	Testing	Training	Testing	Training	Testing
Compressors	Y1- Efficiency	$2.03\ 10^{-4}$	$5.64\ 10^{-4}$	$3.26\ 10^{-3}$	$7.68\ 10^{-3}$	$3.91\ 10^{-3}$	$1.18\ 10^{-2}$
	Y2- Power	$2.92\ 10^{-6}$	$2.94\ 10^{-6}$	5.7110^{-7}	$1.52\ 10^{-6}$	$3.72\ 10^{-7}$	$2.81\ 10^{-6}$
Heat Exchangers	Y3- Heat Duty	$1.73\ 10^{-4}$	$1.64\ 10^{-4}$	$1.55\ 10^{-4}$	$1.33\ 10^{-4}$	$3.12\ 10^{-4}$	$2.77\ 10^{-4}$

5. Conclusion

The cooling and liquefaction of natural gas involves complex networks that require the use of refrigerants within a rigorously controlled environment. Those refrigerants shall meet certain temperature and pressure points via undergoing cooling cycle, called refrigeration cycle. This work considered the MR refrigeration cycle with three compression stages, LP, MP, and HP. This cycle consists mainly of compressors and heat exchangers, whereby such a system is studied considering certain variables of interest associated with the compressors, such as the efficiency and power consumption, in addition to the heat duty for heat exchangers. Complex and nonlinear thermodynamics equations are typically used to calculate those variables. However, this work addresses the use of the surrogate models for the purpose of simplifying the formulas into bilinear and trilinear terms, which can open opportunities for integrating such complex modeling systems with further optimization decision-making approaches. The methodology employed uses a surrogate model building framework comprised of four steps. First, the input dataset is generated experimentally from the Latin Hypercube Sampling method, which is utilized to calculate

the required output variables using the first principles thermodynamics equations. Then, datasets are enhanced by normalization, in which the data points are normalized to lie between 0 and 1 aiming to avoid numeric issues. The third step builds the surrogates using input-output datasets and generates bilinear and trilinear coefficients used to calculate the required variables. Finally, performance check is carried out to determine the accuracy of the surrogates using testing datasets.

Depending on the complexity of the equations to be predicted, bilinear surrogates can be sufficiently accurate. However, trilinear surrogates typically provide significant higher accuracy than their bilinear counterparts. Overall, both provide good efficiency and accuracy for replacing the thermodynamics equations and can be further utilized for simulation, control, and optimization cases. The proposed surrogate modeling approach builds a data-driven model that can be used in different optimization applications, such as in: a) optimizing the power consumption of compressors considering certain design boundaries for each parameter (e.g., allowable temperature and pressure for the compressors); b) optimizing the efficiency of the compressors, which lead to colder MR after the C3 refrigeration cycle resulting in a higher production of LNG with lower temperature; and c) optimizing the heat exchanger duty after each compression stage aiming to achieve better usage of MR in terms of liquifying the natural gas without requiring makeup MR to provide a better cooling and liquefaction to the natural gas. Such approach would allow the utilization of optimization-based tools instead of their simulation counterparts that rely on complex thermodynamics equations.

References

R. E. Franzoi, T. Ali, A. Al-Hammadi, B. C. Menezes, 2021b, Surrogate Modeling Approach for Nonlinear Blending Processes, 1st International Conference on Emerging Smart Technologies and Applications (eSmarTA), 1-8.

R. E. Franzoi , J. D. Kelly, B. C. Menezes, C. L. E. Swartz, 2021a, An adaptive sampling surrogate model building framework for the optimization of reaction systems. Computers and Chemical Engineering, 152, 107371.

G. Hullen, J. Zhai, S. H. Kim, A. Sinha, M. J. Realff, F. Boukouvala, 2020, Managing uncertainty in data-driven simulation-based optimization, Computers and Chemical Engineering, 136, 106519.

L. Mencarelli, A. Pagot, P. Duchêne, 2020, Surrogate-based modeling techniques with application to catalytic reforming and isomerization processes, Computers and Chemical Engineering, 135, 106772.

S. Mokhatab, J. Y. Mak, J. V. Valappil, D. A. Wood, 2014, In Handbook of liquefied natural gas, pp. 297-300, Gulf Professional Publishing.

D. P. Penumuru, S. Muthuswamy, P. Karumbu, 2020, Identification and classification of materials using machine vision and machine learning in the context of industry 4.0. Journal of Intelligent Manufacturing, 31(5), 1229–1241.

A. P. Tran, C. Georgakis, 2018, On the estimation of high-dimensional surrogate models of steady-state of plant-wide processes characteristics, Computers and Chemical Engineering, 116, 56–68.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50300-6

Prediction for heat deflection temperature of polypropylene composite with Catboost

Chonghyo J.[a,b] , Hyundo P.[a,c] , Seokyoung H.[b] , Jongkoo L.[d] , Insu H.[d] , Hyungtae C.[a] , Junghwan K.[a,*]

[a] *Green Materials and Processes R&D Group, Korea Institute of Industrial Technology, Ulsan 44413, Republic of Korea*
[b] *Department of Chemical Engineering, Konkuk University, Seoul 05029, Republic of Korea*
[c] *Department of Chemical and Biomolecular Engineering, Yonsei University, Seoul 03722, Republic of Korea*
[d] *Research & Development Center, GS Caltex Corporation, Daejeon 34122, Republic of Korea*

kjh31@kitech.re.kr

Abstract

Recently, polypropylene composites (PPCs) are in the spotlight because of their versatilities in composite industries. Properties of PPCs are determined by numerous physical property values (PPV), among which heat deflection temperature (HDT), polymer's resistance to distortion, is a key indicator. However, enormous trial and error is required to produce PPCs with desired PPV because there is no theoretical equations between material composition and PPV. Hence, to reduce the cost and time of finding material composition to meet the desired PPV, we proposed a machine learning-based PPV prediction model. However, some categorical data which can have an influence on the prediction model performance are included in the dataset, because some of data were from repeated experiments. Therefore, algorithm case study (Multiple linear regression (MLR), XGBoost, and CatBoost) was conducted to develop the optimal HDT prediction model which could process the normal data as well as the categorical data. The performances of each prediction model were evaluated with R^2 and RMSE. As a result, the CatBoost-based HDT prediction model was proposed as the optimal model to solve the trial and error problem.

Keywords: PP composites; Categorical data; Machine learning; Catboost

1. Introduction

In recent years, polypropylene composites (PPCs) have been highlighted owing to their versatility in composite industries. PPCs exhibit excellent physical properties, such as high strength, light weight, and high impact resistance. These excellent physical property values are afforded by the addition of additives that can improve the physical properties of polypropylene (PP). For example, fillers such as talc are added to improve the rigidity

and toughness of PP, whereas rubber is added to improve its flexibility and ductility. Hence, because the physical properties depend on the type and composition of PP and the additives, i.e., the "recipe," before synthesizing PPCs, the appropriate recipe should be selected based on the target physical property values.

Among the various physical properties of PPC products, specific heat deflection temperature (HDT) is one of the key indicators in the design of PPC products. HDT provides an indication of the temperature at which materials begin to soften when exposed to a fixed load at elevated temperatures. Hence, HDT is an important physical property in PPC applications because it allows engineers to determine the temperature limit above which the material is not appropriate for a structure. Despite the importance of the HDT, achieving the required HDT remains challenging. The required HDT for applications is achieved via numerous trials and errors because the HDT cannot be calculated using the recipe before synthesizing the PPC specimen and testing the HDT. The numerous experiments revealed time-and cost-consuming problems that should be solved to ensure the efficiency of the PPC development process.

As an alternative to the trial-and-error approach, machine learning (ML) has been proposed as it can reduce the number of trials and errors based on the use of a data-driven model that can predict specific values. The data-driven model extracts the relationship between the input and output data, analyzes the relationship, and predicts the output data using the input data. By predicting the output values, the number of trials and errors can be reduced as additional experiments are not required. However, one of the most significant problems in ML is that the prediction performance of a data-driven model depends on the data quality. For instance, if some data have the same input values with different output values in the dataset, then the predictive performance of the model will be low because of overfitting; this is because most ML algorithms replace different output values with the mean value in the categorical data. Among many ML-based algorithms, the CatBoost algorithm was developed to manage categorical data in data-driven modeling. Unlike other regression algorithms, CatBoost can manage categorical data without overfitting by considering various values instead of replacing the output values with the mean value.

Herein, we propose a CatBoost-based model for HDT prediction to reduce the number of trials and errors in the PPC development process and to solve overfitting by categorical data. First, we discovered that some categorical data existed in the dataset because the HDT values depended on the experimental environment, such as temperature, person, machine, and humidity. The categorical data were extracted by comparing them with a dimensionless number "*A*," which we defined. Second, three data-driven models were developed using multiple linear regression (MLR), XGBoost, and CatBoost, separately, to compare their predictive performances. Finally, their R^2 and RMSE were compared to identify the best data-driven model for HDT prediction.

2. Method

2.1. Categorical data treatment

Because some categorical data existed in the recipe dataset, a method was suggested to analyze them. The method comprised two steps: First, "same recipes" which imply the categorical data in a recipe dataset, are defined; second, a dimensionless number, denoted

as "*A*," is calculated to obtain the distribution of the same recipes. The same recipes were detected in two steps, i.e., encoding and detecting, as shown in Figure 1. Because the recipes contain information regarding the materials and their weight percentages, it was challenging to count the number of same recipes. Therefore, encoding to assign codes to the recipes was performed to count the number of same recipes. In this step, all of the recipes were assigned codes that contained information regarding the weight percentage of the materials. By comparing the codes, the same recipes were detected, as shown in Figure 1.

	P001	P002	…	F001	F002	…	R001	R002	…	OTH1	OTH2	…	OTH9
Recipe A	0	70	…	15	0	…	15	0	…	0	0	…	0
Recipe B	0	30	…	35	0	…	35	0	…	0	0	…	0
Recipe C	5	15	…	15	15	…	20	10	…	5	0	…	15
Recipe D	5	15	…	15	15	…	20	10	…	5	0	…	15
Recipe C	5	5	…	30	10	…	15	15	…	10	0	…	10

Encoding for counting same recipes ⬇

	P001	P002	…	F001	F002	…	R001	R002	…	OTH1	OTH2	…	OTH9
Recipe A	070…150…150…00…0												
Recipe B	030…350…350…00…0												
Recipe C	515…1515…2010…50…15												
Recipe D	515…1515…2010…50…15												
Recipe C	55…3010…1515…100…10												

⬇

	P001	P002	…	F001	F002	…	R001	R002	…	OTH1	OTH2	…	OTH9
Recipe A	070…150…150…00…0												
Recipe B	030…350…350…00…0												
Recipe C	515…1515…2010…50…15												
Recipe D	515…1515…2010…50…15												
Recipe C	55…3010…1515…100…10												

Same recipe

Figure 1 Procedures to detect and count same recipes in dataset

After detecting the same recipes, the HDT distribution of the same recipes was obtained by calculating the dimensionless number "*A*" using Eq. (1). Subsequently, "*A*" was used to compare the differences in the HDT for the same recipe dataset. Using the minimum HDT in the same recipe as the denominator and the HDT of a recipe as the numerator, the differences in the same recipes can be obtained, as listed in Table 1.

$$A = \frac{HDT\ of\ the\ recipe}{The\ minimum\ HDT\ of\ same\ recipes} \tag{1}$$

Table 1 Example of calculating *A* from the same recipe

	Materials (wt%)				HDT	*A*
	P006	**R013**	**OTH1**	**OTH7**	**(℃)**	
Recipe 369	87.977	9.775	1.955	0.293	79.5	1.174 (79.5/67.7)
Recipe 378	87.977	9.775	1.955	0.293	80.4	1.188 (80.4/67.7)
Recipe 617	87.977	9.775	1.955	0.293	67.7	1.000 (67.7/67.7)
Recipe 679	87.977	9.775	1.955	0.293	79.2	1.170 (79.2/67.7)

2.2. Catboost

Unlike other regression algorithms, because CatBoost uses ordered encoding instead of mean encoding, it can solve overfitting when a categorical dataset are used for data-driven modeling. The two encoding methods for categorical data are presented in Figure 2. The regression algorithms using mean encoding replace all categorical data with the mean value of the data. Overfitting is incurred in this process because the mean value is proposed as the criterion when the loss of the regression is calculated in every iteration. By contrast, CatBoost, which uses ordered encoding, proposes various mean values that are calculated using random samples in a categorical dataset. Using different loss criteria in each iteration, regression can be generalized.

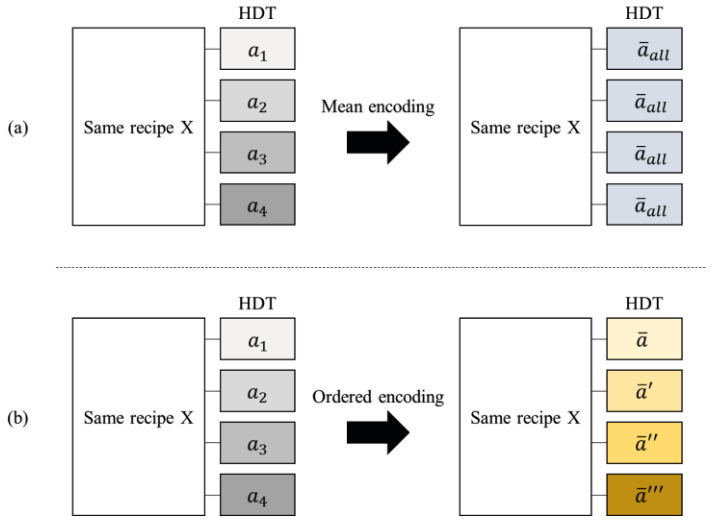


Figure 2 Procedure to process categorical data: (a) mean encoding and (b) ordered target encoding

3. Results and discussions

3.1. Dimensionless number "A" for categorical data treatment

By encoding to count the number of same recipes, 199 recipes among 993 recipes were detected. Because each recipe belonged to a different categorical data group and the groups had different HDT ranges, a criterion to normalize the categorical values was

required to visualize the differences in the HDT. Therefore, a dimensionless number "*A*" was calculated in this study. The calculation results are shown in Figure 3. As shown, the minimum and maximum *A* values were from 1 to 2 which implies that some recipes have twice as large HDT values in the same recipe groups at the maximum. In the same recipe groups, 59 recipes were randomly selected and then used as the test dataset to compare the predictive performances of the three ML algorithms in Section 3.2. The remaining recipes were segregated for data-driven modeling, and the ratio of the segregated data is shown in Table 2.

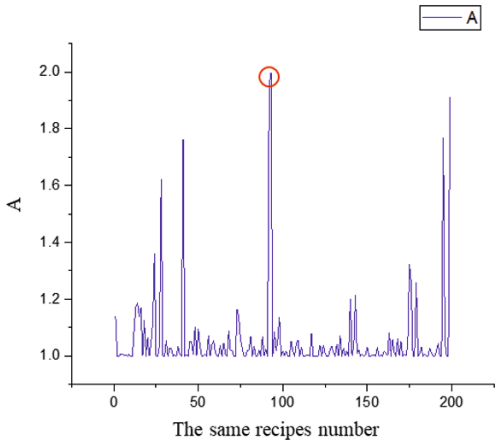


Figure 3 Calculation result of *A*

Table 2 Ratio of data segregated for data-driven modelling

	Number of recipes	Number of the same recipes
Train dataset	638	119
Validation dataset	152	21
Test dataset	203	59
Total	993	199

3.2. Comparison of the three algorithms

In this study, three different algorithms (MLR, XGBoost, and CatBoost) were used to perform data-driven modeling to predict HDT values. MLR is a simple linear regression algorithm that uses mean encoding. XGBoost and CatBoost are classification and regression tree (CART)-based nonlinear regression algorithms based on the boosting method. However, CatBoost uses ordered encoding, whereas XGBoost uses mean encoding.

The predictive performances of the three models are shown in the evaluation results in Figure 5. The R^2 and RMSE were calculated to compare the model performance quantitatively. We discovered that the MLR-based model yielded the lowest R^2 (0.8162) and the highest RMSE (9.7934) when all the recipes in the test dataset were tested. However, the XGBoost-based model indicated a higher R^2 (0.8578) and a lower RMSE (8.7007) than the MLR-based model. Meanwhile, the CatBoost-based model yielded the highest R^2 (0.8965) and lowest RMSE (7.3477). Moreover, the R^2 and RMSE were calculated for only 59 same recipes. The MLR-based model indicated the lowest R^2

(0.8256) and the highest RMSE (7.8899) for 59 same recipes. By contrast, the XGBoost-based model indicated a higher R^2 (0.9690) and a lower RMSE (2.6105) than the MLR-based model. Meanwhile, the CatBoost-based model indicated the highest R^2 (0.9801) and lowest RMSE (2.6105). This shows that the CART-based nonlinear regression algorithms are more appropriate for the data-driven modeling of HDT prediction than the linear regression algorithm, and that the CatBoost-based model can predict the HDT with better performance than the XGBoost-based model even when the dataset includes some categorical data with a high value of *A*.

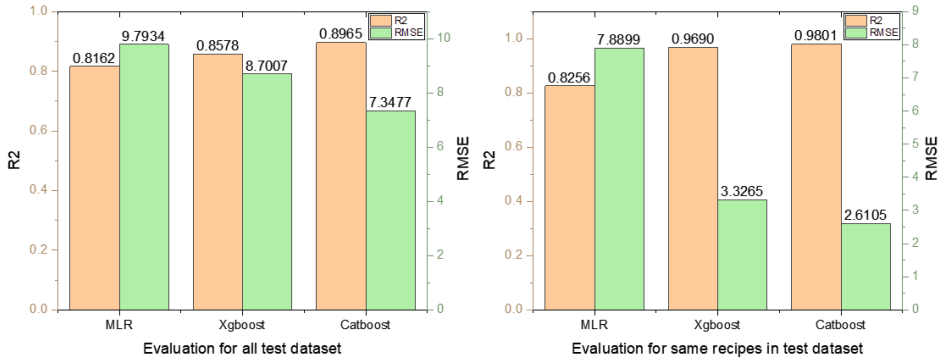


Figure 4 Evaluation result for all recipes in test dataset (left), and for only the same recipes in test dataset (right)

4. Conclusion and future work

In this study, we proposed a CatBoost-based HDT predictive model to reduce the number of trials and errors in the PPC development process. In addition, to detect categorical data, a new approach was proposed, where a recipe is encoded to a code, and the difference in the HDT is calculated for the same recipe group using a dimensionless number "*A*." The results indicated that although the HDT in the same recipes was different, the CatBoost-based predictive model performed better than the MLR and XGBoost-based models. Therefore, if the proposed model is applied to the PPC development process, then the number of trials and errors can be reduced.

In future studies, we will use Shapley additive explanations to further explain the model as well as extend this study to other properties of the PPC.

Acknowledgements

This study has been conducted with the support of the Korea Institute of Industrial Technology as "Development of AI Platform Technology for Smart Chemical Process (kitech JH-21-0005)" and "Development of Global Optimization System for Energy Process (kitech EM-21-0022)"

References

K. Harutun, Impact Behavior of Polypropylene, Its Blends and Composites. Handb. Polypropyl. Polypropyl. Compos. Revis. Expand., CRC Press; 2003, p. 150–213

L. Prokhorenkova, G. Gusev, A. Vorobev, A. Dorogush, A. Gulin, Catboost: Unbiased boosting with categorical features, Adv Neural Inf Process Syst 2018, 6638–43

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50301-8

A New Machine Learning Framework for Efficient MOF Discovery: Application to Hydrogen Storage

Teng Zhou[a,b,*], Zihao Wang[b], Kai Sundmacher[a,b]

[a] *Process Systems Engineering, Otto-von-Guericke University Magdeburg, Universitätsplatz 2, D-39106 Magdeburg, Germany*

[b] *Process Systems Engineering, Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1, D-39106 Magdeburg, Germany*

zhout@mpi-magdeburg.mpg.de

Abstract

Metal-organic frameworks (MOFs) are recognized as promising materials for gas storage and separation due to their structural diversity, high porosity, and tailorable functionality. Considering the large number of possible MOFs, an integrated machine learning framework is proposed to discover promising candidates with desirable adsorption properties. The framework consists of structure decomposition, feature integration, and predictive modelling. Unlike most of the previous studies employing solely structural or geometric descriptors, our method integrates both structural and chemical features of MOFs for adsorption property prediction using the graph convolutional network (GCN) and feed-forward neural network (FNN) approaches. The machine learning framework is first introduced and then applied to hydrogen storage. Promising MOF candidates exhibiting respectable hydrogen storage capacities are successfully identified, which potentially outperform the existing porous materials for hydrogen storage.

Keywords: machine learning, metal-organic framework discovery, graph convolutional network, hydrogen storage

1. Introduction

Metal-organic frameworks (MOFs) are an important type of porous materials with large structural diversity, high porosity, and tailorable functionality. In the past two decades, MOFs have been attracting wide attentions in many applications, especially gas storage and separation (Gándara et al., 2014; Cui et al., 2016; Zhou et al., 2020; Zhang et al., 2021; Chen et al., 2018). Through the combinations of numerous metal nodes and organic linkers under specific topologies, we can in principle synthesize an infinite number of different MOFs. This makes MOF discovery via experimental trail-and-error extremely challenging. High-throughput screening techniques in tandem with molecular simulations or ab-initio calculations are being used to calculate properties of MOFs. Although the grand canonical Monte Carlo (GCMC) simulation has shown remarkable accuracy (Moghadam et al., 2018; Chung et al. 2016) for MOF adsorption property prediction, it is computationally inefficient for finding the best MOFs from a large number of candidates.

Over the past few years, many researchers have applied various supervised machine learning (ML) methods to predict MOF properties (Chong et al., 2020; Altintas et al., 2021; Zhou et al., 2019) from available data. With the established ML models, one can

perform in-silico predictions on a large number of new MOFs and quickly find the best candidates. So far, most of the previous ML works have employed only geometric descriptors, e.g., void fraction, surface area and pore diameters, to correlate MOF adsorption properties (Fernandez et al., 2013; Shi et al., 2020; Yuan et al., 2021; Thornton et al., 2017). Although geometric descriptors largely affect the adsorption performance, chemical diversity of building blocks can also play a crucial role (Moosavi et al., 2020). Unfortunately, given the massive chemical descriptors (e.g., number of atom, atomic charge and dipole moment), applying them for ML requires a lot of domain knowledge and labour for descriptor (or feature) generation and optimal selection. In contrast, representation learning allows the machine to automatically learn important features directly from material structures without any human input.

In this contribution, we propose an end-to-end ML framework where the MOF structure is imported as input, chemical and structural features are learned automatically, and the corresponding adsorption property is finally predicted in one step. The framework is trained with a relatively small amount of GCMC-derived data and the resulting ML model can then be used for a fast MOF discovery from a much larger set of candidates. The proposed method is applied to hydrogen storage aiming to find MOFs with superior volumetric storage capacity.

2. Methods

Prior to GCMC simulations and ML modelling, MOF database pre-treatment is conducted. We select the hypothetical MOF (hMOF) database (Wilmer et al., 2012) as our basis. First, MOFid and MOFkey identifiers are obtained for each hMOF in the database based on their CIF files (Bucior et al., 2019). Second, the metal nodes, organic linkers, and underlying topological networks are extracted from the identifiers. Finally, data cleaning is performed to remove those MOFs sharing duplicate and incomplete identifiers, with invalid organic linkers, and consisting of more than three types of linkers. This finally results in 9156 unique MOFs.

Figure 1 summarizes the integrated ML framework. As indicated, features of the organic molecule are automatically generated by the graph convolutional network (GCN) approach. Initially, each atom in the molecule is assigned with a fixed-length vector of randomized features. Afterwards, atoms learn their representations by aggregating features from their neighbors (i.e., connected atoms). With two GCN layers, two unconnected atoms can learn features from each other if both of them are attached to the same atom. In this way, with multiple GCN layers, all the atomic features are iteratively updated several times to capture the local and global information about the whole molecular structure. Finally, a global pooling is performed on the graph to generate the overall molecular features by aggregating all the atomic features with a certain mathematical rule. After obtaining the organic linker's features, a feedforward neural network (FNN)-based ML model can be finally constructed to predict MOF adsorption uptakes using all the chemical and structural features as inputs. The chemical features include directly embedded metal node and GCN-based organic linker's features, while the structural features incorporate both embedded topology and five key MOF geometric descriptors. Notably, the initial atomic features, GCN layers, and the FNN model are optimized or trained simultaneously to minimize the overall prediction error of the model. After the ML model is built and successfully validated, it can replace the traditional molecular simulation for providing in-silico predictions on thousands or

millions of new MOFs and quickly find the best candidates possessing the most desirable adsorption properties.

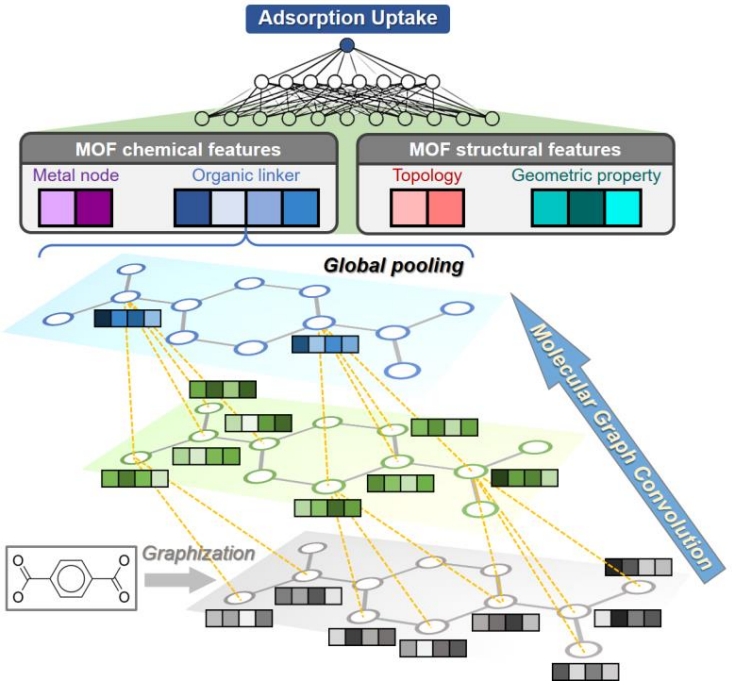


Figure 1. Schematic diagram of the proposed ML framework

3. Case Study

Hydrogen is an appealing energy carrier due to its high gravimetric energy density and low environmental impact. It is industrially stored at around 700 bar, requiring substantial amount of energy and special safety considerations (Gómez-Gualdrón et al., 2016). Recently, MOFs have emerged as promising materials for adsorption-based hydrogen storage at a much moderate pressure around 100 bar. The objective of this work is to use the ML method to find potential MOF candidates that possess high volumetric hydrogen storage capacity. To do so, GCMC simulations are first carried out to compute the hydrogen uptakes at 100 bar/77 K and 2 bar/77 K for the 9156 MOFs. Based on the obtained data, two different ML models are established according to the framework in Figure 1 to predict the two uptakes separately. Using the ML models, the hydrogen storage capacity (maximal amount of hydrogen that can be stored/released for one charging/discharging process), is directly calculated as the difference between the two uptakes.

In order to obtain reliable ML models, the entire dataset is divided into training, validation and test sets accounting for 80%, 10% and 10% of the 9156 data points. These three sets are used for model training, hyper-parameter optimization and early stopping, and model assessment, respectively. Considering all possible hyper-parameter combinations, the optimal ML configuration is first determined by the grid search method using the validation set. After determining the best hyper-parameters, the ML models are trained with the training and assessed with the test data. Model performance

is evaluated with mean absolute error (MAE) and coefficient of determination (R^2), as summarized in Table 1. The parity plot and error distribution of the obtained ML models is visualized in Figure 2. In general, the two ML models achieve accurate predictions for hydrogen uptakes at both 100 bar and 2 bar, with an MAE of 1.04 g/L (1.25 g/L) and 1.03 g/L (1.29 g/L) for the training set (test set), respectively.

Table 1. Model performance in the prediction of hydrogen uptakes

Prediction target	Dataset	MAE (g/L)	R^2
H_2 uptake at 100 bar/77 K	Training	1.04	0.984
	Validation	1.24	0.975
	Test	1.25	0.976
H_2 uptake at 2 bar/77 K	Training	1.03	0.961
	Validation	1.29	0.927
	Test	1.29	0.928

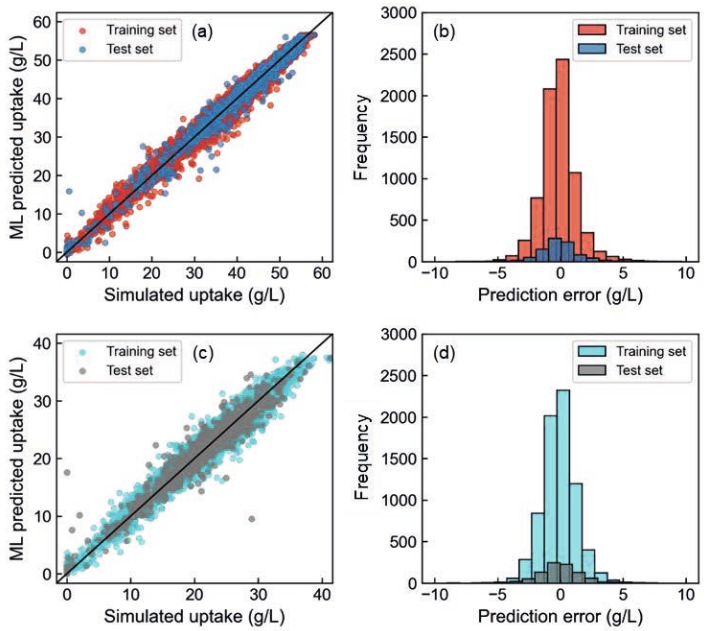


Figure 2. Parity plot and error distribution of the simulated and ML predicted hydrogen uptakes at 100 bar/77 K (a, b) and 2 bar/77 K (c, d)

The ultimate goal is to employ the ML models to discover potential MOFs for efficient hydrogen storage. For this purpose, we collect another much larger MOF database, consisting of 21,384 new MOFs. The two established ML models are then used to predict hydrogen uptakes for these new MOFs, based on which the best candidates possessing the highest storage capacities are screened out. Top 100 candidates are identified, whose storage capacities are between 45.44 g/L and 47.20 g/L. Verification on these top 100 MOFs by GCMC simulations leads to similar capacities. The best two MOFs showing the highest GCMC-derived capacities are illustrated in Figure 3. As indicated, their computed hydrogen storage capacities are higher than the best experimentally verified MOF that shows a capacity of ~ 42 g/L (Ahmed et al., 2017). This proves the great potential of the identified two MOFs for practical applications. As

indicated in Figure 3, these two MOFs have similar structures. For instance, the same topology *pcu* is shared and three types of similar organic linkers are found as well. This provides some useful insights for optimal MOF synthesis.

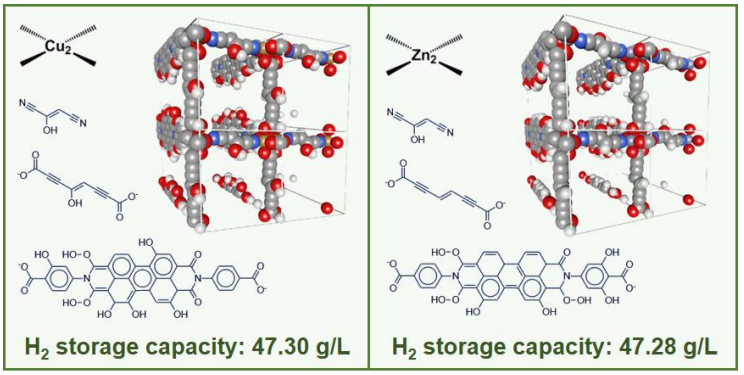


Figure 3. Top two MOF candidates identified for hydrogen storage

4. Conclusions

An integrated ML framework is proposed for the prediction of gas adsorption capacities using both chemical information and structural characteristics of MOFs. The method has been successfully applied to a hydrogen storage case study. High-potential MOFs are successfully identified by large-scale database screening using the obtained ML models. Their superior performances have been validated by GCMC simulations and should be further verified by experiment. It is anticipated that the approach can be used for other applications, e.g., CO_2 capture and methane storage, with a variety of porous materials including zeolites, porous polymers, and covalent-organic frameworks.

Unlike other works using engineered features to train ML models, our approach automatically learns useful features most relevant to adsorption properties. This is efficient, reliable, and can help to discover promising materials. However, it is difficult to draw useful insights on how MOF chemical and structural characteristics influence their performance. This knowledge can be acquired by using the so-called interpretable ML technique, which certainly deserves future studies. Besides, our MOF discovery is achieved by large-scale screening on existing databases. An alternative and probably more efficient way for direct targeting of optimal MOFs is to formulate and solve an optimization-based reverse design problem based on the ML model.

References

A. Ahmed, Y. Liu, J. Purewal, L.D. Tran, A.G. Wong-Foy, M. Veenstra, A.J. Matzger, D.J. Siegel, 2017, Balancing gravimetric and volumetric hydrogen density in MOFs, Energy Environ. Sci., 10, 11, 2459-2471.

C. Altintas, O.F. Altundal, S. Keskin, R. Yildirim, 2021, Machine learning meets with metal organic frameworks for gas storage and separation, J. Chem. Inf. Model., 61, 5, 2131-2146.

B.J. Bucior, A.S. Rosen, M. Haranczyk, Z. Yao, M.E. Ziebel, O.K. Farha, J.T. Hupp, J.I. Siepmann, A. Aspuru-Guzik, R.Q. Snurr, 2019, Identification schemes for metal-organic frameworks to enable rapid search and cheminformatics analysis, Cryst. Growth Des., 19, 11, 6682-6697.

Y. Chen, Z. Qiao, H. Wu, D. Lv, R. Shi, Q. Xia, J. Zhou, Z. Li, 2018, An ethane-trapping MOF PCN-250 for highly selective adsorption of ethane over ethylene, Chem. Eng. Sci., 175, 110-117.

S. Chong, S. Lee, B. Kim, J. Kim, 2020, Applications of machine learning in metal-organic frameworks, Coord. Chem. Rev., 423, 213487.

Y.G. Chung, D.A. Gómez-Gualdrón, P. Li, K.T. Leperi, P. Deria, H. Zhang, N.A. Vermeulen, J.F. Stoddart, F. You, J.T. Hupp, O.K. Farha, R.Q. Snurr, 2016, In silico discovery of metal-organic frameworks for precombustion CO2 capture using a genetic algorithm, Sci. Adv., 2, 10, e1600909.

X. Cui, K. Chen, H. Xing, Q. Yang, R. Krishna, Z. Bao, H. Wu, W. Zhou, X. Dong, Y. Han, B. Li, 2016, Pore chemistry and size control in hybrid porous materials for acetylene capture from ethylene, Science, 353, 6295, 141-144.

M. Fernandez, T.K. Woo, C.E. Wilmer, R.Q. Snurr, 2013, Large-scale quantitative structure-property relationship (QSPR) analysis of methane storage in metal-organic frameworks, J. Phys. Chem. C, 117, 15, 7681-7689.

F. Gándara, H. Furukawa, S. Lee, O.M. Yaghi, 2014, High methane storage capacity in aluminum metal-organic frameworks, J. Am. Chem. Soc., 136, 14, 5271-5274.

D.A. Gómez-Gualdrón, Y.J. Colón, X. Zhang, T.C. Wang, Y.S. Chen, J.T. Hupp, T. Yildirim, O.K. Farha, J. Zhang, R.Q. Snurr, 2016, Evaluating topologically diverse metal-organic frameworks for cryo-adsorbed hydrogen storage, Energy Environ. Sci., 9, 10, 3279-3289.

P.Z. Moghadam, T. Islamoglu, S. Goswami, J. Exley, M. Fantham, C.F. Kaminski, R.Q. Snurr, O.K. Farha, D. Fairen-Jimenez, 2018, Computer-aided discovery of a metal-organic framework with superior oxygen uptake, Nat. Commun., 9, 1, 1-8.

S.M. Moosavi, A. Nandy, K.M. Jablonka, D. Ongari, J.P. Janet, P.G. Boyd, Y. Lee, B. Smit, H.J. Kulik, 2020, Understanding the diversity of the metal-organic framework ecosystem, Nat. Commun., 11, 1, 1-10.

Z. Shi, H. Liang, W. Yang, J. Liu, Z. Liu, Z. Qiao, 2020, Machine learning and in silico discovery of metal-organic frameworks: Methanol as a working fluid in adsorption-driven heat pumps and chillers, Chem. Eng. Sci., 214, 115430.

A.W. Thornton, C.M. Simon, J. Kim, O. Kwon, K.S. Deeg, K. Konstas, S.J. Pas., M.R. Hill, D.A. Winkler, M. Haranczyk, B. Smit, 2017, Materials genome in action: identifying the performance limits of physical hydrogen storage, Chem. Mater., 29, 7, 2844-2854.

C.E. Wilmer, M. Leaf, C.Y. Lee, O.K. Farha, B.G. Hauser, J.T. Hupp, R.Q. Snurr, 2012, Large-scale screening of hypothetical metal-organic frameworks, Nat. Chem., 4, 2, 83-89.

X. Yuan, X. Deng, C. Cai, Z. Shi, H. Liang, S. Li, Z. Qiao, 2021, Machine learning and high-throughput computational screening of hydrophobic metal-organic frameworks for capture of formaldehyde from air, Green Energy Environ., 6, 5, 759-770.

T. Zhou, Z. Song, K. Sundmacher, 2019, Big data creates new opportunities for materials research: A review on methods and applications of machine learning for materials design, Engineering, 5, 6, 1017-1026.

Y. Zhou, T. Zhou, K. Sundmacher, 2020, In silico screening of metal-organic frameworks for acetylene/ethylene separation, In: S. Pierucci, F. Manenti, G.L. Bozzano, D. Manca, eds., Comput.-Aided Chem. Eng., Elsevier, 48, 895-900.

X. Zhang, T. Zhou, K. Sundmacher, 2022, Integrated metal-organic framework and pressure/vacuum swing adsorption process design: Descriptor optimization, AIChE J., e17524.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50302-X

Data-driven Modeling for Magma Density in the Continuous Crystallization Process

Nahyeon An[a, b], Hyukwon Kwon[a] ,Hyungtae Cho[a] and Junghwan Kim[a*]

[a]*Green Materials and Processes R&D Group, Korea Institute of Industrial Technology, Ulsan 44413, Republic of KOREA*
[b]*Department of Chemical and Biomolecular Engineering, Yonsei University, Seoul 03722, Republic of KOREA*
kjh31@kitech.re.kr

Abstract

Crystallization processes have been widely used for separation in many fields, such as food, pharmaceuticals, and chemicals. The crystallization process is a highly nonlinear system, owing to complex crystallization dynamics; therefore, it is difficult to model the process to control the crystal product quality. In this study, a data-driven neural network was implemented to predict the magma density of the continuous crystallization process that produces maleic acid crystals from the mother liquor. Three neural network algorithms, namely deep neural network, long short-term memory, and gated recurrent unit (GRU), were applied for magma density prediction. Process variables, such as the feed flow rate, pressure, and steam flow rate were defined as input, while magma density, the most important control variable in continuous crystallization, was defined as an output variable. The grid search method was used to select suitable hyperparameters for each method, and the predictive accuracy of the models was compared with the root mean square error (RMSE). The GRU-based model afforded the best prediction accuracy among the applied models, with an RMSE of 2.04. Consequently, the developed predictive model can be used as a proper control strategy.

Keywords: Crystallization predictive model; product density prediction; artificial neural network; machine learning and big data

1. Introduction

The crystallization process is used in many fields to produce high-purity products (Velásco-Mejía et al., 2016). Indeed, the quality of the produced crystals has a significant influence on the efficient operation of the downstream process. Therefore, it is crucial to maintain a high and stable quality of the crystallization products. However, the crystallization process consists of several complex mechanisms, such as nucleation, crystal growth, and agglomeration. Moreover, it is difficult to solve the model equation because the mechanisms consist of nonlinear algebraic and partial differential equations. Thus, reliable modeling of the crystallization process remains challenging (Griffin et al., 2016).

To overcome the limitations of equation-based process modeling, the application of artificial neural networks (ANNs) to crystallization process modeling has been studied. Thus, ANNs have been used to model the nonlinear relationship between the input and output variables with high performance. Meng et al. (2021) attempted to monitor the process using a hybrid soft sensor model capable of predicting the mother liquor purity, supersaturation, particle size distribution, and crystal content, which are difficult to

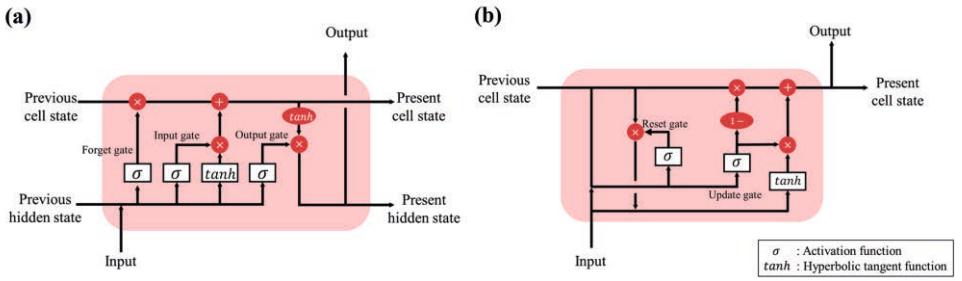


Figure 1 Cell structure of (a) LSTM and (b) GRU

measure in continuous cane sugar crystallization. On the other hand, Manee et al. (2019) developed a model to measure the particle size distribution (PSD) in batch crystallizers through deep learning. Furthermore, in many studies related to the crystallization process, models have been developed to predict the mother liquor purity, particle size, and distribution.

Although the size of each particle is important in the batch crystallization process, it is more crucial to produce products with stable magma density in a continuous process for high productivity. Therefore, in continuous crystallization, the magma density, rather than the PSD, is used as the primary process control variable. However, it is difficult to control the magma density because of the nonlinearity and instability of the process. To solve this problem, this study developed a dynamic prediction model based on an artificial neural network (ANN) to predict the magma density in the continuous crystallization process. Magma density predictive models were developed using three algorithms: deep neural network (DNN), long short-term memory (LSTM), and gated recurrent unit (GRU), wherein LSTM and GRU are recurrent neural network algorithms that reflect the time series of the process. The most suitable model for the continuous crystallization process was selected by comparing the accuracies of the three algorithms for magma density prediction. In addition, the developed data-driven model was applied to control the steam flow rate, and its applicability was verified. Thus, the off-spec of the process was significantly reduced, and the crystals were produced more stably.

2. Preliminaries

2.1. Deep neural network

A DNN is an ANN with multiple layers consists of input, output, and hidden layers. DNNs have been applied to forecast many problems with relatively high performance. Each layer is given the output from the previous layer and transfers it to the next layer. The hidden layers are trained by a backpropagation stochastic gradient descent. The model accuracy highly depends on the algorithms, hyperparameters, the property of the data, and the learning scheme. The outputs ($\mathbf{h}$) of the first, hidden, and output layers are expressed as

$$\mathbf{h}_i = \sigma(\mathbf{W}_i^T \mathbf{x} + \mathbf{b}_i) \tag{1}$$

$$\mathbf{h}_n = \sigma(\mathbf{W}_n^T \mathbf{h}_{n-1} + \mathbf{b}_n) \tag{2}$$

$$\hat{\mathbf{y}} = \mathbf{W}_o^T \mathbf{h}_N + \mathbf{b}_o \quad (3)$$

where $\mathbf{W}$ and $\mathbf{b}$ represent the weight matrix and bias vector of the nth hidden layer, respectively. For the input layer, the input variable vector ($\mathbf{x}$) is used instead of $\mathbf{h}_{n-1}$, while for the output layer, the predicted values of the output layer ($\hat{\mathbf{y}}$) are used instead of $\mathbf{h}_n$.

2.2. Recurrent neural network : LSTM and GRU

An RNN stores the past data and forwards the information to calculate the output of the next step. Unlike the DNN, temporal dynamics can be considered, which is commonly used for time series prediction. Two RNNs, LSTM and GRU, were developed to solve the gradient vanishing problem of standard RNNs. Figure 1 shows the cell structure of the developed LSTM and GRU, comprising three and two gates, respectively, for long-term memory to be efficiently stored. Both algorithms are described in detail in the literature (Hochreiter and Urgen Schmidhuber, 1997).

3. Development of the magma density prediction model

3.1. Process and data description

The target process of this study (Figure 2) is a continuous crystallization process in which an approximately 60% maleic acid–water mixture is concentrated to 78% maleic acid, and 30% of the feed flow rate is crystallized (Ulsan, Republic of Korea). A forced circulation crystallizer was used for crystallization, in which a feed enters the equipment after being heated by 45 ℃ in a heat exchanger. The crystallizer was vacuumed to 50 mbar using a vacuum ejector. The maleic acid crystal was discharged with the mother liquor as magma, where the density of magma was used as a control variable in the process. For stable product production, the magma density was maintained constant at 1,330 kg/m^3 as the set point.

In this study, 10 process variables collected every hour from January to June 2020 were used to develop the data-driven model. Because missing values and outliers in the data

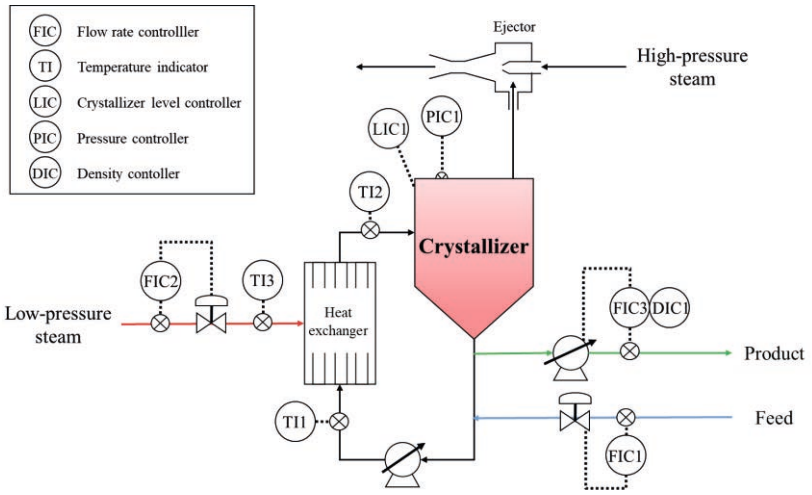


Figure 2 Target process description

Table 1 Parameters of each network

	DNN	LSTM	GRU
Hidden layers	2	1	3
Hidden neurons	20	100	90
Batch size		64	
Early stopping patience		30	
Loss function		Mean square error	

due to instrumental failure adversely affect predictive model training, they were removed and used. The model was developed and evaluated using 80% of 3,300 data as learning data and 20% as test data.

3.2. Model structure

Three algorithms, DNN, LSTM, and GRU, were used to develop a magma density prediction model for the continuous crystallization process. The input variables used in each model were the feed flow rate, low-pressure steam flow rate, pressure, and temperature instruments measured for process monitoring to control the density. The predicted performance was compared using RMSE and calculated using Equation (4). The RMSE, defined as the average absolute ratio error, indicates a higher performance as it approaches 0. The structures of the three predictive models were optimized using grid research, and each model structure is listed in Table 1. In addition, the model with the best performance among the three was derived by analyzing the prediction error distribution.

$$RMSE = \sqrt{\frac{1}{N}\sum_{k}(y_k - \hat{y}_k)^2} \tag{4}$$

where y_k and $\hat{y}_k$ are the actual and predicted data, respectively, and N is the total number of data samples.

4. Results and discussion

4.1. Density prediction accuracy

The magma density prediction performance in the continuous crystallization process using the three models is presented in Table 2. Comparison of the RMSEs reveals that the difference in accuracy between the models is not significant. Figure 3 shows the results predicted by each model, wherein the red line represents the conditions under which the predicted and actual values are equal. Thus, the closer the points are to this line, the more accurate is the model. A point marked in red refers to a data sample with a relatively large error, which is far from the red line. In the three models, the positions marked with red dots were similar, which indicates that a significant error occurred due to an abnormal

Table 2 Prediction accuracy

Model	DNN	LSTM	GRU
RMSE	2.02	2.23	2.04

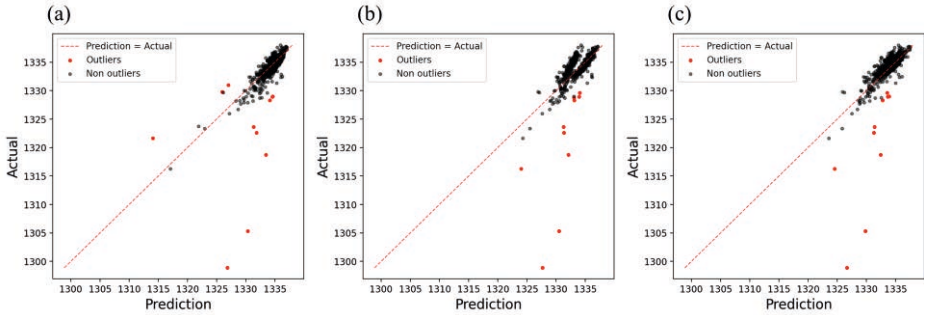


Figure 3 Predicted vs. actual value plots of (a) DNN, (b) LSTM, and (c) GRU models

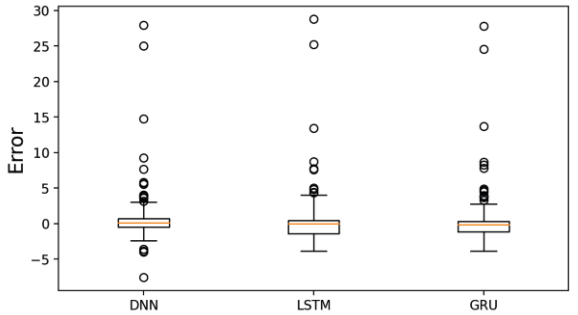


Figure 4 Error distribution boxplot

state of the process. As shown in Figure 3, the DNN and GRU models afforded better predictions than did the LSTM model. Additionally, the box plots in Figure 4 show the distribution of absolute errors in each model, wherein the model accuracy increases as the box at the 0 point becomes narrower. The plot reveals that the absolute error of the GRU is distributed in a narrower area than that of the DNN. Therefore, based on the above analysis, we concluded that the most accurate model for magma density prediction was the GRU model.

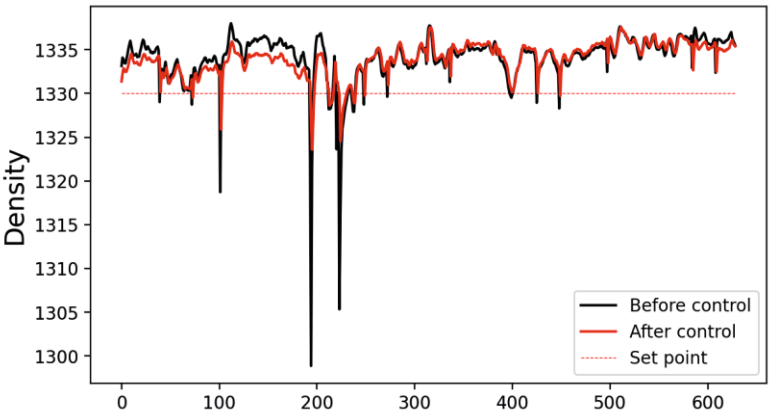


Figure 5 Magma density of before and after steam flow rate control

4.2. Application for steam flow rate control

The density prediction model developed in this study was applied to the steam flow rate control of the target process and compared with the actual steam data. In the current process state, the steam flow rate was changed from -30% to +30%, and the steam flow rate satisfying the set point of the process was derived. In the actual process, an average of 897.2 kg/h of steam was used. However, the prediction model developed in this study revealed that a product that satisfies the setpoint can be produced with an average of only 771.9 kg/h of steam. In addition, as shown in Figure 5, the off-specs generated in the process were also significantly reduced, even though the amount of steam used was decreased by 14%.

5. Conclusion

In this study, we developed an artificial neural network-based dynamic prediction model that can predict the magma density of a continuous crystallization process, which we subsequently applied to control steam usage. We determined that the most suitable model in this process was an RNN-based model using GRU, which presented an RMSE of 2.04. Furthermore, by controlling the steam flow rate using the developed model, 14% of the existing steam usage could be reduced. The dynamic prediction model developed in this study exhibits a high performance, but still has room for improvement. Data used for model development cannot reflect real-time information owing to long sampling rate. Therefore, in future research, if a model is developed by reducing the data sampling interval through data interpolation, it can be applied to an actual process with a more reliable prediction.

Acknowledgement

This study has been conducted with the support of the Korea Institute of Industrial Technology as “Development of digital-based energy optimization platform for manufacturing innovation (KITECH IZ-21-0063)” and “Development of Global Optimization System for Energy Process (KITECH IZ-21-0052, IR-21-0029, EM-21-0022)”

References

Griffin, D.J., Grover, M.A., Kawajiri, Y., Rousseau, R.W., 2016, Data-Driven Modeling and Dynamic Programming Applied to Batch Cooling Crystallization. Industrial and Engineering Chemistry Research. 55(5), 1361–1372.

Hochreiter, S., Urgen Schmidhuber, J., 1997, Long Shortterm Memory. Neural Computation.

Manee, V., Zhu, W., Romagnoli, J.A., 2019, A Deep Learning Image-Based Sensor for Real-Time Crystal Size Distribution Characterization. Industrial and Engineering Chemistry Research. 58(51), 23175–23186.

Meng, Y., Yao, T., Yu, S., Qin, J., Zhang, J., Wu, J., 2021, Data-driven modeling for crystal size distribution parameters in cane sugar crystallization process. Journal of Food Process Engineering. 44(4), 1–15.

Velásco-Mejía, A., Vallejo-Becerra, V., Chávez-Ramírez, A.U., Torres-González, J., Reyes-Vidal, Y., Castañeda-Zaldivar, F., 2016, Modeling and optimization of a pharmaceutical crystallization process by using neural networks and genetic algorithms. Powder Technology. 292, 122–128.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50303-1

Gaussian Process Regression Machine Learning Models for Photonic Sintering

Ke Wang[a], Mortaza Saeidi-Javash[b], Minxiang Zeng[b], Zeyu Liu[b], Yanliang Zhang[b], Tengfei Luo[b], Alexander W. Dowling[a*]

[a]*Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN 46556, USA*
[b]*Department of Aerospace and Mechanical Engineering, University of Notre Dame, Notre Dame, IN 46556, USA*
[*]*adowling@nd.edu*

Abstract

Novel solid-state thermoelectric (TE) materials have the potential to improve energy efficiency by converting waste heat into electricity. However, the performance of many state-of-the-art TE materials remains inadequate for adoption beyond niche applications. Current efforts to optimize photonic sintering, an important step in additive manufacturing of TE devices, rely on expert-driven trial-and-error search which is often extremely time-consuming and without the guarantee of improvement. Emerging Bayesian optimization frameworks offer a principled approach to intelligentially recommend optimized experimental conditions by balancing exploitation and exploration. In this paper, we develop a Gaussian Process Regression (GPR) machine learning model to predict the thermoelectric power factor of aerosol jet printed n-type $Bi_2Te_{2.7}Se_{0.3}$ TE films. We compare hyperparameter tuning methods and perform retrospective analysis to quantify the predictivity of GPR. Finally, we discuss the challenges and opportunities of adopting Bayesian optimization for photonic sintering and fabrication of high-performance TE devices.

Keywords: Additive Manufacturing; Data Science; Bayesian Optimization; Machine Learning; Gaussian Process Regression

1. Introduction

1.1. Background

Discover functional materials with desired properties is a central goal of material science and engineering; yet materials discovery and optimization is often slow and expensive. For example, out of the 10^{23} possible drug-like molecules, only 10^{8} have been synthesized. (Elton et al., 2019) Computer-aided molecular design (CAMD) is frequently used to design new functional material, however, its success is usually limited by the accuracy and efficiency of the physical models. (Austin et al., 2016) Supervised machine learning has demonstrated great promise for predicting the physical properties of material and revolutionizing the design process. (Lookman et al., 2019) For example, Gaussian Process Regression (GPR) and Bayesian optimization have been shown over the past decade to accelerate the design and manufacturing of new functional material. (Wang et al., 2022)

In this paper, we apply GPR to model the thermoelectric power factor of aerosol jet printed n-type $Bi_2Te_{2.7}Se_{0.3}$TE films under different photonic sintering variables. With the urgent demand for waste energy recovery and wearable electronic devices, developing high-performance TE generators has attracted much attention. For example, TE devices can harvest heat from the automobile exhaust to improve its overall efficiency. (Han et al., 2018) They can also convert heat generated by the human body into energy for the electronic device, e.g., wearable medical monitors. (Jiang et al., 2020) The thermoelectric performance of TE materials is typically evaluated by a dimensionless figure of merit, $zT = \frac{S^2\sigma T}{\kappa}$, which depends on the electrical conductivity (σ), Seebeck coefficient (S), thermal conductivity (κ), and absolute temperature (T). Among these, the power factor, ($S^2\sigma$), is usually of the utmost importance in developing new TE materials platforms.

2. Method

2.1. Decision Variables and Data

Photonic sintering is a well-known technique for fabricating high-performance thin-film TE material. (Yu et al., 2017) For photonic sintering, there are four key process variables to optimize: voltage (x_{i1}), pulse duration (x_{i2}), number of pulses (x_{i3}), and pulse delay (x_{i4}). For each experiment i, photonic sintering is performed at conditions $\boldsymbol{x}_i$ and the power factor (y_i) is measured. Saeidi-Javash et al. (2019) described the experimental procedure used to collect the thermoelectric properties under various photonic sintering conditions in this paper. Figure 1 summarizes the photonic sintering dataset. In Group 0, seven experiments were performed with one pulse, and voltage and pulse duration systematically varied based on expert intuition. The remaining groups, 1 – 4, are one dimension sensitivity analysis in which voltage (x_{i1}), pulse duration (x_{i2}), and pulse delay (x_{i4}) are fixed in each group and the number of pulses (x_{i3}) is systematically varied. The GPR machine learning models are then trained using these data in this work.

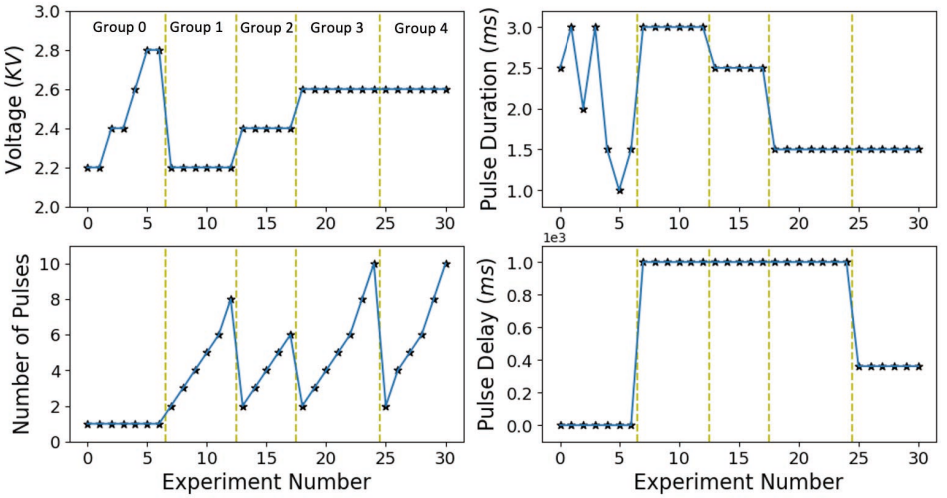


Figure 1: The four photonic sintering optimization variables were systematically varied, based on expert intuition, across 31 experiments which are divided into five groups.

Let $\mathrm{D} = \{(\boldsymbol{x}_i, y_i), |\boldsymbol{x}_i \in \mathrm{R}^4, y_i \in \mathrm{R}, i \in 1,\ldots,31\}$ be a collection of 31 photonic sintering experiments (Figure 1). For convenience, we denote the data $D = (\boldsymbol{X}, \boldsymbol{y})$ using matrix $\boldsymbol{X} = (\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)^T$ and vector $\mathbf{y} = (y_1,\ldots,y_n)^{\mathrm{T}}$. However, each dimension of $\boldsymbol{x}_i$, as well as y_i, has different units. To address this, each dimension of D is standardized using the mean (expected value) and standard deviation where $\boldsymbol{x}_j$ is the jth column of $\boldsymbol{X}$:

$$\boldsymbol{y} \leftarrow \frac{\big(\boldsymbol{y} - E(\boldsymbol{y})\big)}{\sqrt{Var(\boldsymbol{y})}}, \qquad \boldsymbol{x}_j \leftarrow \frac{\Big(\boldsymbol{x}_j - E\big(\boldsymbol{x}_j\big)\Big)}{\sqrt{Var\big(\boldsymbol{x}_j\big)}} \tag{1}$$

2.2. Gaussian Process Regression

Gaussian Processes (GPs) are non-parametric probabilistic models that are well-known to emulate expensive continuous functions, $f(.)$, by interpolating between training data.

$$f \sim GP\big(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')\big) \qquad \boldsymbol{x}, \boldsymbol{x}' \in R^p \tag{2}$$

The output of a GP model is a normally distributed random variable fully specified by the mean function, $m(\boldsymbol{x}) = E[f(\boldsymbol{x})]$, and the kernel function $k(\boldsymbol{x}, \boldsymbol{x}') = E\big[\big(f(\boldsymbol{x}) - m(\boldsymbol{x})\big)\big(f(\boldsymbol{x}') - m(\boldsymbol{x}')\big)\big]$. (Rasmussen, 2003) The kernel function determines how the GP model interpolates between the encoded data D. In doing so, the kernel function also specified the uncertainty in GP predictions. $k(.,.)$ contains hyperparameters which often include the length-scales, denoted by $\boldsymbol{l} \in R^p$, for each dimension of the input data $\boldsymbol{X}$. The smaller the l_j, the more important that corresponding feature (x_j). Training the hyperparameters is often performed by using log marginal likelihood or cross-validation methods. The optimal length-scales $\boldsymbol{l}$ help identify which features are most important. For simplicity, we set $m(\boldsymbol{x})$ to zero and use Radial Basis Function (k_{RBF}) defined in Eq. (3), where $\boldsymbol{l} = (l_1, l_2, l_3, l_4)^T$.

$$k_{RBF}(x, x') = e^{-\frac{1}{2}\sum_{j=1}^{p}\left(\frac{x_j - x'_j}{l_j}\right)^2} \qquad \theta = \boldsymbol{l} \tag{3}$$

We define new inputs values $\boldsymbol{X}_*$ with corresponding prediction $\boldsymbol{f}_*$. Given training data $(\boldsymbol{X}, \boldsymbol{y})$ and values of the hyperparameters θ, we can write the outputs $\boldsymbol{y}$ and $\boldsymbol{f}_*$ as a multivariate normal (Gaussian) distribution, Eq. (4), where $\boldsymbol{K}(.,.)$ kernel function $k(.,.)$ is evaluated elementwise. Moreover, we assume each measurement is corrupted by normally distributed observation error ε with zero mean and variance σ^2, $\varepsilon \sim N(0, \sigma^2)$.

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{f}_* \end{bmatrix} \sim \boldsymbol{N}\left(\begin{bmatrix} \boldsymbol{m}(\boldsymbol{X}) \\ \boldsymbol{m}(\boldsymbol{X}_*) \end{bmatrix}, \quad \begin{bmatrix} \boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}) + \boldsymbol{\sigma^2 I} & \boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}_*) \\ \boldsymbol{K}(\boldsymbol{X}_*, \boldsymbol{X}) & \boldsymbol{K}(\boldsymbol{X}_*, \boldsymbol{X}_*) \end{bmatrix}\right) \tag{4}$$

The conjugacy properties of multivariate Gaussian distribution give (Bishop, 2006):

$$E(\boldsymbol{f}_*) = \boldsymbol{m}(\boldsymbol{X}_*) + \boldsymbol{K}(\boldsymbol{X}_*, \boldsymbol{X})[\boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}) + \sigma^2 \boldsymbol{I}]^{-1}\big(\boldsymbol{y} - \boldsymbol{m}(\boldsymbol{X})\big) \tag{5a}$$

$$Var(\boldsymbol{f}_*) = \boldsymbol{K}(\boldsymbol{X}_*, \boldsymbol{X}_*) - \boldsymbol{K}(\boldsymbol{X}_*, \boldsymbol{X})[\boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}) + \sigma^2 \boldsymbol{I}]^{-1}\boldsymbol{K}(\boldsymbol{X}, \boldsymbol{X}_*) \tag{5b}$$

2.3. Hyperparameter Tuning

A key step in GP modeling is training the hyperparameters. We start by comparing the performance of log marginal likelihood (LML) Eq.(6) and cross-validation (CV) Eq.(7) for training the length scales of each dimension (l_1, l_2, l_3, l_4) and the optional observation

error σ as hyperparameters. LML (a.k.a. maximum likelihood estimation, MLE) uses all the training data D to find the hyperparameter values which maximize the log-likelihood function:

$$\log p(\boldsymbol{y}|\boldsymbol{X},\theta) = -\frac{1}{2}\boldsymbol{y}^T[\boldsymbol{K}(\boldsymbol{X},\boldsymbol{X}|\theta) + \sigma^2\boldsymbol{I}]^{-1} - \frac{1}{2}\log|\boldsymbol{K}(\boldsymbol{X},\boldsymbol{X}|\theta) + \sigma^2\boldsymbol{I}| - \frac{n}{2}\log 2\pi \quad (6)$$

In contrast, CV uses only a subset of the data to reduce the variance of the prediction evaluation. The LML is computed with data $D_{-i} = (\boldsymbol{X}_{-i}, \boldsymbol{y}_{-i})$ where $-i$ denotes all data except sample i:

$$\log P(y_i|X_{-i}, y_{-i}, \theta) = -\frac{1}{2}\log\sigma_i^2 - \frac{(y_i - \mu_i)^2}{2\sigma_i^2} - \frac{1}{2}\log 2\pi \quad (7a)$$

The conjugacy property of GPR greatly reduces the computation cost of evaluating Eq. (7a). The overall leave-one-out CV (Loo-CV) likelihood function is computed by averaging all the leave-one-out samples:

$$L_{Loo-CV}(X, y, \theta) = \frac{1}{n}\sum_{i=1}^{n}\log P(y_i|X_{-i}, y_{-i}, \theta) \quad (7b)$$

The domain knowledge of experimentalists believed that the four proposed variables are all influential for determining power factor (y_i). To incorporate this prior knowledge, we bounded hyperparameter values, including $\boldsymbol{l}$ and σ, between 0 and 1 for this preliminary analysis, since a large value for l_j would imply that dimension j is not important.

3. Results

3.1. Log Marginal Likelihood (LML) and Leave-one-out Cross-Validation (Loo-CV) identify similar hyperparameter values

We start by comparing LML and Loo-CV hyperparameter training approaches for the photonic sintering data. Table 1 shows LML and Loo-CV identify identical optimal hyperparameters using grid search. The first two rows correspond to optimizing $\boldsymbol{l}$ with fixed $\sigma = 0.1$ which is informed by the experimental observation error. The optimal $\boldsymbol{l}$ obtaining with LML and Loo-CV methods are the same which suggests the simpler method, LML, is adequate for this photonic sintering dataset. Conversely, the third and fourth rows consider both $\boldsymbol{l}$ and σ as optimized hyperparameters. With σ considered as a tuneable hyperparameter, σ increases from 0.1 to 0.2, and l_2 increases from 0.635 to 0.687. These changes reflect the trade-off between bias and variance (Bishop, 2006) and correspond to the conclusion that relatively more complicated model (e.g., $l_2 = 0.635$) usually obtaining low observation error (e.g., $\sigma = 0.1$), while a simpler model (e.g., $l_2 = 0.687$) has higher observation error (e.g., $\sigma = 0.2$).

Table 1: Comparison of hyperparameter values from LML and Loo-CV training

	l_1	l_2	l_3	l_4	σ
LML with σ fixed	1	0.635	0.322	1	0.1
Loo-CV with σ fixed	1	0.635	0.322	1	0.1
LML with σ tuned	1	0.687	0.322	1	0.2
Loo-CV with σ tuned	1	0.687	0.322	1	0.2

Next, we consider the predictive uncertainty of the GP model. Figure 2 is a parity plot for LML optimal hyperparameters $(l_1 = 1, l_2 = 0.687, l_3 = 0.322, l_4 = 1, \sigma = 0.2)$. This plot shows the leave-one-out predictions with the GP model. The x-axis and y-axis are experimental and predicted power factor, respectively. The five symbols demark groups

of experiments. The error bars show one prediction standard deviation from the GP model. We observed that 27 out of 31 predictions are within one prediction standard deviation of the parity line, suggesting the GPR model successfully emulates the relation between experimental conditions $\boldsymbol{x}_i$ and target y_i.

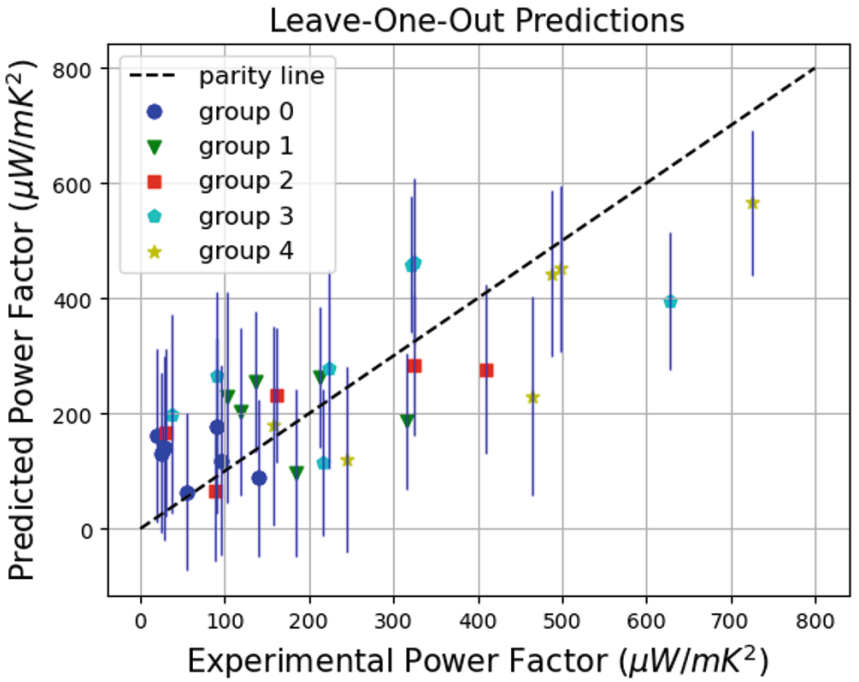


Figure 2: Parity plot of GPR prediction and experimental power factor R in photonic sintering.

3.2. Retrospective Analysis

Figure 3 illustrates the retrospective analysis of GPR prediction in the photonic sintering dataset. The predicted power factors (squares) are generated iteratively (with hyperparameters $l_1 = 1, l_2 = 0.687, l_3 = 0.322, l_4 = 1, \sigma = 0.2$ fixed) using all previous data (to the left of each square). For example, the GPR prediction for experiment 5 uses 5 prior observations for training. The red diamonds show the experimentally measured power factor, and the dashed lines demark each experimental group. In this analysis, we observed 25 out of 30 samples fallen into the predicted (within one standard deviation) bounds. This result demonstrates the GPR model predictions improve as additional data are incorporated into the model.

4. Conclusion

In this work, we successfully develop a GP model to predict the power factor of sintered n-type $Bi_2Te_{2.7}Se_{0.3}$TE films as a function of four photonic sintering variables. This analysis shows that LML and Loo-CV hyperparameter tuning methods identify the same optimal hyperparameters. Through both the parity plot (Figure 2) and retrospective analysis (Figure 3), we show the accuracy of the GPR predictions (both mean and uncertainty estimates). These results suggest that the GP models can be integrated into a Bayesian optimization framework to identify photonic sintering

experimental conditions to maximize the thermoelectric power factor. We thank the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) and Advanced Manufacturing Office Award DE-EE0009103 for funding.

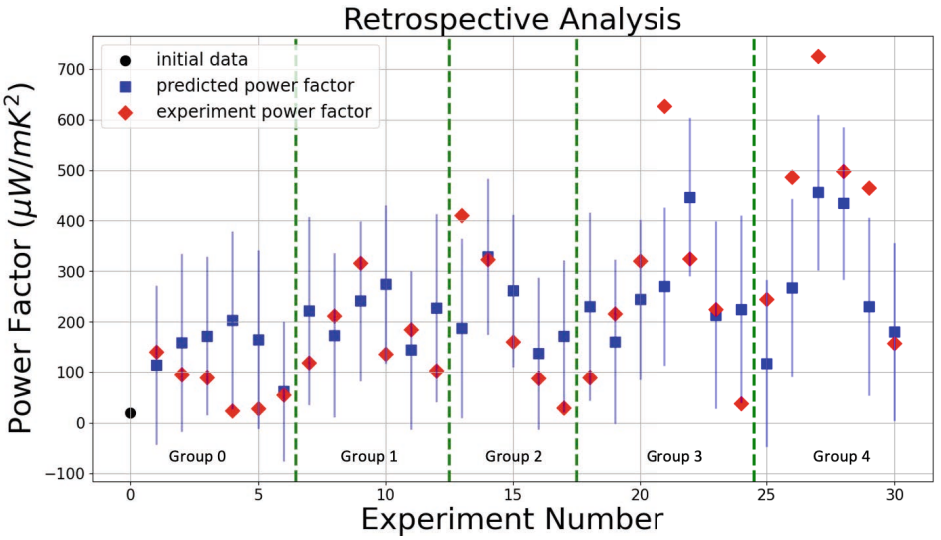


Figure 3: Retrospective analysis of GPR in photonic sintering

References

Elton DC, Boukouvalas Z, Fuge MD, Chung PW. Deep learning for molecular design—a review of the state of the art. Molecular Systems Design & Engineering. 2019;4(4):828-49.

Austin ND, Sahinidis NV, Trahan DW. Computer-aided molecular design: An introduction and review of tools, applications, and solution techniques. Chemical Engineering Research and Design. 2016 Dec 1;116:2-6.

Jiang C, Ding Y, Cai K, Tong L, Lu Y, Zhao W, Wei P. Ultrahigh performance of n-type Ag2Se films for flexible thermoelectric power generators. ACS applied materials & interfaces. 2020 Feb 3;12(8):9646-55.

Saeidi-Javash M, Kuang W, Dun C, Zhang Y. 3D Conformal Printing and Photonic Sintering of High-Performance Flexible Thermoelectric Films Using 2D Nanoplates. Advanced Functional Materials. 2019 Aug;29(35):1901930.

Bishop CM. Pattern recognition. Machine learning. 2006 Feb;128(9).

Rasmussen CE. Gaussian processes in machine learning. InSummer school on machine learning 2003 Feb 2 (pp. 63-71). Springer, Berlin, Heidelberg.

Yu M, Grasso S, Mckinnon R, Saunders T, Reece MJ. Review of flash sintering: materials, mechanisms and modelling. Advances in Applied Ceramics. 2017 Jan 2;116(1):24-60.

Lookman T, Balachandran PV, Xue D, Yuan R. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. npj Computational Materials. 2019 Feb 18;5(1):1-7.

Ke Wang, Alexander W Dowling, Bayesian optimization for chemical products and functional materials, Current Opinion in Chemical Engineering, Volume 36, 2022, 100728

Han C, Tan G, Varghese T, Kanatzidis MG, Zhang Y. High-performance PbTe thermoelectric films by scalable and low-cost printing. ACS Energy Letters. 2018 Feb 23;3(4):818-22.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50304-3

Development of Dye Exhaustion Behavior Prediction Model using Deep Neural Network

Jonghun Lim[a,b§], Soohwan Jeong[a,c§], Sungsu Lim[c], Hyungtae Cho[a], Jae Yun Shim[d], Seok Il Hong[d], Soon Chul Kwon[d], Heedong Lee[d], Il Moon[C] , Junghwan Kim[a*]

[a]*Green Materials and Process R&D Group, Korea Institute of Industrial Technology, 55, Jonga-ro, Jung-gu, Ulsan 44413, Republic of Korea*
[b]*Department of Chemical and Biomolecular Engineering, Yonsei University, 50, Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea*
[c]*Department of Computer Science Engineering, Chungnam National University, 99, Daehak-ro, Yuseong-gu, Daejeon 34134, Republic of Korea*
[d]*ICT Textile and Apparel R&BD Group, Korea Institute of Industrial Technology, Ansan, 15588, Republic of Korea*

[§]*Lim J. and Jeong S. contributed equally to this work as first authors.*
kjh31@kitech.re.kr

Abstract

The textile dyeing process consumes a significant quantity of energy as it is necessary to maintain the water temperature between 60–120 °C during the dyeing of reactive dyes. Therefore, to reduce the overall cost of the process through reducing the quantity of waste energy, it is crucial to increase the right first time (RFT) rate, which corresponds to the rate at which the target color is imparted through a single dyeing process (Park et al., 2009). To improve the RFT rate, the proper operation with following the optimal dye exhaustion behavior in consideration of the color difference and dyeing uniformity is a critical factor. The color difference is determined according to maximal absorption and the dyeing uniformity is decided by dye exhaustion behavior [Bouatay et al., 2016]. In this study, we developed a model for predicting dyeing exhaustion behavior, and utilized the model to predict optimal dye exhaustion behavior under various dyeing conditions. A deep neural network-based on the dye exhaustion behavior prediction model was developed through regression analysis, the model was further developed and evaluated by dividing the entire dataset into learning and evaluation data. The model's performance was evaluated using the root mean square error (RMSE) parameter alongside the coefficient of determination (R^2) which acted as performance evaluation metrics. Using these performance metrics, it was found that the proposed DNN regression exhibited the highest performance and the smallest error in comparison with established models, with root mean square error RMSE and R^2 values of 0.016 and 0.994, respectively. The results reported in this study demonstrate that the proposed model exhibits superior performance in predicting the dye exhaustion behavior.

Keywords: Textile industry, re-dyeing, right-first-time, deep neural network-based prediction

1. Introduction

Within the textile industry, dyeing forms a cost-intensive process requiring considerable volumes of hot water and the chemicals required for reactive dyeing. In addition, this process generates significant volumes of wastewater, which forms a severe environmental pollutant. The exhaust method is frequently used within the textile industry to dye cellulose fibers with reactive dyes as it achieves the highest productivity over a short period of time. However, the exhaust method exhibits a disadvantage in that the quality of the final product exhibits significant variations as a result of slight fluctuations in dyeing conditions, such as the dye ratio, temperature, and the Na_2SO_4 and Na_2CO_3 concentrations. Therefore, to obtain the product's target quality, a re-dyeing procedure is often required, which increases the overall dyeing cost by 98–169 % in comparison to a one-time dyeing process, while environmental pollution is also increased owing to the increased wastewater volume. Thus, to address this problem, it is crucial to increase the right first time (RFT) rate, which is the rate at which the target color is reproduced with only one dyeing cycle in the dyeing machine.To increase the RFT rate during the textile dyeing process, active research has recently been conducted to increase the compatibility of the dye and also to determine the optimal dyeing conditions to increase the final exhaustion rate, which impacts the target color change. To increase the dye compatibility, Kim et al. determined the optimal pH conditions for mononicotinic acid triazine-type dyes. It was found that the highest final exhaustion rate was achieved at pH values greater than 9. Kim et al. proposed the optimal dyeing conditions for the application of both reactive and acidic dyes to Angora fibers to improve the final exhaustion rate. The optimal dyeing conditions were reported to be a dye concentration of 8 % o.w.f. for both the reactive and acidic dyes, a pH of 3–4 at 110 °C for reactive dyes, and a pH of 3–4 at 70 °C for acidic dyes.

Despite the considerable number of reports aiming to increase the RFT rate, a significant limitation still remains. During the textile dyeing process, the dyeing quality was determined based on the color difference relative to the target color and dyeing uniformity. If the dye is adsorbed into the fabric at a greater rate than the optimal exhaustion behavior, the dye alkalizes rapidly during the reaction stage. This results in uneven fixation and dyeing, which degrades the dyeing uniformity. Thus, in order to increase the RFT rate, it is crucial to consider dyeing uniformity, which simultaneously determines both the exhaustion behavior and color difference. At present, no relevant studies have been conducted in this respect.

To address these limitations, we propose a dye exhaustion behavior prediction model, utilizing a deep neural network (DNN), to determine the optimal dyeing conditions. The aim of this particular study was to increase the RFT rate by deriving the optimal dyeing conditions, which are obtained through a determination of the optimal exhaustion behavior which considers both the color difference and dyeing uniformity to overcome the significant limitations imparted by excessive costs and environmental contamination.

2. Process description

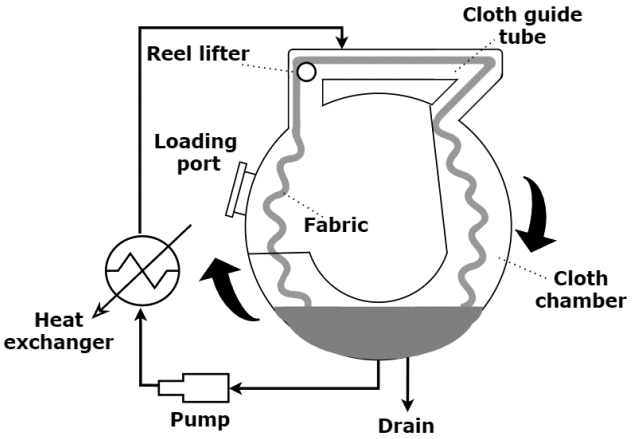


Figure 1. Schematic diagram of a round-configuration jet dyeing machine

Figure 1 shows a schematic diagram of a round-configuration jet-dyeing machine. In this particular apparatus, dyeing proceeds under high temperature and pressure to facilitate the reactive dyeing procedure. In general, the reactive dyeing process proceeds as follows. Initially, the dye is ejected through a nozzle attached to the cloth guide tube within the closed tube system of the cloth chamber. During the continuous dye ejection through the nozzle, the dye is absorbed into the fabric, which is rotated over the reel lifter through the application of the injection pressure. The resultant turbulence facilitates dye penetration into the fabric, while simultaneously reducing the mechanical impact on the fabric. Finally, the dyeing solution containing the reactive dye is heated using a heat exchanger to obtain a suitable temperature for reactive dyeing.

3. Methodology

3.1. Data generation and preprocessing

In order to develop the dye exhaustion behavior prediction model with respect to the dyeing procedure, 615 datasets were extracted detailing time, temperature, and Na_2SO_4 and Na_2CO_3 input quantities. The time values ranged between 0 and 120 min at 3 min intervals, and the temperature fluctuated according to certain set values over this time period. Three Na_2SO_4 input quantities were evaluated, specifically, 10, 30, and 50 g; while five Na_2CO_3 input quantities were evaluated, specifically, 0, 5, 10, 15, and 20 g. Following this, the extracted datasets were preprocessed in two steps. First, any datasets that were not required for the dye exhaustion behavior prediction model were removed. As a result, since no trend was observed at an Na_2CO_3 input quantity of 0 g, the corresponding dataset was eliminated. Secondly, each dataset was normalized as the units ascribed to each data point vary, which, in turn, hinders the learning process. In this study, the z-score normalization method was applied, while the mean and standard deviation for each value of the same parameter were used to scale the data to exhibit a normal distribution with a mean of 0 and a standard deviation of 1. In addition, any time warping or alignment of the data sets were ignored.

3.2. Development of DNN-based prediction model

A DNN-based prediction model was developed to predict the exhaustion behavior exhibited during textile dyeing procedures. Typically, a DNN is a machine learning algorithm based on an artificial neural network (ANN) which mimics the principles and structure of a human neural network. An ANN is composed of an input layer, a hidden layer, and an outer layer. If the number of hidden layers is greater than or equal to three, the system is denoted as a DNN. Figure 2 shows a typical DNN structure. During regression analysis or classification problems, a linear estimation function of the type $y = w^T x + b$ is used to solve the linear problem. Typically, a DNN applies an activation function to a linear estimation function to solve nonlinear problems.

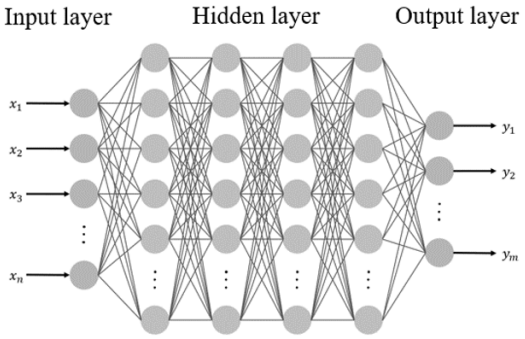


Figure 2. Typical DNN structure

In order to develop the DNN-based dye exhaustion behavior prediction model, the datasets obtained after preprocessing were used. These were the exhaustion rates according to time, temperature, and the Na_2SO_4 and Na_2CO_3 input quantities. The pre-processed datasets were divided into training (75 %) and test (25 %) sets. With respect to the DNN hyper-parameter, the model consisted of three hidden layers (h_1, h_2, and h_3), with the unit corresponding to each hidden layer set to 100. ReLU was used as the activation function for h_1 and h_2, while a sigmoid was used for h_3.

$$\mathrm{ReLU}(x) = \max(0, x) \tag{1}$$

$$\mathrm{sigmoid}(x) = (1 + e^{-x})^{-1} \tag{2}$$

Finally, Adam was applied as the optimizer function, while the mean squared error (MSE) was determined to form an appropriate loss function for this regression analysis.

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} \tag{3}$$

4. Results and discussion

4.1. Performance of the DNN-based prediction model

This section discusses the performance of the exhaustion behavior prediction model developed using a DNN-based regression analysis, in which the MSE and R^2 were used as performance indicators. To evaluate the performance of the DNN model, other

regression analysis methods, such as lasso, ridge, and support vector regression, were applied to identical datasets. Each model was subjected to 100 regression experiments, and the training and test data were varied randomly each time. Table 1 lists the performance of each model.

Table 1. Performance of each regression model

Regression model	R^2	$RMSE$
Lasso regression	0.726 ± 0.040	0.110 ± 0.015
Ridge regression	0.723 ± 0.037	0.273 ± 0.020
Support vector regression	0.902 ± 0.035	0.066 ± 0.017
DNN regression	0.994 ± 0.004	0.016 ± 0.006

From the data shown in table 1, it can be seen that the DNN regression exhibited the highest performance and the smallest error, with root mean square error (RMSE) and R^2 values of 0.016 and 0.994, respectively.

4.2. Prediction results of the dye exhaustion behavior

Based on the proposed DNN-based prediction model, the dye exhaustion behavior as a function of time, temperature, and the Na_2SO_4 and Na_2CO_3 input quantities was predicted. Figures 3 correspond to the cases for which the Na_2CO_3 input quantities were 5 and 10 g, and 15 and 20 g, respectively.

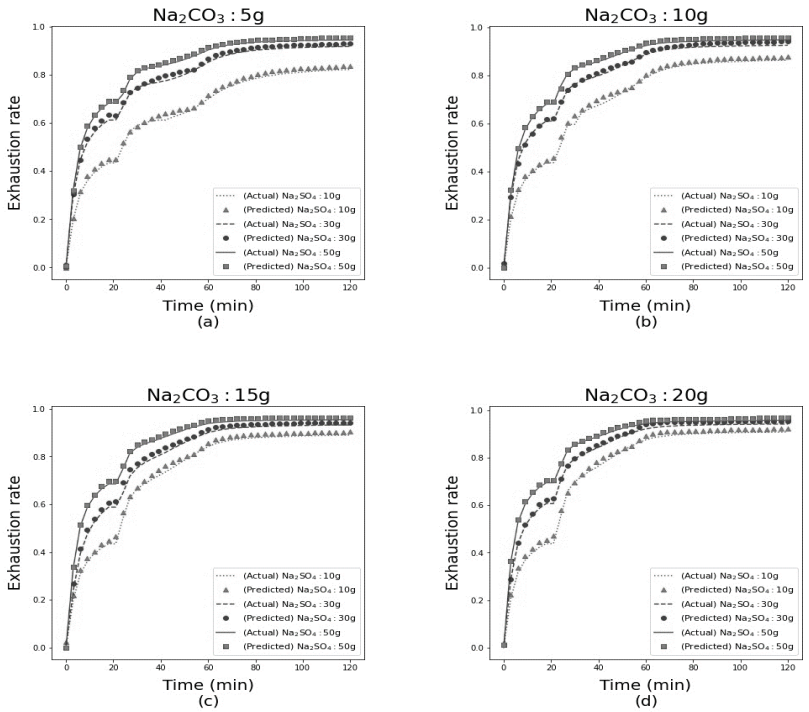


Figure 3. Dye exhaustion behavior prediction

It is clear that the data obtained from the DNN-based prediction model strongly agrees the exhaustion behavior exhibited over time. Hence, the prediction model which is defined using the true dataset can be considered reliable. Through the application of the proposed DNN-based prediction model, the optimal dye exhaustion behavior can be predicted with high accuracy under various dyeing conditions. Therefore, it is possible to derive the optimal dyeing conditions that are derived from the targeted optimal dye exhaustion behavior without the need to perform the dyeing process several times.

5. Conclusions

In this study, we developed a dye exhaustion behavior prediction model which utilized a DNN to determine the optimal dyeing conditions. As the proposed DNN-based prediction model was used to predict the dye exhaustion rate under various dyeing conditions, it was then possible to derive the optimal dyeing conditions through an evaluation of the optimal dye exhaustion behavior. Thus, it will be possible to increase the RFT rate by considering both the color difference and dyeing uniformity to overcome the significant hinderances of excessive cost and environmental contamination through the generation and release of wastewater. As a result, the application of the proposed DNN-based prediction model reduces the re-dyeing rate through increasing the RFT rate. Therefore, the proposed model facilitates a significant improvement in the environmental and economic impact imparted during the dyeing process, while also providing valuable insight into the textile dyeing process.

Acknowledgements

This study has been conducted with the support of the Korea Institute of Industrial Technology with respect to "Development of complex parameter smart analysis modules for color customering (kitech EH-21-0008)" and "Development of Global Optimization System for Energy Process (kitech EM-21-0022), (kitech IZ-21-0052), (kitech IR-21-0029)".

References

J. Park, and J. Shore, 2009, Evolution of right-first-time dyeing production, Coloration technology, 125, 133-140

F. Bouatay, N. Meksi, S. Adeel, F. Salah, and F. Mhenni, 2016, Dyeing behavior of the cellulosic and jute fibers with cationic dyes: process development and optimization using statistical analysis, Journal of Natural Fibers, 13, 423-436

J. Kim, K. Lee, H. Park, N. Yoon, 2004, The One-bath One-step Dyeing of Nylon / Cotton Blends with Acid Dyes and Mononicotinic acid-triazine type Reactive Dyes, Textile Coloration and Finishing, 16, 1-7

Y. Kim, S. Lee, Y. Son, 2010, Dyeing Properties of Acid and Reactive Dye for Super Soft Angora / PET, Nylon Blended Fabric. Textile Coloration and Finishing, 22, 332-40

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50305-5

Guaranteed Error-bounded Surrogate Modeling and Application to Thermodynamics

Ashfaq Iftakher, Chinmay M. Aras, Mohammed Sadaf Monjur, M. M. Faruque Hasan*

Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX 77843-3122, USA.
hasan@tamu.edu

Abstract

We present a data-driven surrogate modeling technique and demonstrate its applicability to replace complicated thermodynamic models for efficient process simulation and synthesis. We employ data-driven edge-concave underestimators and edge-convex overestimators to provide guaranteed error-bounded approximation over the entire domain. A surrogate model is then achieved by performing a parameter estimation that ensures the approximation to be bounded between the vertex polyhedral under- and over-estimators of the original model. We also present GEMS (Guaranteed Error-bounded Modeling of Surrogates) framework, which is a package with automated dataflow for sample evaluation, Hessian bound estimation and parameter estimation to obtain the surrogate models. We apply the technique to predict the solubility of hydrofluorocarbon (HFC) refrigerants in ionic liquids.

Keywords: Surrogate Modeling, Simulation-based Optimization, Data-driven Modeling

1. Introduction

Thermodynamic models are key to design realistic chemical processes to address climate change, decarbonization and other grand challenges. For example, ionic liquid (IL)-assisted innovative separation processes using extractive distillation require rigorous solubility modeling. Also, due to the nature of ILs as 'designer solvents', selection of optimal IL from many candidate solvents for the same separation task requires the understanding of phase behavior. In general, computer aided process intensification guides the discovery of innovative process units that may result in dramatic performance improvement. To attain confidence in the proposed design, the underlying mathematical model must be able to sufficiently capture the physical phenomena. To achieve this, rigorous thermodynamic models are incorporated while solving process synthesis problems. However, attaining globally optimal solutions has been a challenge due to the nonlinear and nonconvex nature or large size of thermodynamic models.

In simulation based optimization approach, a thermodynamic model is treated as black-box. The sampling data is generated over the operating domain and the output data is utilized by an optimizer which performs derivative-free optimization (Bajaj et al., 2021). To obtain guaranteed convergence to optimal solution, one requires dense sampling. If the black-box model is very large, the samplings can become computationally expensive. To increase efficiency, a set of sampling data can be used

to generate surrogate (reduced order) models (Cozad et al., 2014; Boukouvala et al., 2017). Such simpler models can achieve computational efficiency by replacing the original computationally expensive models. However, the prediction accuracy remains a challenge. For example, single/piecewise linear approximations, polynomial response surfaces, Artificial neural networks (ANN) suffer from not being able to reliably predict the type of approximation (under vs overestimation) (Jones, 2001). Also, Kriging (Jones et al., 1998), radial basis functions (RBF) and other interpolating surrogates (Bhosekar and Ierapetritou, 2018) exactly predict the training points (Wang et al., 2014) while providing no guarantee on the quality of prediction over the entire domain of interest.

It has been recently shown that theoretically guaranteed lower bounds can be tractably obtained just by data-driven black-box sampling (Bajaj and Hasan, 2019). The only information required is the global upper bound on the diagonal Hessian elements which can be obtained by either physical intuition, solving an NLP procedure or automatic differentiation over the whole discretized domain. Specifically for the application of thermodynamics, since it is possible to derive the bounds on the diagonal Hessian elements for known models, the edge concavity-based relaxation provides an attractive way towards developing surrogate thermodynamic models with theoretically guaranteed bounds on the prediction errors. In this work, we extend the underestimator formulation and propose a new data-driven surrogate modeling technique that provide theoretically guaranteed tight error bounding (under and overestimation) of blackbox models over the entire domain. We also present a framework that performs sampling of the blackbox models, calculates Hessian via automatic differentiation, and performs globally bounded parameter estimation through GAMS thereby facilitating the data transfer and allowing a single flexible and user friendly package.

2. THEORETICAL BOUNDED APPROXIMATION OF BLACKBOX MODELS

2.1. Edge-concave underestimation and edge-convex overestimation based bounding

We adopt the edge-concave underestimator (Hasan, 2018; Bajaj and Hasan, 2019) and utilize vertex polyhedral property (Tardella, 2004) to construct linear facets of the convex envelope solely based on evaluation of a given greybox/blackbox function $f(\mathbf{x})$ at the domain bounds and interior (sampled) points. Assuming twice-differentiability of $f(\mathbf{x})$, its edge-concave underestimator, $L(\mathbf{x})$ is given by:

$$L(\mathbf{x}) = f(\mathbf{x}) - \sum_{i=1}^{n} \theta_i^L (x_i - x_i^{Int})^2 \tag{1}$$

where, x_i^{Int} is the value of sampled variable x_i, and the parameter θ_i^L is defined as:

$$\theta_i^L = \max\left\{0, \frac{1}{2}\left[\frac{\partial^2 f}{\partial x_i^2}\right]^U\right\} \tag{2}$$

Similarly, edge-convex overestimator, $U(\mathbf{x})$ is expressed as follows:

$$U(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^{n} \theta_i^U (x_i - x_i^{Int})^2 \tag{3}$$

where, the parameter θ_i^U is defined as:

$$\theta_i^U = \max\left\{0, \frac{1}{2}\left[\frac{-\partial^2 f}{\partial x_i^2}\right]^U\right\} \tag{4}$$

For an n-dimensional problem, J simulations result in J underestimators and J overestimators. The linear facets of which result in $2J$ number of $n+1$ dimensional simplices (polytopes) each having 2^n vertices at the domain bounds and 1 interior vertex (simulation point) pertaining to that particular simplex j.

2.2. Bounded Surrogate Model Parameter Estimation Formulation

We limit our focus on generating regression based non-interpolating polynomial surrogates. To attain guarantee in the type of approximation, we bound the surrogate prediction using the linear facets of the estimators, i.e., enforce necessary constraints so that the surrogate prediction lies within a prescribed error bound. To that end, we generate the under and over-estimators from f_i^L and f_i^U which correspond to the shifted points (below and above respectively) of the sampled point f_i. The mathematical formulation is as follows:

$$\min \sum_{n=1}^{N}\left[\left(\frac{\hat{f}_i - s_i}{\hat{f}_i}\right)^2 + \left(\frac{f_i^U - \hat{f}_i}{\hat{f}_i}\right)^2 + \left(\frac{\hat{f}_i - f_i^L}{\hat{f}_i}\right)^2\right] \tag{5}$$

s.t.

$$s_i = \alpha + \sum_{n=1}^{N}\beta_n \hat{x}_{i,n} + \sum_{n=1}^{N}\sum_{m=n}^{N}\gamma_{n,m}\hat{x}_{i,n}\hat{x}_{i,m} \tag{6}$$

$$f_i^L - \sum_{n=1}^{N}\theta_n^{f,L}\cdot(x_{i,n,v} - x_i^{Int})^2 = s_i - \sum_{n=1}^{N}\theta_n^{s,L}\cdot(x_{i,n,v} - x_{i,n}^{Int})^2 \qquad i \in I, v \in V \tag{7}$$

$$f_i^U + \sum_{n=1}^{N}\theta_n^{f,U}\cdot(x_{i,n,v} - x_i^{Int})^2 = s_i + \sum_{n=1}^{N}\theta_n^{s,U}\cdot(x_{i,n,v} - x_{i,n}^{Int})^2 \qquad i \in I, v \in V \tag{8}$$

$$\theta_n^{s,L} \geq \gamma_{n,n} \qquad n \in N \tag{9}$$

$$\theta_n^{s,U} \geq -\gamma_{n,n} \qquad n \in N \tag{10}$$

The first term in the objective function (Eq. 5) minimizes the error of the surrogate fit from the sampled data while the second and third terms reduce the shifts of f_i^L and f_i^U from the original sampled point. Eq. 6 denotes the general form of an n-dimensional quadratic function where α, β_n and $\gamma_{n,m}$ are the estimated parameters. Eqs. 7 and 8 ensure that the linear facets of the under- and over-estimator of the surrogate fit are bounded by the linear facets of the under- and over-estimator of the original function shifted by some value at each sampled point. Here, $\theta_n^{f,L}$ and $\theta_n^{f,U}$ are the parameters required to construct the under- and over-estimators of the original function respectively. Similarly, $\theta_n^{s,L}$ and $\theta_n^{s,U}$ are the variables required to generate the estimators of the surrogate function. Eqs. 9 and 10 guarantee the edge-concavity and edge-convexity of the under and over-estimators of the surrogate function respectively. For estimating n-dimensional functions using a higher order (> 2) polynomial, the model formulation essentially remains the same. However, in this case, the upper bounds of the second derivatives of the surrogate function depend upon the sign of the estimated parameters. This justifies incorporating mixed-integer logic in Eqs. 9 and 10 for evaluating $\theta_n^{s,L}$ and $\theta_n^{s,U}$.

3. GEMS FRAMEWORK

For a specified model defined by closed functional forms or a system of equations, and the bounds on the independent variables, the GEMS framework (shown in Figure 1) allows data sampling (e.g, via Latin Hypercube sampling). To generate the upper bound of the Hessian (θ), the framework utilizes two different approaches: 1) If the model is of closed form and the symbolic differentiation is tractable, the NLP with an objective function of maximizing the second derivatives subject to the domain bounds, can be solved through GAMS. 2) If the model is not of closed form, or the number of independent variables is high, then the GEMS framework allows efficient computation of the Hessian by Automatic differentiation (AD). AD employs techniques similar to backpropagation and provides numerical values of derivatives (Griewank, 2003).

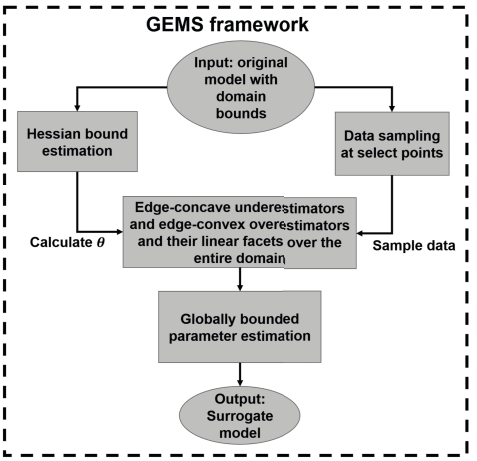


Figure 1: GEMS framework.

The calculated θ values, together with the sampled data are utilized to solve the globally bounded parameter estimation problem (see Section 2.2). The output is a reduced order surrogate polynomial that is guaranteed to not overshoot above and below a certain threshold defined by the relaxed piecewise linear bounding of the edge-concave under and edge-convex overestimators of the original blackbox model. The main advantage of the framework is that it allows the linking between Python (PyTorch), C++ (ADOL-C) for calculating Hessians via automatic differentiation, sampling of the blackbox model via either uniform sampling or Latin hypercube sampling and parameter estimation in GAMS thereby handling the required data transfer and facilitating a user friendly package.

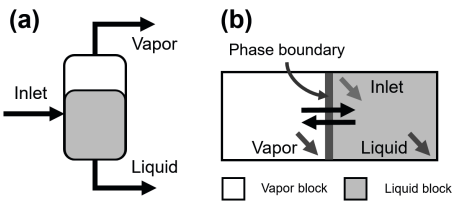


Figure 2: Flash separator: (a) conventional representation and (b) building block-based representation.

4. APPLICATION TO SURROGATE THERMODYNAMIC MODELING

To combat global warming, ILs have garnered significant attention as a potential solvent for the separation of high global warming potential refrigerants, such as hydrofluorocarbons (HFCs). To design innovative and intensified processes for such separation task, one must accurately represent the HFC/IL binary system through

thermodynamic models. Rigorous models such as Gamma-Phi or equation of state (EOS) have higher prediction accuracy. However, these models are highly non-linear and nonconvex, which significantly increase the computational burden for any process synthesis and optimization frameworks. Therefore, a surrogate model with a simpler functional form can evade the model complexity. To that end, we focus on generating surrogate for the well-known Gamma-Phi thermodynamic model, which for the prediction of the amount of HFC absorbed in IL, can be simplified as follows:

$$f(P,T,\widetilde{x_1}) = \ln P + \frac{(P-P_1^s)(B_1-\widetilde{V_1})}{RT} - \ln P_1^s - \ln \widetilde{x_1} - \ln \gamma_1 \tag{11}$$

Here, P_1^s is the saturated vapor pressure, B_1 is the second virial coefficient, $\widetilde{V_1}$ is the molar volume of the ionic liquid, P is the total pressure of the system, $\widetilde{x_1}$ is the equilibrium liquid phase composition of HFC. Activity coefficent model for component i, (γ_i) can be represented via well-known Margules or NRTL model. More details on the thermodynamic modeling can be found elsewhere (Shiflett and Yokozeki, 2006).

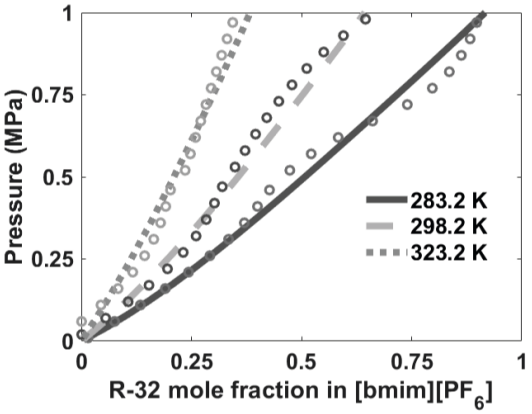


Figure 3: Solubility isotherms of R-32 with Margules activity coefficient model. The lines represents the solubility isotherms by the Gamma-phi based method, the symbol (o) represent the solubility predicted by the surrogate model.

Here we consider R-32/[bmim][PF_6] binary system. We derive the analytical expression of the Hessian from the thermodynamic model (see Eq. 11.) Given the variable bounds, i.e., $[P^L,P^U],[T^L,T^U],[\widetilde{x_1}^L,\widetilde{x_1}^U]$, we determine the required $\theta_P^L,\theta_T^L,\theta_{\widetilde{x_1}}^L,\theta_P^U,\theta_T^U,\theta_{\widetilde{x_1}}^U$ (see Eq. 2 and 4) which are used to construct the piecewise linear bounding of the Gamma-Phi model. Since the goal is to predict mole fraction, $\widetilde{x_1}$, for a given P,T; we divide the entire P,T space into four subregions as follows: $R1 = \{P,T : P \in \{0.01,0.5\}, T \in \{280,330\}\}$, $R2 = \{P,T : P \in \{0.5,1\}, T \in \{280,330\}\}$, $R3 = \{P,T : P \in \{0.01,0.5\}, T \in \{330,375\}\}$, $R4 = \{P,T : P \in \{0.5,1\}, T \in \{330,375\}\}$. For each of the subregions, we apply the model formulation as described in Section 2.2, and obtain a cubic surrogate polynomial.

After that, we incorporate the surrogate to a flash-separator (see Figure 2) through SPICE (Monjur et al., 2021a,b) framework that leverages building-block based representation followed by superstructure optimization. To represent the vapor-liquid phases using building blocks, we require two blocks. The phase contact is represented by a semi-restricted boundary depicted by the thick vertical line in Figure 2b. Mass transfer between the phases takes place through this boundary which is represented by the VLE model (see Eq. 11).

When the surrogate prediction from the flash separator is compared to the Gamma-Phi prediction (see Figure 3), we observe that the averaged prediction error lies within $\approx$ 8.45%. We obtain four different best fitted surrogates for each of the subregions which explains the discontinuity at P = 0.5 MPa for each of the isotherms in Figure 3. It can be seen that the surrogates intersect the Gamma-Phi based isotherms more than once. This may suggest that the surrogate model may be of higher order than the original function. Since we assumed a simple polynomial form, i.e. cubic, the possibility of exponential or logarithmic terms in the actual function may also lead to this behavior. The main takeaway point is that through the GEMS framework, we are able to replace the nonlinear thermodynamics by a simpler surrogate with guaranteed error bounds, allowing us to use the same problem formulation for process synthesis applications.

5. CONCLUSIONS

The data-driven approach presented through the GEMS framework can be efficiently applied for the approximation of VLE models. It also shows a promising pathway for solving general data-driven global optimization problems. Incorporation of automatic differentiation in the framework allows the calculation of the upper bound on the Hessian even if a system of equations rather than an explicit functional form of the model is available. The approach could be an efficient way for accelerating computationally demanding process simulations employing complex models by proposing simple and more computationally favorable accurate enough surrogates, thereby providing a means to attain the global convergence of data-driven process optimization problems.

6. Acknowledgment

The authors gratefully acknowledge support from the NSF EFRI DCheM (2029354) and NSF CAREER (CBET-1943479) grants.

References

I. Bajaj, A. Arora, M. F. Hasan, 2021. Springer, pp. 35–65.

I. Bajaj, M. F. Hasan, 2019, 1–16.

A. Bhosekar, M. Ierapetritou, 2018. Computers Chemical Engineering 108, 250–267.

F. Boukouvala, M. F. Hasan, C. A. Floudas, 2017. Journal of Global Optimization 67 (1-2), 3–42.

A. Cozad, N. V. Sahinidis, D. C. Miller, 2014. AIChE Journal 60 (6), 2211–2227.

A. Griewank, 2003. Acta Numerica 12, 321–398.

M. M. F. Hasan, 2018. Journal of Global Optimization 71 (4), 735–752.

D. R. Jones, Dec 2001. Journal of Global Optimization 21 (4), 345–383.

D. R. Jones, M. Schonlau, W. J. Welch, Dec 1998. Journal of Global Optimization 13 (4), 455–492.

M. S. Monjur, S. E. Demirel, J. Li, M. M. F. Hasan, 2021a. Comput. Aided Chem. Eng 50, 287–293.

M. S. Monjur, S. E. Demirel, J. Li, M. M. F. Hasan, 2021b. Ind. Eng. Chem. Res.

M. B. Shiflett, A. Yokozeki, 2006. AIChE Journal 52 (3), 1205–1219.

F. Tardella, 2004. Frontiers in Global Optimization, 563–573.

C. Wang, Q. Duan, W. Gong, A. Ye, Z. Di, C. Miao, 2014. Environmental Modelling Software 60, 167–179.

Proceedings of the 14thInternational Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022,Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50306-7

Development of an ANN-based soft-sensor to estimate pH variations in Intelligent Packaging Systems with visual indicators

Isadora F. Brazolin[a]*, Felipe Matheus Mota Sousa [b]; Flavio Vasconcelos Silva [b]; Viktor O. C. Concha[a]; Cristiana M. P. Yoshida[a]

[a]*UNIFESP - Federal University of São Paulo, Institute of Ambiental, Chemical, and Pharmaceutical Science, São Paulo, BRAZIL*
[b]*UNICAMP – University of Campinas, School of Chemical Engineering, São Paulo, BRAZIL*
ibrazolin@unifesp.com.br

Abstract

Based on the experimental data of colorimetric indication, an artificial neural network was first established to classify the pH ranges of the intelligent food packaging device. An intelligent packaging system monitors the package product's condition to provide information about the quality and/or safety during transport, distribution, and storage. The intelligent packaging senses and informs the conditions of the product in an easy and accessible manner, without opening the package. Food pH is strongly related to the quality of food packaged products, indicating deterioration, microbial growth, and adulteration. In the case study, the development and training of an artificial neural network (ANN) aimed to easy quality control of food products that can present alterations/adulterations from pH variation reactions, based on a functional colorimetric indicators' response from a sustainable, intelligent packaging device (biopolymeric chitosan films) of easy and renewable source manufacturing. Chitosan intelligent films were formulated with different chitosan and natural colorimetric indicator (anthocyanin) concentrations, forming the intelligent device. The intelligent devices were immersed in a wide pH range (1.0 to 13.0) solutions, and color parameters (L*, a*, b*) variations were measured. An empirical multivariable model was developed based on artificial intelligence (ANN) to classify pH ranges through the indicator's color variation and the chitosan and anthocyanin concentrations. The ANN of chitosan intelligent films device could ensure acceptable food quality and safety levels to provide adequate protection for consumers and facilitate trade.

Keywords: Intelligent Packaging; Machine Learning; IA; colorimetric indicator

1. Introduction

Packaging technologies are being developed to improve products preservation, quality, and safety. Among recent technologies, intelligent packaging is products' condition monitoring systems, providing quality information during transportation, distribution, and storage. The intelligent packaging device senses the environment inside or outside the package and informs the manufacturer, retailer, and consumer regarding the product's condition (Kuswandi et al., 2011).

Food products' shelf-life tests demand time and cost, while colorimetric indicators can assist the consumer or retailer when buying the products, ensuring quality and food safety. Supply chain management based on pH measurements can significantly decrease food

waste, a critical environmental and social concern. An efficient supply chain management could save food disposal, water, energy, increase return-on-investment, improve consumer satisfaction and support regulatory requirements (Mercier & Uysal, 2018). Yoshida et al. (2014) developed an easy-manufacturing and sustainable colorimetric indicator using anthocyanin as a pH colorimetric indicator incorporated into a natural polymer matrix (chitosan). The intelligent films presented final properties to be applied as intelligent devices material, with a color variation due to pH range, which was observed pink color to acid pH, green-blue color to neutral pH, and yellow color to basic pH.

The food industry, like many others, benefits a lot from modernization and the use of technology. Industry 4.0 brings connections and interactions between machines and operators from different sectors using artificial intelligence, including artificial neural networks (ANN). ANN is a computational model built from several simple processing units (neurons) capable of assimilating data and information presented and, from the acquired knowledge, estimating solutions that were not known until now. In this way, an ANN simulates the nervous system behavior of a living being. The definitions involved make an analogy to the components and processes related to the functioning of the human brain (Haykin, 2009).

This work aimed to apply artificial intelligence, specifically artificial neural networks, to determine the pH value range displayed on the sustainable colorimetric indicating device. The empirical pH model was obtained by varying the formulation (concentrations of chitosan and anthocyanin) and measuring the respective color parameters measured in a colorimeter.

2. Materials and methods

2.1. Chitosan Intelligent films

The chitosan intelligent films were obtained accordingly to (Yoshida et al., 2014), using different anthocyanin (Cath, 0.5, 1.0 e 2.0 % m/m) and chitosan (Cch, 0.5, 1.0 e 2.0 % m/m) concentrations.

2.2 Color indication intelligent device

The different pH values were measured using standard solutions in a wide range of pH from 1.20 (HCl 1 mol/L) to 13.29 (NaOH 1 mol/L) using MilliQ water, generating the data required to train the artificial neural network. Buffer solutions were prepared to obtain intermediate pH solutions: McIlvaine, Kolthoff, boric acid-potassium chloride-sodium hydroxide. Chitosan intelligent films were immersed in pH solutions for ninety seconds. Instrumental color parameters (L*, a*, b*) were measured using a portable colorimeter (Konica Minolta, CR-400, Osaka, Japan) in three random positions.

2.3 *Development of empirical mathematical models using Artificial Neural Networks*

Experimental tests were carried out to create a database used to determine the empirical model of the pH of the chitosan film. Chitosan and anthocyanin concentrations and measured color parameters (L *, a *, and b *) were used as ANN input variables. The pH value of the chitosan film was used to create three distinct classes for the network output. The database was divided into three sets for use in the training, validation, and testing phases of the ANNs. The database consists of 430 data, each one containing 5 inputs (Cch, Cath, L *, a *, and b *) and 1 output (pH class).

The ANNs were developed using the dedicated libraries Tensorflow and Keras. Adam optimization was used as an optimizer, and the categorical cross-entropy was chosen as the objective function. The experimental database was divided randomly into training and

test datasets with 80 and 20 % of the original set. The input variables were normalized in the range of -1 and 1 to ensure unbiased models. Three pH ranges were selected as labels for the outputs corresponding to pH under 4, between 4 and 8, and over 8. Early-stopping was introduced during the training stage to avoid model overfitting, using 10 % of the training dataset after performing cross-validation with the selected hyper-parameters summarized in Table 1. It was used 10-fold cross-validation to improve the ANN model's choosing, eliminating usually found consequences of random choosing.

Different architectures of multilayer feedback artificial neural networks (hidden layers and an output layer) were proposed. Python, an interpreted language guaranteeing a free tool with easy installation and platform independence, was used to implement computational models. The hyper-parameters were optimized using a grid-search algorithm. The output layer is a three-neuron-densely-connected layer with a SoftMax activation function to determine the likelihood of each category. For each input vector, the output with the highest value was selected as the predicted output for direct comparison with the real data.

The methodology applied to this case study is illustrated in Figure 1.

Table 1. Hyper-parameters for this classification problem.

Hyper-parameter	Values
Hidden layers	1 and 2
Neurons	10, 20, 30, 40, 50, 60, 70, 80, 90, 100
Weight initialization	Identity, random normal, and random uniform
Activation function	Hyperbolic tangent, sigmoid and ReLU

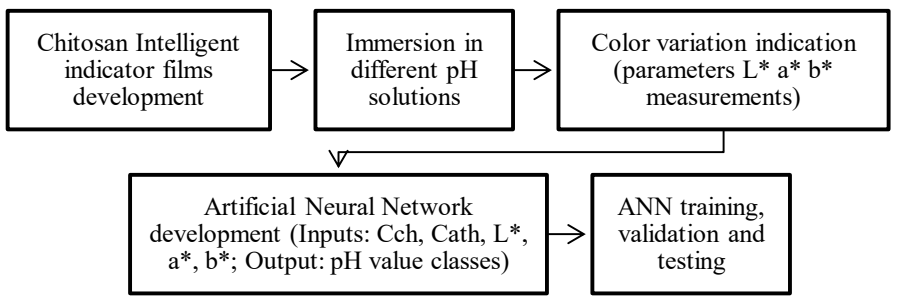
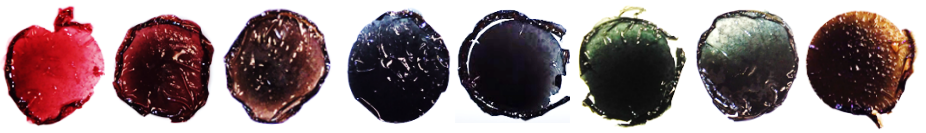


Figure 1. Scheme of the methodology adopted.

3. Results

Chitosan intelligent films containing different concentrations of anthocyanin presented a violet color. Films with lower chitosan concentration were more transparent and flexible, facilitating visual color variation. The chitosan intelligent films were immersed in different solutions in a wide pH range (1.20 to 12.58), from acidic to alkaline (Figure 2). The chitosan intelligent film device presented a reddish color for the pH range from 1 to 5 (acid condition), a blue-greenish color for the range 6 to 10, changing to a yellowish color in a more alkaline range (pH > 12). For the pH values of 7.77 and 8.87 and 3.79 and

5.87, the films presented a very similar color even though representing different pH values.



(a) 1.20 (b) 3.79 (c) 5.87 (d)7.77 (e) 8.87 (f) 9.96 (g)11.40 (h)12.58

Figure 2. Visual color variation of chitosan intelligent film device in contact with different pH conditions (a-h).

In this study, an ANN was designed to classify the pH range of a chitosan film from the concentrations of chitosan and anthocyanin used in the film formulation and from the measured color parameters (Cath, Cch, L *, a *, b *), as seen in Figure 3.

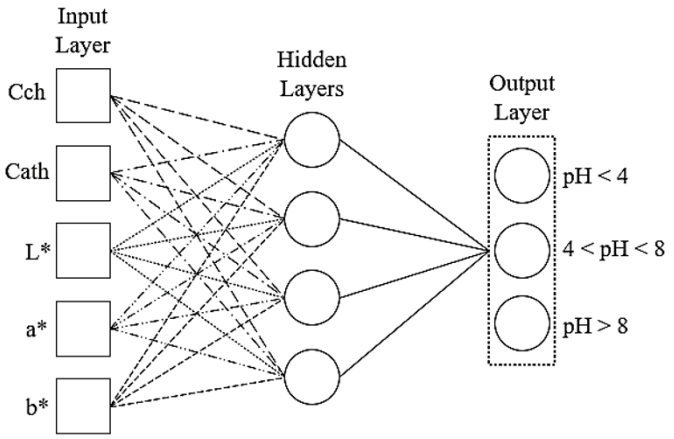


Figure 3. Representation of a neural network used to predict the pH at chitosan intelligent from the device.

The database used for the cross-validation consists of the pH values of 1.20, 3.79, 5.87, 7.77, 9.96, and 12.58 due to the previously exposed problem of having similar colors representing different pH values. This way was observed a better-quality training of the neural network, consequently giving better prediction results.

Several models were generated involving changes in the number of hidden layers (1 or 2), number of neurons (10 to 100), weight initialization (identity, random normal and random uniform), and activation function (hyperbolic tangent "tanh," rectified linear unit "ReLU" and sigmoid). Table 2 summarizes the best model regarding test accuracy.

Table 2. Best generated model

Hidden Layers	Neurons 1st layer	Neurons 2nd layer	Weight initialization	Activation function	Training accuracy	Precision Mean
2	80	20	Random normal	tanh	0.79	0.62

The following results represent the best model illustrated in Table 2. The confusion matrix is shown in Figure 4.

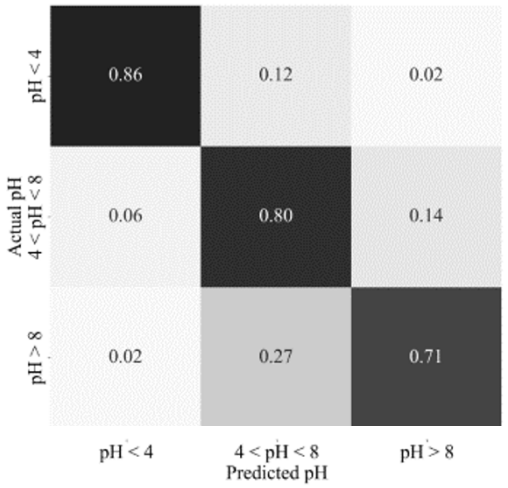


Figure 4. Confusion matrix: a summary of the prediction results of this study's best model.

The accuracy for the first and second classes was over 80 %, while the "pH > 8" class was misclassified as the "4 < pH < 8" class at a maximum rate of 27 %. This misclassification can result from similar color parameters (inputs L*, a*, b*) of the intelligent films immersed in pH values solutions closest to the edge of the two last classes (pH values 7 to 9, as seen in Figure 2.). Previous types show more accurate results due to more significant color changes in the films, and consequently, distinct color parameters, allowing the network to distinguish the pH ranges more efficiently. Again, the accuracy presented can result from similar film colors on the edge of both pH ranges.

Due to the relatively low test accuracy of the model (79 %) and the favorable misclassification rate of the third class, classification algorithms such as decision trees and support vector machines could be tested to obtain better pH-classification models. Mainly, the ANN applied to intelligent packaging is unexplored. The devices are based on biopolymer matrix formation that could not form a standard linkage between the chains. It is necessary to get more experimental data for successful of AI techniques. However, this is a very promising area in the future.

4. Conclusions

The artificial neural network was developed with data obtained from the color parameters for all formulations of chitosan intelligent films devices in contact with different pH solutions (from 1.20 to 12.58) and chitosan and anthocyanin concentration values. The chitosan intelligent films formulation with the most efficient colorimetric results were Cch = 0.5 % and Cath = 0.5 % (w/w).

A classification model to identify the devices out of the non-spoilage range (pH values from 4 to 8) ensures a better standard of quality and safety of the food product during the supply chain, also allowing the use of the model as a software sensor, assisting in the decision-making of changes. The best ANN showed a decent generalization accuracy,

about 79 %, but still below the desirable rate. This could have resulted from similar sensor colors on the classes' edges, leading to a misconception of the ANN. However, the results showed that classification algorithms based on colorimetric measurements could be explored to indicate alterations/adulterations from pH variation reactions.

The sensor device formulation needs to be improved to show the most significant distinction between color changes; that way, it would be possible to use a more extensive database with more pH values, leading to a better trained ANN.

The commercial implementation of the sustainable, intelligent device is still a challenge, but a global market of food products with strict laws can be good support.

Acknowledgments

Capes - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior and FAPESP – Fundação de Amparo a Pesquisa

References

Haykin, S. (2009). *Neural networks and learning Third Edition*. *Institute of Physics Conference Series* (Vol. 127).

Kuswandi, B., Wicaksono, Y., Jayus, Abdullah, A., Heng, L. Y., & Ahmad, M. (2011). Smart packaging: Sensors for monitoring of food quality and safety. *Sensing and Instrumentation for Food Quality and Safety*, *5*(3–4), 137–146. https://doi.org/10.1007/s11694-011-9120-x

Mercier, S., & Uysal, I. (2018). Neural network models for predicting perishable food temperatures along the supply chain. *Biosystems Engineering*, *171*, 91–100. https://doi.org/10.1016/j.biosystemseng.2018.04.016

Yoshida, C. M. P., Maciel, V. B. V., Mendonça, M. E. D., & Franco, T. T. (2014). Chitosan biobased and intelligent films: Monitoring pH variations. *LWT - Food Science and Technology*, *55*(1), 83–89. https://doi.org/10.1016/j.lwt.2013.09.015

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50307-9

Process Systems Engineering Guided Machine Learning for Speech Disorder Screening in Children

Farnaz Yousefi Zowj[a], Kerul Suthar[a], Marisha Speights Atkins[b], Q. Peter He[a*]

[a]*Department of Chemical Engineering, Auburn University, Auburn, AL 36849, USA*
[b]*Communication Sciences & Disorders, Northwestern University, Evanston, IL 60201, USA*

qhe@auburn.edu

Abstract

Auditory perceptual analysis (APA) is the primary method for clinical assessment of speech-language deficits, one of the most prevalent childhood disabilities. Due to multiple limitations of APA including being susceptible to intra- and inter-rater variabilities, automated methods such as Landmark (LM) analysis that quantify speech patterns for diagnosing speech disorders in children are developed. This work investigates the utilization of LMs for automatic speech disorder detection in children. Leveraging the similarities between disease detection in medical/clinical research and fault detection in process systems engineering (PSE), we propose to improve the detection of speech disorder in children via PSE principles. Specifically, the parsimony principle is followed for reducing feature and parameter spaces. Domain knowledge is utilized for generating a set of novel knowledge-based features to address the challenge of large within-class variations in LM measurements. A systematic study and comparison of different linear and nonlinear machine learning classification techniques are conducted to assess the effectiveness of the novel features in classifying speech disorder patients from normal speakers.

Keywords: systems engineering, machine learning, feature engineering, child speech disorder, landmark

1. Introduction

Speech-language deficits are one of the most prevalent childhood disabilities affecting about 1 in 12 children between the ages of three and five years old. Approximately 40 % of children with speech and language disorders do not receive intervention because their impairment goes undetected (Nelson et al., 2006). Auditory perceptual analysis (APA) is the main method for clinical assessment of disordered speech; however, results from APA are susceptible to intra- and inter-rater variabilities. Another factor to consider is that some children may be reluctant to participate in long testing sessions, and even if they do, transcription of large data sets of audio recordings is time-consuming and requires a high level of expertise from therapists. These limitations of manual or hand transcription based diagnostic assessment methods have led to an increasing need for automated methods to quickly and consistently quantify child speech patterns. Landmark (LM) analysis is such an approach that characterizes speech with acoustic markers that are developed

based on the LM theory of speech perception. LM analysis has been suggested as the basis for automatic speech analysis (Ishikawa et al., 2017). Therefore, in this work we focus on the utilization of LMs for automatic speech disorder detection in children, with LMs extracted using SpeechMark® toolbox (Boyce et al., 2012). The description of each landmark detected by this tool and used in this study are presented in Table 1.

Different LM-based features have been proposed in the literature. The most common one is the counting of individual LMs, a.k.a. unigrams, which does not consider the specific order or sequence of the LMs. n-gram, which is a generalization of unigrams and is defined as a sequence of n consecutive LMs, takes the specific LM order into consideration when $n \geq 2$. It was found that n-gram counts (n=1, 2, 3, 4) were good features for depression detection. Besides n-gram count, time based LM features have also been proposed in the literature. These time based LM features include durations of the bigrams (i.e., 2-grams) and LM pairs (i.e., onset and offset of a LM as defined in Table 1) (Huang et al., 2019), and speech rate, which is defined as the number of phonetic units, such as syllables or words, uttered per unit time (Huici et al., 2016). Syllabic cluster (SC) analysis, which clusters LMs into syllabic units, has also been found to be an important feature for speech disorder detection (Boyce et al., 2012; Atkins et al., 2019).

Systems engineering principles have been instrumental in analyzing various biological data. For instance, we have recently reviewed a large body of research that utilizes PSE principles and techniques to address some of the technical challenges in Big Data analytics for biological, biomedical and healthcare applications, including the principle of parsimony in addressing overfitting, the dynamic analysis of biological data and the role of domain knowledge in biological data analytics (He and Wang, 2020). This work aims to improve speech disorder detection following PSE principles. The basic idea is that, despite the extremely different physical implementations, a human body can be viewed as a complicated biochemical plant and they share many common features at the system level. For example, a human disease or disorder can be viewed as an anomaly (or "fault" in a PSE term) in a human body. As a result, the principles and techniques developed for fault detection and diagnosis in the PSE community can be adopted to address some of the challenges in disease detection. The first step when developing a model is to select input features or variables from the PSE perspective. In particular, many PSE applications have demonstrated that raw features are often not the best features for capturing process characteristics, while engineered features with statistical and/or physical meanings are often more informative and robust in characterizing process behavior, and therefore are more effective in fault/disease detection (He and Wang, 2011, 2020; Lee et al., 2020; Shah et al., 2020; Suthar and He, 2021; Wang and He, 2010). Specifically, we propose novel knowledge-based features that are the ratios of the count of n-grams ($n \geq 2$) to that of unigrams. Ratios are usually better features than the absolute individual values in addressing the individual variations of samples within the same class. In addition, the parsimony principle of PSE leads us to develop robust models by reducing feature space (through feature selection) and parameter space (e.g., through the utilization of simple linear models). The final contribution of this work is a systematic study and comparison of different linear and nonlinear machine learning classification techniques and their effectiveness in classifying speech disorder patients from normal speakers.

The remainder of this work is organized as follows: Section 2 describes materials used

in this study, the features proposed in this work, and introduces the analytical methods. Section 3 presents results and discussions of this work, and Section 4 draws conclusions.

2. Materials and Methods

2.1. Data description

The speech of 51 children ages 33 - 94 months was retrieved from the Speech Evaluation and Exemplars Database (SEED) (Speights Atkins et al., 2020). 39 were typically developing, and 12 with speech sound disorder without language impairment. Speech samples retrieved for this study were recordings of children uttering the word "flower" which is one of the 11 triage words of the set from Anderson and Cohen (2012).

2.2. Methodology

The raw features extracted from audio recordings using the SpeechMark MATLAB Toolbox include time stamp and strength of each LM listed in Table 1, plus SC count. We have adopted all LM and SC based features proposed in the literature, including n-gram counts, and duration and rate features based on LMs and n-grams. In addition, we explore LM strength based features and propose n-gram ratio based features to better address within-class variations. We have 189 engineered features after removing illegitimate or trivial features (e.g., n-gram counts that are all zeros, or ratios with a denominator of zero) and redundant features (i.e., the features that are highly correlated with an existing feature- Pearson correlation coefficient of 0.99). Recursive feature elimination with cross-validation (RFECV) from scikit-learn is utilized for feature selection with the default 5-fold cross-validation with a linear discriminant analysis (LDA) model as the classifier. Result indicates that only 10 features are needed to obtain the optimal cross-validation score. Nine out of the ten features (seven ratio-based and two strength-based features) are new features proposed in this work that have not been utilized before.

There is approximately a 3:1 class imbalance between the normal speaker samples and the disordered ones. The synthetic minority over-sampling technique (SMOTE) is utilized in this work, in which new samples are synthesized from the existing samples. Once the training set is balanced using SMOTE, we train four different classification algorithms, namely linear discriminant analysis(LDA), support vector machine (SVM), extreme gradient boosting (XGBoost), and random forest (RF), tune their hyperparameters with a

Table 1: Description of landmarks used in this study

Landmark	Description
g (glottis)	Onset (+) and offset (-) of sustained motion of vocal fold
b (burst)	Onset (+) and offset (-) of frication or bursts in an unvoiced segment
s (syllabicity)	Release (+) and closure (-) of sonorant consonant in a voiced segment
f (unvoiced frication)	Onset (+) and offset (-) of frication in an unvoiced segment
v (voiced frication)	Onset (+) and offset (-) of frication of in a voiced segment

10-fold stratified cross-validation (CV), apply the models to the left-out test samples, and report the sensitivity and specificity which are two most commonly used critical metrics when dealing with binary classification problems in healthcare. This whole procedure is referred to as one Monte-Carlo validation and testing (MCVT). We report the mean and standard deviation of sensitivity and specificity of 50 such MCVT runs, which is a robust way of comparing different modeling techniques and assessing their performances. Sensitivity is the true positive rate, i.e., the classifier's ability to detect diseased patients correctly, and specificity is the true negative rate, i.e., the classifier's ability to detect normal controls (i.e., the ones without diseases) correctly. We also use accuracy as a single measure when we need to evaluate the overall performance of a classifier. Throughout the modeling procedure, grid search and random search are used for hyperparameter tuning.

3. Results and discussion

In this work, we conduct investigation from two perspectives: (1) comparing classification performance when different feature sets are used, and (2) comparing classification performance when different classification techniques are used. When comparing different features, the following two feature sets are studied: (a) the original 21 features directly obtained from the SpeechMark Toolbox, which include the counts and strengths of the ten LMs (listed in Table 1, considering both onset and offset) for each sample, plus one syllabic count per sample; and (b) The ten selected features selected via RFECV.

As shown in Table 2, when the 21 raw features are used, SVM with RBF kernel provides the best overall classification performance with 75.0 % accuracy (i.e., 75.0 % of the samples are classified correctly) and has somewhat a balanced specificity and sensitivity of 80.0 % and 70.0 %, respectively. LDA provides the second-best result with 71.0 % accuracy. The overall performances of all methods, linear or nonlinear, are relatively poor, indicating that the raw features are not very informative in classifying the two classes.

Next, we apply different classification methods to the selected ten features obtained through rational feature engineering and selection. The results are listed in Table 3 and shown in Figure 1. By comparing Table 2 and 3, we can see that the performances of all methods have significantly improved classification accuracy. The significantly improved performance with these features across all classification methods demonstrates that the proposed features are more informative than the raw features. Since there are seven features that are

Table 2: Comparison of classification performance based on raw features

Method	Sensitivity	Specificity	Accuracy
LDA	64	54	59
SVM (Linear)	68	74	71
SVM (Poly)	78	32	55
SVM (RBF)	70	80	75
SVM (Sigmoid)	80	54	67
XGBoost	50	86	68
RF	28	76	52

Table 3: Comparison of classification performance based on rationally engineered and selected features

Method	Sensitivity	Specificity	Accuracy
LDA	94.0	92.0	93.0
SVM (Linear)	86.0	92.0	89.0
SVM (Poly)	84.0	94.0	89.0
SVM (RBF)	72.0	94.0	83.0
SVM (Sigmoid)	88.0	88.0	88.0
XGBoost	60.0	88.0	74.0
RF	76.0	94.0	85.0

ratio based, the improved performance is most likely due to our hypothesis that ratio-based features are better at addressing individual variations of samples from the same class. In particular, LDA classifier achieves 94.0 %, 92.0 % and 93.0 % in sensitivity, specificity, and overall accuracy respectively. Several other methods also achieve nearly 90.0 % in sensitivity, specificity and overall accuracy, including SVM with linear, polynomial and sigmoid kernels. In comparison to the raw features, the sensitivity and specificity based on the selected engineered features are much more balanced.

4. Conclusions

In this work, we propose an automated computer-assisted screening method for detecting speech disorder in children following PSE principles. Specifically, the proposed knowledge-based features have been found particularly informative in characterizing audio recordings for speech disorder detection, and the parsimonious models derived from these features are found to be not only accurate but also robust. It is demonstrated that, with raw features, all classification methods fail to achieve high classification performance. In comparison, with the ten selected features, which contain nine features proposed in this work, the performances of all classification methods are significantly improved indicating that the proposed features are more effective for characterizing speech disorder using speech LMs. With knowledge-based features, LDA achieves a classification sensitivity of 94.0 %, speci-

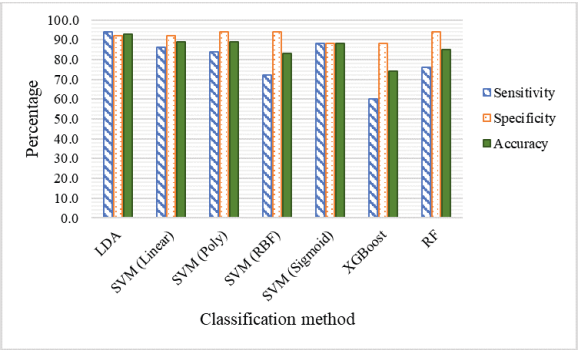


Figure 1: Comparison of classification performance when selected features are used

ficity of 92.0 %, and overall accuracy of 94.0 % compared to the SVM with RBF kernel using raw features reaching 70.0 % sensitivity, 80.0 % specificity, and 75.0 % overall accuracy. This work demonstrates that integration of domain knowledge into ML techniques can significantly improve the performance of purely data-driven or data-centric methods.

References

Anderson, C. and Cohen, W. (2012). Measuring word complexity in speech screening: single-word sampling to identify phonological delay/disorder in preschool children. *International journal of language & communication disorders*, 47(5):534–541.

Atkins, M. S., Boyce, S. E., MacAuslan, J., and Silbert, N. (2019). Computer-assisted syllable complexity analysis of continuous speech as a measure of child speech disorders. In *Proceedings of the 19th International Congress of Phonetic Sciences,(ICPhS 2019), Melbourne, Australia*, pages 4–10.

Boyce, S., Fell, H., and MacAuslan, J. (2012). Speechmark: Landmark detection tool for speech analysis. In *Thirteenth Annual Conference of the International Speech Communication Association*.

He, Q. P. and Wang, J. (2011). Statistics pattern analysis: A new process monitoring framework and its application to semiconductor batch processes. *AIChE journal*, 57(1):107–121.

He, Q. P. and Wang, J. (2020). Application of systems engineering principles and techniques in biological big data analytics: A review. *Processes*, 8(8):951.

Huang, Z., Epps, J., and Joachim, D. (2019). Investigation of speech landmark patterns for depression detection. *IEEE Transactions on Affective Computing*.

Huici, H.-D., Kairuz, H. A., Martens, H., Van Nuffelen, G., and De Bodt, M. (2016). Speech rate estimation in disordered speech based on spectral landmark detection. *Biomedical signal processing and control*, 27:1–6.

Ishikawa, K., MacAuslan, J., and Boyce, S. (2017). Toward clinical application of landmark-based speech analysis: Landmark expression in normal adult speech. *The Journal of the Acoustical Society of America*, 142(5):EL441–EL447.

Lee, J., Kumar, A., Flores-Cerrillo, J., Wang, J., and He, Q. P. (2020). Feature based fault detection for pressure swing adsorption processes. *IFAC-PapersOnLine*, 53(2):11301–11306.

Nelson, H. D., Nygren, P., Walker, M., and Panoscha, R. (2006). Screening for speech and language delay in preschool children: systematic evidence review for the us preventive services task force. *Pediatrics*, 117(2):e298–e319.

Shah, D., Wang, J., and He, Q. P. (2020). Feature engineering in big data analytics for iot-enabled smart manufacturing–comparison between deep learning and statistical learning. *Computers & Chemical Engineering*, 141:106970.

Speights Atkins, M., Bailey, D. J., and Boyce, S. E. (2020). Speech exemplar and evaluation database (seed) for clinical training in articulatory phonetics and speech science. *Clinical linguistics & phonetics*, 34(9):878–886.

Suthar, K. and He, Q. P. (2021). Multiclass moisture classification in woodchips using iiot wi-fi and machine learning techniques. *Computers & Chemical Engineering*, 154:107445.

Wang, J. and He, Q. P. (2010). Multivariate statistical process monitoring based on statistics pattern analysis. *Industrial & Engineering Chemistry Research*, 49(17):7858–7869.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50308-0

Emission and mitigation of CO_2 and CH_4 produced by cattle: a case study in the Brazilian Pantanal

Victor G. Moretti[a], Celma O. Ribeiro[a*], Claudio A. Oller do Nascimento[b], Julia Tomei[c], Alison J. Fairbrass[c]

[a] *Department of Production Engineering, EPUSP - Polytechnic School of the University of São Paulo, SP, Brazil*
[b] *Department of Chemical Engineering, EPUSP - Polytechnic School of the University of São Paulo, SP, Brazil*
[c] *UCL Institute for Sustainable Resources, University College London, London, UK*
v.gmoretti94@gmail.com

Abstract

This study aims to analyse the effects on emissions and land use resulting from the increase of productivity in cattle farming in Brazil, focusing on the Brazilian Pantanal region. Considering public data from the municipalities comprising the Pantanal, the study considers alternative scenarios for the reduction of ranching areas through sustainable intensification and technological improvements, identifying the effects of alternative policies on emissions and natural vegetation preservation. By employing a System Dynamics model, the study analyses the effects of these policies and identifies the relationship between cattle intensive practices, land use distribution and CH_4 emissions.

Keywords: Energy, Food and Environmental Systems, Emissions, Brazilian Pantanal, System dynamics

1. Introduction

The increase in the demand for agricultural commodities and meat products in the world has a significant impact on the carbon footprint and in the land use. According to FAO, the global livestock is responsible for 7.1 Gigatons of CO2-equiv per year, with cattle being responsible for about 65 % of the livestock sector's emissions (Gerber, 2013). Global methane emissions were addressed at the COP-26 event, held in November/2021 in Glasgow, Scotland, where the Methane Agreement was signed, establishing the global commitment to cut the gas emissions by 30% by 2030, bringing great challenges to the Brazilian cattle production chain.

Brazil is the second largest producer and the largest beef exporter in the world, with large part of the livestock's herd raised on pasture. The use of pastures reduces financial costs for the producers but results in a significant reduction of natural vegetation and productivity. The evolution of Brazilian cattle ranching practices in recent decades has increased productivity mainly through technological improvements, but the numbers are still far from ideal. The increase of pasture area from 135 million hectares in 1990 to 167 million hectares in 2019 (MapBiomas, 2020) indicates that there is room for significant improvement in the country.

The Pantanal, surrounded by the Amazonia, Cerrado, and Chaco biomes, is a megadiverse tropical wetland. It supports numerous valuable ecosystem services including the provision of wild foods, environmental regulation, and maintenance services such as carbon storage and sediment retention, a diverse tourism industry, water supply and a growing cattle ranching industry. The biome has recently received global attention due to an alarming increase in the frequency and extent of wildfires, the causes of which are diverse and complex. A challenge for decision-makers is how to report on the environmental and socio-economic impacts of these competing human uses of the biome, to inform decisions that maximize benefits to local, regional, and global communities while maintaining ecosystem integrity.

Based on system dynamics modelling, this study indicates how technological changes in cattle ranching can improve productivity and can lead to a decrease in emissions, and simultaneously to the preservation on the concept of natural capital.

2. Livestock in the Brazilian Pantanal region

The Pantanal encompasses an area of 150,355 km², occupying 1.76% of the Brazilian territory and comprising about 3% of the entire world's wetlands (IBGE, 2004). Brazilian beef cattle development is historically based on the expansion of the agricultural frontier, through the deforestation of regions without infrastructure and lands depleted by agriculture. Brazil has the second largest cattle herd in the world and is the largest meat exporter, exporting around 20% of its production. Even so, it still has productivity rates below other great producers' countries. In Pantanal, the cattle herd had an impressive increase in the last three decades, growing from around 3.5 million to 4.2 million heads (21.8 %) from 1990 to 2018 in the municipalities covered by the bioma (IBGE, 2004), resulting in a conversion of around 2.1 Mha of land to new pastures, that went from 2.6 Mha to 4.7 Mha. The decrease in the savanna and forest formations during the period, from 9,0 Mha to 7.0 Mha (MapBiomas, 2020), indicates the use of part of these lands to ranching activities through deforestation. Although cattle raising methods have improved in some municipalities with intensification of cattle practices, use of new technologies and high-quality production, the efforts have not been enough to assure the Pantanal natural resources conservation.

This study focuses on livestock activities as one of the major factors driving land change and emissions in the Brazilian Pantanal. Negative impacts on ecosystem services resulting from the decrease in native vegetation, methane and CO_2 emissions and socio-economic effects are assumed to be the main results from the significant increase of new pastures of the region. Enteric methane emissions from cattle are responsible for an expressive amount of world's GHG emissions and improving animal productivity is recognized as an important pathway to achieve global sustainable goals.

Alternative public policies can modify the global greenhouse emissions resulting from livestock as presented by Avery Cohn *et al.* (2014), who examined policies to encourage cattle ranching intensification in Brazil as a strategy to reduce GHG emissions. Semi intensive practices may not be the best alternative for Pantanal region, but several alternatives related to productivity increase and emission reductions can be adopted. The strength of the policy depends on wellbeing and the economic value of ecosystem services.

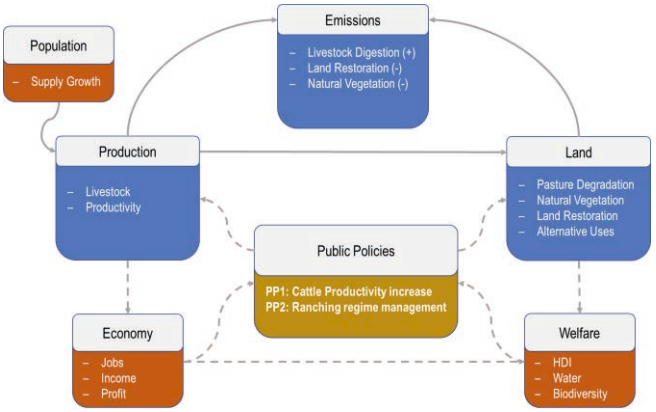


Figure 1 – Main subsystems. Source: Own Elaboration

3. The system dynamics (SD) model

System Dynamics model applies for dynamics complexity related problems in which links of different natures are made through various types of variables, flows, auxiliary variables, and parameters, with simulation modelling based on feedback systems theory. In this paper, the interaction of seven subsystems (land, economy, policies, emissions, welfare, livestock, and population) were considered, based in Fiddaman (2012).

The proposed system dynamics (SD) model illustrates the stock and flow of cattle pasture and pastureland, focusing on the emissions subsystems and measuring the relationships among indicators identified in each subsystem, as showed is Figure 1. For that, it is assumed that the CH_4 resulting from enteric fermentation of the digestive process of the cattle is the main source of emissions in Pantanal. Additionally, the study considers the pastures as the major cause of natural vegetation reduction. Water and biodiversity changes are considered exogenous, possibly resulting from climatic changes. The main driving factors for a policy shift from the Business as Usual (BAU) model to a climate policy focusing on emission reduction through the adoption of good ranching practices and restoration of natural vegetation are assumed to result from social and environmental pressures.

4. Material and methods

For this research, the Insight Maker®, a free simulation tool that runs in the web browser, was used to support the system dynamics application, enabling diagramming, and modelling features and creating representations of the system. The SD model is presented with emphasis on sustainable cattle ranching practices to cover issues on Pantanal over the period of 1990-2019. Through the SD simulation, we can analyse the impact of productivity policies on CH_4 emission and land use and the relationship between cattle intensive practices, natural vegetation restoration and land use distribution.

Historical time series from municipalities comprising the Brazilian Pantanal were gathered from public databases, from 1990 to 2019 in a yearly basis, considering information related to land use, cattle production, and regional socioeconomic indicators. Parameters related to jobs, income, cattle productivity and CH_4 emissions, among others, were obtained from literature.

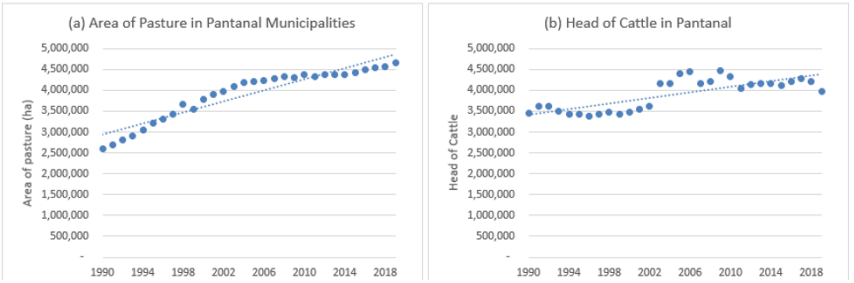


Figure 2 - Main data used in the BAU scenario. Figure 2(a) represents the area of pasture in Pantanal Municipalities. Figure 2(b) represents the Pantanal head of cattle supply growth

Figure 2 represents actual data adopted as initial assumptions for the SD model, to illustrate the BAU scenario. Figure 2(a) depicts the area of pasture in Pantanal municipalities, while Figure 2(b) gives rise to the supply growth of head of cattle in Pantanal, which is assumed to be linear for this study, from 1990 to 2019.

A reduction proxy of 38.5 % in the use of area (ha/head of cattle) was adopted to illustrate a pasture intensification scenario, based on an experience with cattle in the Mato Grosso state. The productivity gain when applying intensive cattle in comparison to the traditional one gives rise to a great number of recovered pastures (Fabiano Alvim Barbosa, 2015).

A report by EMBRAPA (2014) showed that the intensification of cattle, despite increasing methane emissions, increases carbon sequestration in the recovered pastures, resulting in a positive net balance in the total emissions. According to their study, the average annual net emission of enteric methane (CH_4) per animal is 57 CH_4 kg per year per animal and that, through better cattle management practices, such as ranching and feeding conditions, this number can reach 37.7 kg, representing a 33.9 % emission reduction of this gas, which is one of the main global greenhouse gases.

To simulate the impact of cattle intensification policies under emission and land use behaviours, two hypothetical public policies (PP1 and PP2) were proposed. With supply growth and pressure for more sustainable cattle, these policies aim to increase productivity, reconciling rural development, supply growth and environmental conservation, represented here by CH_4 emissions and land use from cattle. PP1 assumes that all supply growth shall be attended by new ranching practices with higher productivity (heads/ha). PP2 considers that, after its implementation, a percentage of the existing cattle will have their ranching regime changed, moving from extensive to semi-intensive cattle practices within a fixed horizon (5 years).

The SD model was tested to evaluate its accuracy in interpreting the actual scenario (BAU – Business as Usual model). The policies were assumed to be adopted in 2000 (year 10) and with a linear ramp-up of 5 years for full implementation of policy PP2 in year 2005 (year 15).

5. Results

By employing the System Dynamics modelling, the study analyses the results of the proposed policies and identifies the relationship between cattle intensive practices, land use distribution and CH4 emissions. For that, three scenarios were considered. Scenario 1 considers the Business as usual (BAU), which replicates what happened with the

number of cattle heads and land use between 1990 and 2019 in Pantanal; Scenario 2 contemplates the application of PP1 and PP2 policies disregarding any supply growth. This is a theoretical scenario aimed at controlling and isolating the effect of the policies and illustrating how supply growth plays an important role in land use and global emissions, when compared to the Scenario 3, which applies, in addition to the policies, an annual supply growth.

The application of policies PP1 and PP2 reflects how ranching time optimization and better management practices impacts on cattle productivity and CH_4 emissions. In comparison to the BAU scenario, it is possible to identify an improvement in the land use area, as showed in Figures 3(a) and 3(b). By comparing Scenario 3 with Scenario 1 (BAU), the natural vegetation, which in 2019 totalled 15.1 Mha, would reach 16.2 Mha (an increase of 6.8 %). Supply growth is responsible for more than 10 % of natural vegetation consumption with pasture, by comparing Scenario 2 and 3. The pastureland would experience a significant reduction of almost 28 % in Scenario 3.

Regarding the number of heads, considering the policies application in Scenario 3, in year 30, intensive cattle practices would represent more than 35 % of the region's cattle supply. The total emissions per year, in tons of CH_4, by year 30, went from 244,100 tons of CH_4 in the BAU scenario to 203,000, by applying policies PP1 and PP2 (Scenario 3), representing an important reduction of 20.4 %. When analyzing the number of tons of CH_4 emitted per ton of meat, considering cattle with average weight of 400 kg, there is a significant reduction of 26.4% (from 2.28 ton of CH_4 per ton of meat to 1.67 ton of CH_4 per ton of meat). Even with supply growth (Scenario 3), this number remains descending in the analyzed horizon, as shown in figure 3(d).

From the perspective of socioeconomic impact, there was an improvement in HDI index. However, it is not possible to correlate this improvement with livestock practices in the region, since it proportionally follows the growth of the country's HDI, resulting from federal socio-economic measures implemented in the last 30 years. In addition, the income, with the new practices would increase by 50%, according to specialists, since intensive techniques would require greater professional qualification of ranching employees, contributing to technical development and quality of life improvement for the region.

6. Conclusion

This study aimed to analyze the impacts of productivity increase in cattle farming on the Brazilian Pantanal region on CH4 emissions and land use. It was considered the application of public policies (PP1 and PP2) that only admitted intensification practices for new cattle and proposed a change in a percentual of existing cattle, from extensive to intensive regime, by year 10. The proposed SD model compared the actual scenario, based on actual data (Scenario 1), with two theoretical scenarios (Scenario 2 and 3), that contemplated these new policies, from 1990 to 2019. The employment of a System dynamics model connects and gives rise to the system’s behavior, enabling better policy decisions and design.

It was found that the application of public policies that are not so restrictive can effectively reduce the number of emissions and offer more efficient livestock practices in the Pantanal. Even considering supply growth, the total emissions per year went from 244,100 ton of CH4 in the actual scenario to 203.000 ton of CH4 in Scenario 3, in year 30 (a reduction of more than 20 %).

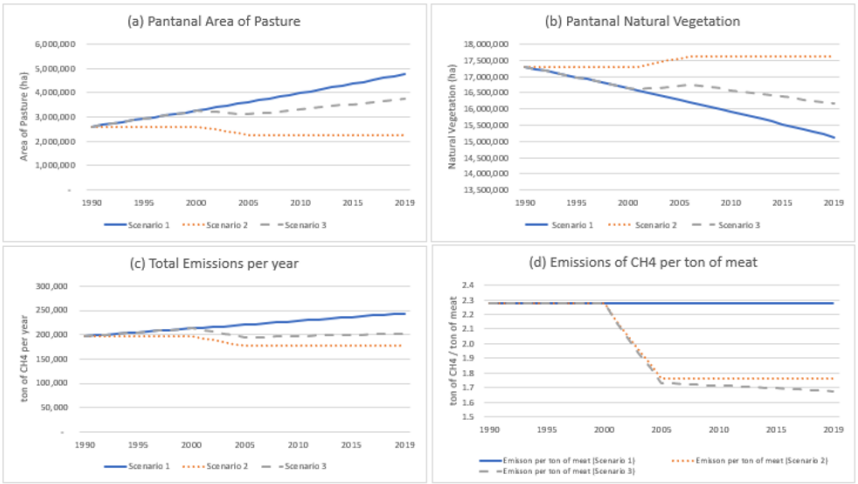


Figure 3 - Results from scenarios analysis for Pantanal cattle case. Figure 3(a): land used for pasture; Figure 3(b): natural vegetation available in Pantanal; Figure 3(c): total CH4 emissions per year, in tons of CH4; Figure 3(d): total emissions of CH4 per ton of meat of cattle

A reduction of 26.4 % in tons of CH4 per ton of meat could also be observed, (from 2.28 ton of CH4 per ton of meat to 1.67 ton of CH4 per ton of meat). The natural vegetation could experience an important achievement, by applying these policies, reaching an increase of almost 7 % (from 15.1 Mha to 16.2 Mha) in comparison to the actual scenario.

From a socioeconomic point of view, these policy changes would lead to an increase in the region's income and a need for more professional qualification. An in-depth study is recommended to study how improvements in livestock production practices could lead to an improvement in socioeconomic conditions in the region.

These important results ensure that improvements in land use and in CH_4 emissions can be made with easy-to-implement measures, helping Brazilian institution to meet the objective of COP-26.

7. References

Avery Cohn, A. M. (2014). Cattle ranching intensification in Brazil can reduce global greenhouse gas emissions by sparing land from deforestation. Proceedings of the National Academy of Sciences, 25.

EMBRAPA. (2014). Revista XXI - Ciência para a vida. 27-29, 14.

Fabiano Alvim Barbosa, B. S. (2015). Cenários para a Pecuária de Corte Amazônica. Belo Horizonte.

Fiddaman, T. S. (2012). Exploring policy options with a behavioral climate–economy model. System Dynamics Review, 18(2), 243-267.

IBGE. (2020). PPM - Pesquisa da Pecuária Municipal.

MapBiomas. (2020). MapBiomas Project - Collection 5.0 of Brazilian Land Cover & Use Map Series. Acesso em 26 de 06 de 2021.

P. J. Gerber, H. S. (2013). Tackling climate change through livestock: A global assessment of emissions and mitigation opportunities. FAO.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50309-2

Promoting phosphorus recovery at livestock facilities in the Great Lakes region: Analysis of incentive policies

Edgar Martín Hernández[a,b*], Yicheng Hu[c], Victor M. Zavala[c], Mariano Martín[a], Gerardo J. Ruiz-Mercado[d,e]

[a] *Department of Chemical Engineering, University of Salamanca, Plza. Caídos 1-5, 37008 Salamanca, Spain*
[b] *Oak Ridge Institute for Science and Education, hosted by Office of Research & Development, US Environmental Protection Agency, 26 West Martin Luther King Drive, Cincinnati, Ohio 45268, Unied States*
[c] *Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, Wisconsin 53706, Unied States*
[d] *Center for Environmental Solutions and Emergency Response (CESER), U.S. Environmental Protection Agency, 26 West Martin Luther King Drive, Cincinnati, Ohio 45268, Unied States*
[e] *Chemical Engineering Graduate Program, University of Atlántico, Puerto Colombia 080007, Colombia*

emartinher@usal.es

Abstract

Intensive farming activities release large amounts of phosphorus into the environment in the form livestock manure, contributing to the eutrophication of waterbodies, and can lead to algal bloom episodes. This work conducts a study on the design and analysis of incentive policies to promote the implementation of phosphorus recovery systems at intensive livestock facilities minimizing their negative impact on the economy of livestock operations. The Great Lakes area is used as case study, analyzing the economic impact of the implementation of phosphorus recovery systems, either considering the deployment of standalone phosphorus recovery processes, or integrated systems combining nutrient recovery with anaerobic digestion for the production of electricity. Moreover, the fair allocation of monetary resources when the available budget is limited has been studied using the Nash allocation scheme.

Keywords: Environmental Policy, Circular Economy, Resource Recovery, Nutrient Pollution, Organic Waste

1. Introduction

Since the 19th century, the agricultural sector has experienced an accelerated industrialization pursuing its intensification to satisfy the demand of food and agricultural products, i.e., increasing the agricultural production per unit of input resources (land, labor, time, fertilizer, seeds, and investment) (FAO, 2004). However, multiple environmental challenges must be faced as a consequence of the industrialization of the agricultural and farming activities. One of the main sources of concern are the nutrient releases from intensive

livestock facilities in the form of manure, which contribute to the nutrient pollution of waterbodies, and contribute to eutrophication and harmful algal blooms (HABs) episodes. Therefore, nutrient recovery and recycling is not only a desirable but also a necessary approach to develop a more sustainable agricultural paradigm.

In this work, the effect of different incentive policies to promote the implementation of P recovery systems at concentrated animal feeding operations (CAFOs) is evaluated. Since P recovery systems can be implemented either as standalone systems, or integrated with biogas production and upgrading processes, the combination of incentives for the recovery of both phosphorus and electricity has also been considered. In addition, we study the allocation of limited monetary resources using a Nash scheme; this determines the break-even point for the allocation of monetary resources based on the availability of incentives.

2. Framework for the assessment of incentive policies

A two-stage framework is proposed for the evaluation of incentive policies, as shown in Figure 1. In the first stage, the size and geographical location of the studied CAFOs are analyzed, selecting the most suitable P recovery process for each CAFO assessed from a pool of six P recovery technologies. We note that these technologies can be implemented either standalone, or integrated with anaerobic digestion (AD) for biogas production. The P recovery selection stage is composed of different models that are fed with data regarding the type and number of animals in the studied CAFO, as well as its geographical location (box *a*). The assessment of the regional environmental vulnerability to nutrient pollution is performed through a geographical information system (GIS) model (box *b*). Additionally, the techno-economic assessment of the different phosphorus recovery technologies, and biogas production in those cases where this process is considered, is performed based on the characteristics of each CAFO under evaluation in parallel (box *c*). The information returned by these models is normalized and aggregated in a multi-criteria decision analysis (MCDA) model to select the most suitable nutrient management technology for the evaluated livestock facility (box *d*). A detailed description of decission-support framework used for the assessment and selection of P recovery systems considering the environmental vulnerability to nutrient pollution can be found in Martín-Hernández et al. (2021).

In a second stage, the effect of incentives on the economic performance of the P recovery systems selected in the first stage is evaluated. This study is performed through an economic model that estimates the profit of the P recovery systems implemented and the total cost of phosphorus recovery. Additionally, a cost-benefit analysis comparing the recovery cost and the economic looses due to nutrient pollution is performed (box *e*).

The incentives considered in this study to promote the implementation of nutrient recovery systems are phosphorous credits (P credits) and renewable electricity certificates (REC). P credits can be articulated as a system for the acquisition of phosphorus emission allowances, or conversely, as an income obtained by recovering phosphorus, which is the P credits definition considered in this work. In addition, in those scenarios where biogas production is integrated, REC are also considered. REC incentives provide a fixed remuneration for the electricity produced, which can result in a higher transaction price of electricity to cover the extra production costs and guarantee long-term price stability.

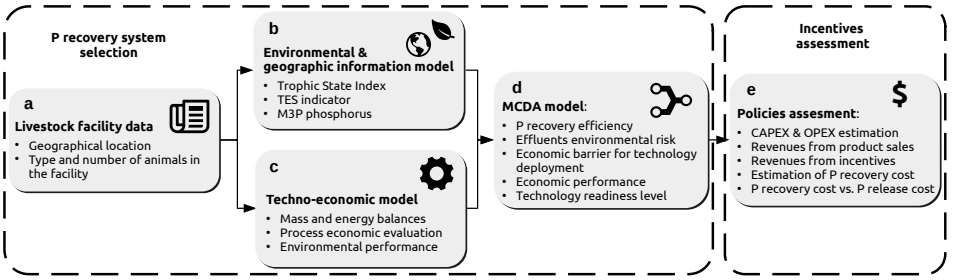


Figure 1: Flowchart of the models for selection, sizing, and evaluation of nutrient recovery systems at livestock facilities.

The states of the Great Lakes area, i.e., Minnesota, Indiana, Ohio, Pennsylvania, Wisconsin, and Michigan, are the study region considered to analyzed the impact of incentive policies on P recovery. The CAFOs considered for the deployment of livestock waste treatment processes are those livestock facilities with more than 300 animal units reported in the National Pollutant Discharge Elimination System (NPDES) by the U.S. Environmental Protection Agency (US EPA) in the states under evaluation. An animal unit is defined as an animal equivalent of 1,000 pounds live weight. 2,217 CAFOs are considered in total.

3. Results and discussion

The results of the implementation and allocation of incentives for phosphorus recovery are shown in this section. The results regarding the techno-economic assessment of the different processes, technology selection, and phosphorus recovered can be found in Martín-Hernández et al. (2021).

3.1. Combined effect of incentives for phosphorus and renewable electricity recovery

The net processing costs obtained for different scenarios combining multiple values of P credits and REC incentives are shown in Figure 2. They show a base cost for the recovery of phosphorus between 5.81 and 12.47 USD per ton of processed manure if no incentives or anaerobic digestion stages are considered. The installation of biogas processes is not profitable by itself, increasing the processing costs by 1.2-1.9 times over the base case, and it is only beneficial for large size CAFOs unser specific scenarios combining moderate P credits (>3 USD/kgP recovered) and electricity incentives (>60 USD/MWh). The scenarios combining states with large CAFOs and high value for P credits, and the optional production of renewable energy from biogas result in negative processing costs, i.e., they are profitable. Since the analysis of the different scenarios is carried out at the state level, this means that the profitable P recovery processes are able to balance out the non-profitable ones in the state.

3.2. Environmental cost-benefit analysis

The cost-effectiveness of the total cost involved in phosphorus recovery, including the amortization of the investment, operating costs, and total cost of incentives for each sce-

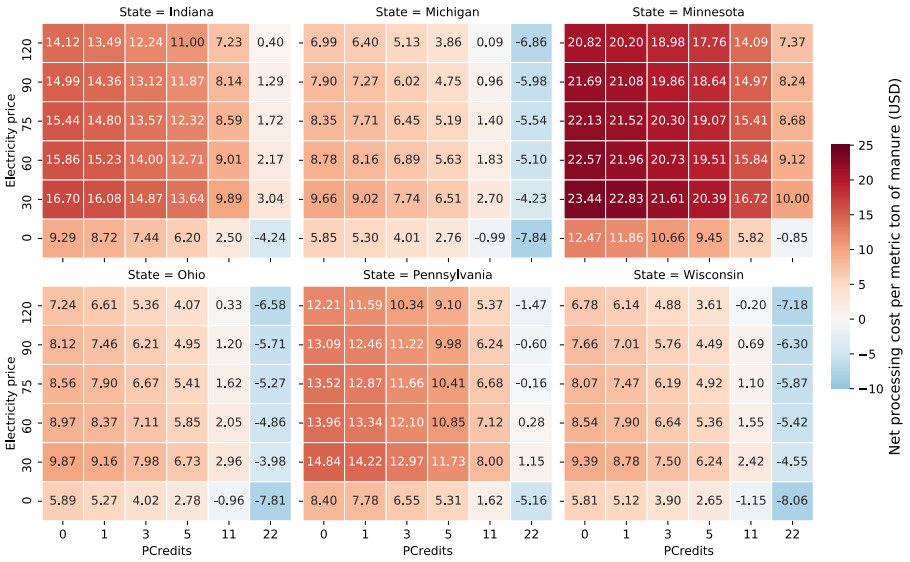


Figure 2: Net processing cost per metric ton of manure (USD).

nario under evaluation has been studied to determine the long-term economic benefits of phosphorus recovery.

Figure 3 shows the total cost of phosphorus recovery under different policies, including all the items previosuly described. The economic losses due to phosphorus releases have been estimated in 74.5 USD per kg of phosphorus released by Sampat et al. (2021). It can be observed that P recovery is economically beneficial in all scenarios considered, even those resulting in the largest P recovery costs as a result of high incentive values. The role of the size of CAFOs in the cost of phosphorus recovery can be also observed in this study. Those states with larger average size of CAFOs, such as Wisconsin, Ohio, and Michigan, have recovery costs significantly lower than the states where medium and small size CAFOs are predominant.

3.3. Fair distribution of incentives

In those scenarios where the available budget is not sufficient to cover the operating expenses of the unprofitable P recovery processes, the fair distribution of incentives is a relevant problem. For the CAFOs in the study region considered in this work, the necessary budget to cover the economic losses of the unprofitable P recovery systems is 222.6 MM USD. We note that Due to the marginal benefits obtained by installing AD processes, as described in Section 3.1., the implementation of only P recovery systems is assumed in both studies, and therefore only incentives for P recovery are considered.

The fair allocation of limited incentives has been addressed by using the Nash allocation scheme. This approach has been selected because this scheme is able to capture the scales of the different stakeholders (CAFO facilities) in order to achieve a fair distribution of a certain resource (incentives), as it was demonstrated by Sampat and Zavala (2019).

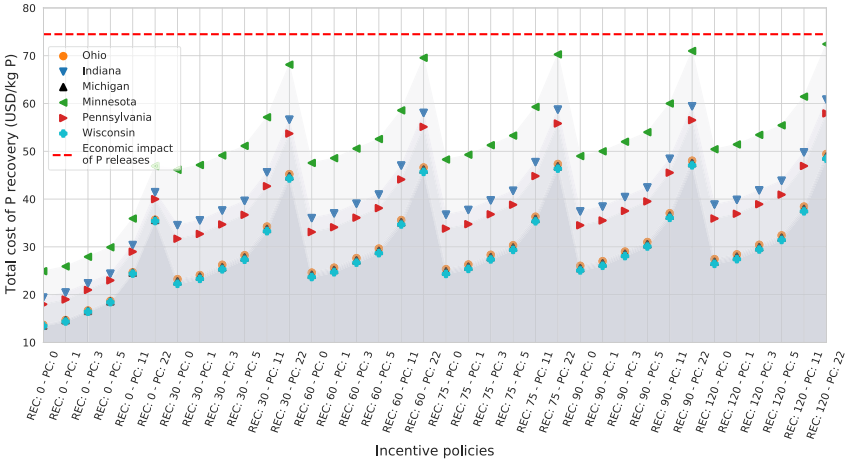


Figure 3: Comparison of the total cost of phosphorus recovery for each scenario assessed and the environmental remediation cost due to phosphorus releases. *REC* denotes the electricity incentive values considered in USD/MWh, and *PC* denotes the value of phosphorus credits in USD/$kg_{P\ recovered}$. The red dotted line represents the economic losses due to phosphorus releases to the environment.

Figure 4 illustrates the distribution of incentives as a function of the net revenues of the P recovery system installed in each CAFO c before any incentive is applied. The cases where the available budget are the 10%, 30%, 50%, 70%, and 100% of the incentives needed to cover the economic losses of unprofitable CAFOs are analyzed (22.3, 66.8, 111.3, 155.8 and 222.6 MM USD respectively). Since the available incentives are limited, a break-even point determining the profitability level of the P recovery systems below which the incentives should be allocated is set for each scenario. As a result, the fewer incentives available, the more restrictive the break-even point is. Additionally, it can be observed that the displacement of the break-even points is progressively reduced as the available incentives increase, resulting in a marginal improvement between the scenarios considering the 50% and 70% of the economic resources needed to guarantee the economic neutrality of the nutrient management systems.

4. Conclusions

This work aims at analyzing incentive policies for the implementation of phosphorus recovery systems for the abatement of nutrient releases from CAFOs. The deployment of phosphorus recovery processes is self-profitable through struvite sales only for the largest P recovery processes, which represent less than the 5% over the total CAFOs in all the studied states. However, the application of P credits increases the fraction of profitable processes around to 100% in the states with large-size CAFOs (Michigan, Ohio and Wisconsin), and up to 80% for the states with medium-size CAFOs (Indiana and Pennsylvania). The incentives necessary for covering the economic losses of unprofitable CAFOs estimated in 222.6 million USD. The integration of phosphorus recovery technologies with

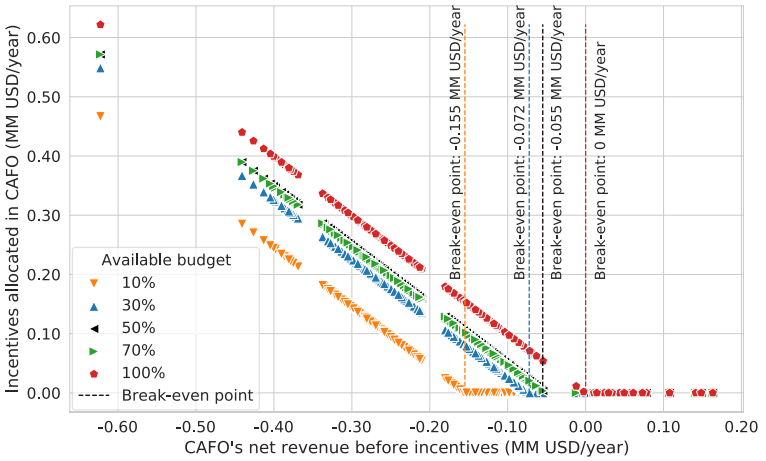


Figure 4: Distribution of incentives considering the Nash allocation scheme. Scenarios assuming available incentives equal to the 10%, 30%, 50%, 70%, and 100% of the incentives needed to cover the economic losses of unprofitable P recovery systems in the Great Lakes area are illustrated.

anaerobic digestion and biogas upgrading processes does not result in any practical improvement in terms of economic performance. The total cost of phosphorus recovery, including the investment amortization, operating costs, and total cost of incentives is lower than the long-term economic losses due to phosphorus pollution for all the evaluated states and policies, proving that sustainable nutrient management systems are economically and environmentally beneficial. Additionally, the fair distribution of limited incentives has been studied, determining the break-even point for the allocation of monetary resources based on the availability of incentives.

Disclaimer: The views expressed in this article are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency. Mention of trade names, products, or services does not convey, and should not be interpreted as conveying, official U.S. EPA approval, endorsement, or recommendation.

References

FAO, 2004, The ethics of sustainable agricultural intensification, FAO.

E. Martín-Hernández, M. Martín, G.J. Ruiz-Mercado, 2021, A geospatial environmental and techno-economic framework for sustainable phosphorus management at livestock facilities, Resources, Conservation and Recycling, 175, 105843.

A.M. Sampat, A. Hicks, G.J. Ruiz-Mercado, V.M. Zavala, 2021, Valuing economic impact reductions of nutrient pollution from livestock waste, Resources, Conservation and Recycling, 164, 105199.

A.M. Sampat, V.M. Zavala, 2019, Fairness measures for decision-making and conflict resolution, Optimization and Engineering 20, 1249–1272.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50310-9

Production of ethanol, xylitol and antioxidants in a biorefinery from olive tree wastes: process economics, carbon footprint and water consumption

Luis David Servian-Rivas[a], Ismael Díaz[a], Elia Ruiz Pachón[b], Manuel Rodríguez[a], María González-Miquel[a], Emilio J. González[a]
[a]*Departamento de Ingeniería Química Industrial y del Medioambiente, Escuela Técnica Superior de Ingenieros Industriales, Universidad Politécnica de Madrid, C/ José Gutierrez Abascal 2, 28006 Madrid, Spain*

[b]*Bioenergy and Energy Planning Research Group, EPFL, Switzerland*
luisdavid.servian.rivas@upm.es

Abstract

This work focuses on the evaluation of the economic performance and carbon and water-related environmental impacts of a biorefinery scheme using olive tree pruning wastes as feedstock. The considered process is based on a multiproduct biorefinery producing 0.66 m^3/h of ethanol, 114 kg/h of xylitol and 144.4 kg/h of antioxidants valorising both the cellulose/hemicellulose and the extractives fraction. The plant is not energetically self-sufficient (even considering the combined heat and power production from the combustion of the waste solids fraction) requiring a supply of natural gas. Nonetheless, the plant shows a positive investment balance with a net present value of 11.56 M€ in a 20-year period being the most important product the antioxidant which represent 66.2 % of total revenues. In addition, the biorefinery also shows a better environmental profile in comparison to the business-as-usual production of ethanol, xylitol and antioxidant.

Keywords: Lignocellulosic biorefinery, Technoeconomic Analysis, Life Cycle Assessment.

1. Introduction

Europe established in the European Green Deal an impulse on circular bioeconomy, replacing fossil-based materials and energy by biobased solutions (European Commission, 2019), in which biowaste is the principal feedstock. This strategy helps to mitigate the problem of the generation of high-volume residues, and at the same time products of great interest are obtained. Furthermore, several biorefinery projects in Europe using biowastes as feedstock have already shown a substantial reduction of energy and greenhouse gas emissions (GHGs) with respect to the reference fossil-based processes (IEA, 2021).
Mediterranean countries in Europe could take advantage of incorporating this circular bioeconomy, because one of the most promising feedstocks for the obtention of a wide range of chemical products is olive tree waste (Lo Giudice et al., 2021). Those countries produce 70 % of the total world olive oil production (21 million t), being Spain the main world producer (1.8 million t) (FAOSTAT, 2021). Furthermore, the olive oil industry (olive cultivation and olive oil production) produces 10 million tonnes of residues per year, with a great potential to be valorised by means of different biorefinery processes.

One of the residues is olive tree pruning (OTP), which is an abundant lignocellulosic residue (1-2 t/ha each year) from olive cultivation. This residue must be removed from the field to prevent the propagation of vegetal pests (Contreras et al., 2020), but it is usually burned in the field, being this burning the main source of fine aerosol in winter in Mediterranean countries (Kostenidou et al., 2013).
OTP has been used to obtain marketable products such as ethanol and xylitol and it is also a good source of natural antioxidants such as hydroxytyrosol, tyrosol, and oleuropein (Conde et al., 2009a). There are previous works that have studied the economic feasibility of plants that manufacture those products (Susmozas et al., 2019); however, an assessment of the associated environmental impacts (carbon and water) has not been carried out so far.
This work focuses on the economic performance and the carbon and water-related impact assessment of a biorefinery plant using OTP as feedstock. The biorefinery was modelled using commercial process simulation software (Aspen Plus). The subsequent economic and environmental analysis carried out were rooted on data from the process simulations. Economic performance is studied using the net present value as economic metric. The carbon footprint and water consumption were accounted using a life cycle assessment (LCA) methodology to compare the impacts of the biorefineries with respect to the business-as-usual solution.

2. Materials and Methods.

2.1. Feedstock
The OTP composition %wt (dry basis)is the same described by Ballesteros et al.(2011): 28.0 cellulose (as cellulose); 20.6 hemicellulose (as xylan); 25.2 lignin; 2.7 acetic groups; 5.9 Ash (as CaO); 7.9 glucose, 0.1 arabinose, 0.1 mannose, 1 galactose, 0.7 xylose, 4 mannitol, and 3.8 antioxidant. (Ballesteros et al.(2011). The OTP is assumed to be already crushed (this will affect the price).

2.2. Process modelling.
Materials and energy balances of the considered scheme were computed using Aspen Plus v.11. Two thermodynamic packages (NRTL and UNIF-LL) were used in this simulation. The semi-empirical NRTL model was used as the default method in the simulation, with the only exception the liquid-liquid extraction of antioxidants. For this purpose, the UNIF-LL package was used to estimate the activity coefficients required in calculations of fluid phase equilibria of the mixtures involving antioxidants, ethyl acetate, sugars, and water. Missing pure component parameters were estimated using the built-in property estimation models (group contribution models from molecular structure). The antioxidant is simulated as hydroxytyrosol because it represent the 90 % of total compounds in this fraction (Conde et al., 2009b). The process is design to work with 96 t/day of OTP, which represent the 0,7 % of total available OTP residue at year. Process parameters such as stoichiometric conversion and operating conditions for pre-treatment, saccharification, and fermentation were retrieved from literature.

2.3. Process description.
In figure 1 can be observed that OTP is firstly subjected to a hot water extraction process with water at 393.15 K and 5 bar, to extract 90 % of the extractives. A mixture of solids and liquids is obtained in this unit operation: a liquid fraction with the extractives and a solid fraction (SF) with the insoluble solids. The liquid fraction is quenched to 308.15 K and pumped to a liquid-liquid extraction process using ethyl acetate solvent in a 3:1 (v/v)

liquid-solvent ratio. Ethyl acetate selectively extracts 99 % antioxidants from the feed to the organic liquid, while sugars remain in the aqueous fraction.
On the other hand, a significant amount of ethyl acetate dissolves in the aqueous stream so it must be separated and sent back to the process. This is required to improve the economy of the process (recovering as much solvent as possible) and to reduce the organic fraction sent to the wastewater treatment plant.
The organic liquid with the antioxidant dissolved is sent to a flash vessel at 373.15 K and 0.1 bar, where the solvent is recovered by evaporation (light component) and the extracted antioxidants are collected from the bottom with a purity of 97.33 wt %. A total recovery of 90 % of the incoming antioxidants from OTP is obtained.
The SF is sent to a steam explosion process with steam at 468.99 K and 14 bar and phosphoric acid (1 %) as catalyst. This operation breaks lignin and eases the subsequent saccharification and fermentation of cellulose. Two fractions are obtained after filtration: water-insoluble solids (WIS) and water-soluble solids (WSS).
The WSS fraction is composed of xylose as the main organic constituent and is sent to the xylitol production route. The WSS is cold down at 323 K and mixed with lime to remove furans and phenolic compounds, followed by neutralization with H_2SO_4. The resulting solids are filtered and disposed. The filtered liquid is fermented at 303 K, obtaining a yield of 75 % of xylose to xylitol, then subjected a filtration to remove yeast, and evaporation process at 313.15 k and 0.5 bar to concentrate xylitol up to 50 wt%. Finally, xylitol is mixed with ethanol to decrease xylitol solubility at a 0.13:1 (w/w) ethanol-dissolution ratio. Finally, a crystallization process takes place at 268.15 K and 1 bar to obtain xylitol crystals with 99 % purity.
The WIS mainly contains insoluble cellulose and lignin. This fraction is converted to ethanol by a conventional sequence of pre-saccharification, saccharification, and fermentation steps (PSSF). Pre-saccharification is carried out at 323 K. Then the resulting stream is further cooled to 308 K and fermented to ethanol (70% of the theoretical yield is obtained). The resulting ethanol stream is firstly purified in a beer column which increases ethanol concentration up to 50 %wt and further concentrated in a purification column which increases ethanol concentration to 93 %. Finally, 99% ethanol is obtained using molecular sieves.
The aqueous wastes from the process are sent to the wastewater treatment system (anaerobic and aerobic processes) where methane, sludge, and biogas are produced and then burned in a combined heat and power section to partially compensate the plant energy requirements.

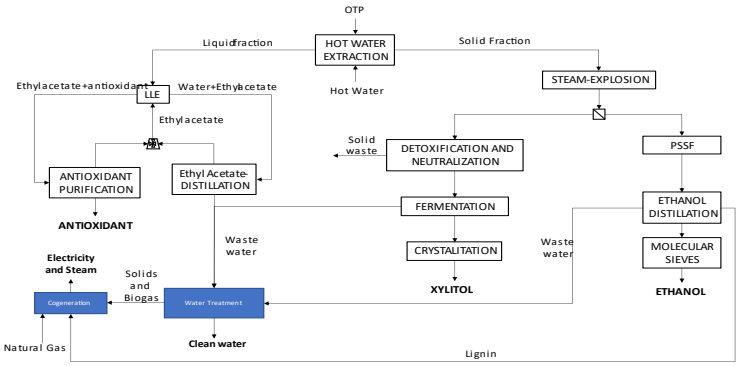


Figure 1. Biorefinery process flow diagram

2.4. Economic evaluation

Net present value (NPV) has been used as the financial metric to measure the profitability of the plant with a lifetime of 20 years, an interest rate of 15 % and 25 % of taxes. The Inside battery limit (ISBL) plant cost is obtained by means of Aspen Economic Analyzer (Aspen Technologies Inc., USA). The cost of all equipment has been increased by 30% to cover possible uncertainties.

A sensitivity analysis is also studied, varying the prices of the OTP and the three products until the NPV is equal to zero. The base prices of the products are: 0.58 €/L ethanol, 3170 €/t xylitol, 11000 €/t antioxidants and the OTP 44.77 €/t (Susmozas et al., 2019)

2.5. Carbon footprint and water consumption

The carbon and water footprints analysis were calculated in line with the Life Cycle Assessment (LCA) principles, which is an internationally standardized methodology that helps to quantify the environmental impact of processes and products.

First, the main goal is to quantify the environmental impacts of producing ethanol, xylitol, and antioxidants from OTP, compared to the conventional counterparts of biorefinery products (reference system). The reference system is made up of ethanol from ethylene, xylitol from corncob, and propyl gallate from spruce bark was considered as the counterpart reference for antioxidant production. The overall functional unit of the whole system is 1 kg of OTP.

As the studied biorefinery is a multiproduct system, an economic allocation approach has been adopted following the value-based methodology proposed by Gnansounou et al.(2015).

Data to build the OTP biorefinery inventory are obtained from the process simulations carried out in this work, this information is quantified and transformed into environmental impacts categories ReCiPe 2016 impact assessment methodology. SimaPro v11 software was used to assist in the calculation of the four environmental impacts selected: global warming (kg CO_2) and water consumption (m^3), The database Ecoinvent 3.5 is used.

3. Results and Discussion

3.1. Process design and simulation.

The modelled biorefinery produces 0.66 m^3/h of ethanol, 114 kg/h of xylitol and 144 kg/h of antioxidants with a raw material consumption of 4040 kg/h. The thermal need of the plant is 61.2 GJ/h, but the plant just produces 23.9 GJ/h from the combustion of the solid organic residues and biogas of the plant, so the plant needs an external supply of energy by means of natural gas. This plant is thermally integrated, and the consumption of heat is already optimized using the Aspen Energy Analyzer tool.

The main areas of thermal energy consumption are steam explosion pretreatment and antioxidant extraction (63 % of thermal consumption). Besides, these processes consume 73 % of total process water in the entire plant.

3.2. Economic evaluation.

The capital expenditure of the plant is 39.5 M€ and the operating expenses are 7.14 M€/y, being OTP the most important operating cost (23 %). Revenues are accounted for 20.3 M€/y, being the most important product the antioxidant that accounts for 66.2 % of the total revenues. The NPV is 11.53 M€, with a pay-back time of 4 years. Meaning that the plant is profitable.

In the sensitivity analysis, the price of OTP could increase by 224.45 % with respect its original price, even being the most important operating cost, the plant could resist price variations keeping positive profits. Furthermore, the price of ethanol could be reduced by

a maximum of 98 %, and xylitol could be sold for free, and the plant would reach a NPV of 1.5 M€. On the other hand, the antioxidant price just could be reduced by 27 %.
This means that the plant profits are supported by the antioxidant market, making this part of the production scheme the most important in terms of process economics. So future optimization of the plant should focus on the improvement of the antioxidant production or even on the consideration of a standalone antioxidant production plant.

3.3. Carbon footprint and water consumption

The indirect and direct emissions in terms of kg CO_2 equivalent (carbon footprint) and the total water consumptions were calculated for both the considered biorefinery and the reference system (Table 2). It can be seen a clear reduction in the carbon footprint of the biorefinery products with respect to the individual reference counterparts. The global warming impact is reduced from 8.17 to 0.42 kg CO_2 equivalent. However, in terms of water consumption the advantages of the biorefinery are not so clear. On the one hand, there is a global reduction of water use which is a consequence of the higher water consumption of the reference antioxidant production process (propyl gallate). On the other hand, both ethanol and xylitol produced in the biorefinery have a higher water footprint than the individual reference systems.

Table 1. LCIA results of the OTP biorefinery and the reference system for 1 kg of OTP.

Biorefinery		**Reference System**	
Global warming (kg CO_{2eq})			
Total	0.42	8.17	Total
Antioxidant	0.16	7.62	Propyl Gallate
Ethanol	0.14	0.16	Ethanol (fossil)
Xylitol	0.12	0.38	Xylitol (corncob)
Water consumption (m^3)			
Total	0.12	0.16	Total
Antioxidant	0.071	0.150	Propyl Gallate
Ethanol	0.026	0.006	Ethanol (fossil)
Xylitol	0.019	0.001	Xylitol (corncob)

As stated above, antioxidants are the most economically appealing product in the considered multiproduct biorefinery scheme. However, in terms of environmental aspects they are also responsible for the highest impacts of the plant in comparison to ethanol and xylitol. This is mainly due to the high water and energy consumptions required in the antioxidant hot water extraction process (Section 3.1). At the same time, the antioxidants produced from OTP are the only product of the plant which reduces both the carbon and water footprints with respect to its reference (propyl gallate).

4. Conclusions

From the results shown, it can be clearly seen that the use of OTP as feedstock for the production of ethanol, xylitol and antioxidants is both economically profitable and less environmentally harmful in terms of carbon footprint and water consumption than the business-as-usual solution. At the same time, antioxidants are the most economically interesting product of the plant (66.2 % of revenues). From an environmental point of view, they also reduce drastically all the impacts with respect to the reference propyl gallate production.

References

Ballesteros, I., Ballesteros, M., Cara, C., Sáez, F., Castro, E., Manzanares, P., Negro, M. J., & Oliva, J. M. (2011). Effect of water extraction on sugars recovery from steam exploded olive tree pruning. *Bioresource Technology*, *102*(11), 6611–6616. https://doi.org/10.1016/j.biortech.2011.03.077

Comission, E. (2019). *A European Green Deal | European Commission*. European Commission.

Conde, E., Cara, C., Moure, A., Ruiz, E., Castro, E., & Domínguez, H. (2009a). Antioxidant activity of the phenolic compounds released by hydrothermal treatments of olive tree pruning. *Food Chemistry*, *114*(3), 806–812. https://doi.org/10.1016/j.foodchem.2008.10.017

Conde, E., Cara, C., Moure, A., Ruiz, E., Castro, E., & Domínguez, H. (2009b). Antioxidant activity of the phenolic compounds released by hydrothermal treatments of olive tree pruning. *Food Chemistry*, *114*(3), 806–812. https://doi.org/10.1016/j.foodchem.2008.10.017

Contreras, M. del M., Romero, I., Moya, M., & Castro, E. (2020). Olive-derived biomass as a renewable source of value-added products. *Process Biochemistry*, *97*(March), 43–56. https://doi.org/10.1016/j.procbio.2020.06.013

FAOSTAT. (2021). *Crops*. http://www.fao.org/faostat

Gnansounou, E., Vaskan, P., & Pachón, E. R. (2015). Comparative techno-economic assessment and LCA of selected integrated sugarcane-based biorefineries. *Bioresource Technology*, *196*, 364–375. https://doi.org/10.1016/j.biortech.2015.07.072

IEA. (2021). *Net Zero by 2050 A Roadmap for the Global Energy Sector*. 222.

Kostenidou, E., Kaltsonoudis, C., Tsiflikiotou, M., Louvaris, E., Russell, L. M., & Pandis, S. N. (2013). Burning of olive tree branches: A major organic aerosol source in the Mediterranean. *Atmospheric Chemistry and Physics*, *13*(17), 8797–8811. https://doi.org/10.5194/acp-13-8797-2013

Lo Giudice, V., Faraone, I., Bruno, M. R., Ponticelli, M., Labanca, F., Bisaccia, D., Massarelli, C., Milella, L., & Todaro, L. (2021). Olive trees by-products as sources of bioactive and other industrially useful compounds: A systematic review. *Molecules*, *26*(16). https://doi.org/10.3390/molecules26165081

Susmozas, A., Moreno, A. D., Romero-García, J. M., Manzanares, P., & Ballesteros, M. (2019). Designing an olive tree pruning biorefinery for the production of bioethanol, xylitol and antioxidants: A techno-economic assessment. *Holzforschung*, *73*(1), 15–23. https://doi.org/10.1515/hf-2018-0099

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50311-0

Application of CAPE Tools into Prospective Life Cycle Assessment: A Case Study in Acetylated Cellulose Nanofiber-Reinforced Plastics

Yasunori Kikuchi[a,b,c*] and Yuichiro Kanematsu[b]

[a] *Institute for Future Initiatives, The University of Tokyo, Tokyo 113-8654, Japan*
[b] *Presidential Endowed Chair for "Platinum Society", The University of Tokyo, Tokyo 113-8656, Japan*
[c] *Department of Chemical System Engineering, The University of Tokyo, Tokyo 113-8656, Japan*
ykikuchi@ifi.u-tokyo.ac.jp

Abstract

In this study, we are tackling systems design with assessments for emerging technologies. Computer-aided process engineering (CAPE) tools such as process design heuristics, process simulation, optimization, parametric analysis for characterizing sensitivity and alternative generation, and decision making with uncertainties have huge potential to compensate the data limitation of emerging technology and jump up to the deep technology assessments with quantified results. A case study on the application of CAPE tools into prospective life cycle assessment was conducted for the acetylated cellulose nanofiber-reinforced plastics, which has been developed to replace the conventional structural materials, e.g., steel or fossil-based plastics, in applications automobile or home appliances. We performed simulation-based life cycle inventory analysis to reveal the environmental and economic performance of CNF-reinforced plastics considering the future scale-up of production processes. Through this case study, it was demonstrated that the application of CAPE tools into prospective LCA enables the strategic technology assessments for systems design. Especially in the proof of concept on technology implementation can be verified and validated with the ranged values of uncertainties in emerging technology under development.

Keywords: LCA, CNF, greenhouse gas emission, production cost

1. Introduction

In order to achieve the decarbonization target by 2050 with defossilization, we must focus on the early introduction and diffusion of state-of-the-art elemental technologies. However, many promising elemental technologies are still under development, and even if they are expected to be commercialized, there is uncertainty about their decarbonizing effects in implementation. Therefore, for these promising elemental technologies, an early technology assessment on the economic and environmental aspects of the technology should be carried out before large-scale implementation, and a roadmap for the diffusion of the technology should be formulated, taking into account technological characteristics such as the maturity of technological development, technological change, and economies of scale, while limiting uncertainties. In recent years, there has been an increase in the number of case studies using prospective life cycle assessment (LCA), which take into account the future potential of the technology and aim to predict the

environmental impacts on the technology under development. (Arvidsson et al., 2018; Moni et al., 2020; Thonemann et al., 2020)

In this study, we are tackling systems design with assessments for emerging technologies. Because of the data limitation on the systems and processes adopting emerging technologies, their design and assessments have uncertainties and difficulties to implement them into society smoothly. Computer-aided process engineering (CAPE) tools such as process design heuristics, process simulation, optimization, parametric analysis for characterizing sensitivity and alternative generation, and decision making with uncertainties have huge potential to compensate such data limitation and jump up to the deep technology assessments with quantified results. In this paper, we examine the applicability of CAPE tools for systems design and assessment adopting emerging technologies with a case study in acetylated cellulose nanofiber-reinforced plastics (AcCNF-RP). AcCNF-RP has been developed to replace the conventional structural materials, e.g., steel or fossil-based plastics, in applications automobile or home appliances. Cellulose nanofibers (CNF) can be produced from plant-derived renewable resources and have advantage of mechanical properties in lightness and strength when it was applied as the filler of the composites. Examining the proof-of-concept, mitigating fossil use and greenhouse gas (GHG) emissions, is strongly needed before such emerging technologies spread to the market and society. In this study, we performed simulation-based life cycle inventory analysis to reveal the environmental and economic performance of AcCNF-RP considering the future scale-up of production processes.

2. Materials and methods

2.1. Application of CAPE tools into prospective LCA

Figure 1 shows the description of systems assessments applying CAPE tools for prospective LCA. In management activity and resource provider, data estimation and interpretations are assigned to CAPE tools considering the conditions in prospective LCA.

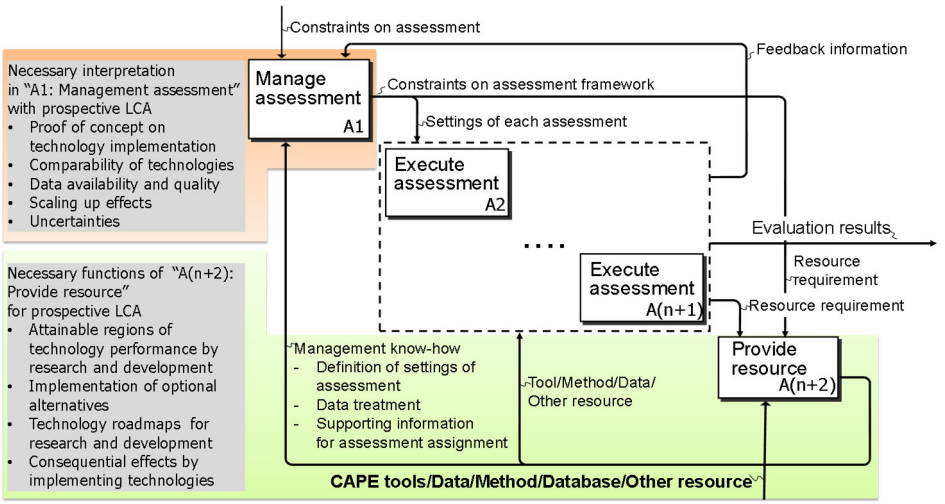


Figure 1 Description of multiple assessment activities with the necessary conditions for prospective LCA. n is the number of assessment methods. (Modified from previous studies (Kikuchi, 2014; Kikuchi et al., 2010; Kikuchi and Hirao, 2009))

Conventional LCA does not take into account changes in technology level, because it refers to information on the current technology level and specifically estimates the environmental impacts of each process related to the provision of products and services. On the other hand, efforts to tackle climate change have become more active in recent years, and new products and technologies are changing concepts and models more rapidly, making the transition to a low-emission society more urgent.

The significance of conducting a strategic LCA of emerging technologies for the 30-year time horizon up to the target year of 2050 arose regarding the issues on the climate change. Emerging technologies, as defined by Rotolo et al. (2015), are; "innovative and rapidly growing technologies that have the potential to have a significant social and economic impact in the domains in which they are structured, with some degree of persistent coherence, actors, institutions, ways of interacting with them and related knowledge production processes. It is characterised by its potential to have significant social and economic impacts. However, its most prominent impact lies in the future and is therefore somewhat uncertain and ambiguous at the stage at which the technology emerges." These technologies are characterized as "innovative", "rapid growth", "consistent", "significant impact" and "uncertain", which makes technology assessment difficult due to lack of existing data and knowledge.

Four main issues were identified as needing to be addressed in conducting prospective LCAs of emerging technologies (Thonemann et al., 2020; Moni et al., 2020). (1) comparability of technologies; (2) availability and quality of data; (3) scale-up challenges; and (4) uncertainty of assessment results. Process modeling and simulation are effective in estimating the missing process inventories in industrial scale production, because these technologies are under development in lab or pilot scale.

2.2. Case study: Acetylation of pulp for AcCNF-RP

Figure 2 shows the boundaries examined in this study. AcCNF-RP have been developed as substitutes for conventional structural materials (Eichhorn et al., 2010). Although kneading with polymers is required for pulp disintegration into nanofibers, it was excluded in this paper to focus on the chemical processes of acetylation applying CAPE tools (see also the previous paper (Kanematsu et al., 2021) on kneading process).

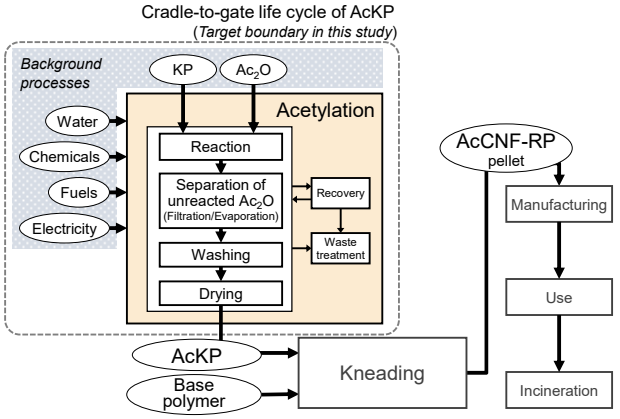


Figure 2 Related boundaries for AcCNF-RP life cycles in this study. KP: kraft pulp, AcKP: acetylated KP, AcCNF-RP: acetylated CNF-reinforced plastic, Ac_2O: acetic anhydride. (Modified from the previous literature (Kanematsu et al., 2021))

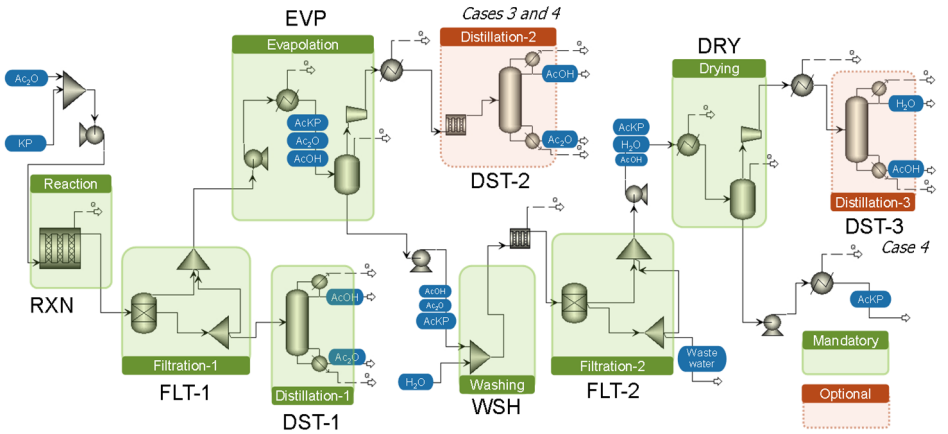


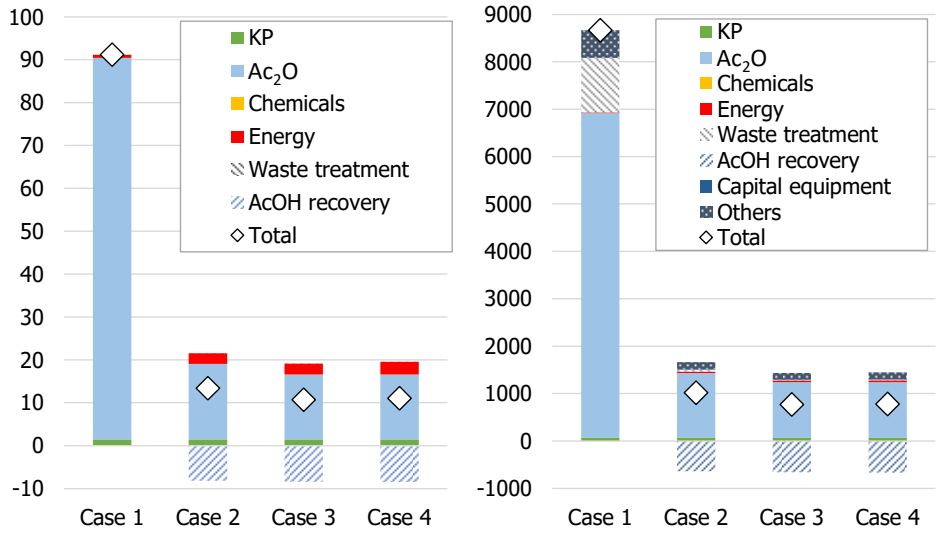
Figure 3 Process flow including all designed alternatives of the AcKP production process at an industrial scale. DST-1, -2, and -3 are optional processes, and the effects of their addition were compared through the process simulation. All reflux ratios in distillation column were set as the 1.3 times the theoretical minimum reflux ratio. This process consists of the process sections of reaction (RXN), filtration (FLT), evaporation (EVP), distillation (DST), washing (WSH), and drying (DRY). (Modified from the previous literature (Kanematsu et al., 2021))

As shown in Figure 2, Acetylation was defined as foreground processes, and kraft pulp (KP) obtained from a paper mill was used as pulp feedstock for CNF. The chemical modification process was examined in a scaled-up industrial process system implemented in the process simulator AspenPlus™ as well as in a lab-scale production (Kanematsu et al., 2021). The model constructed on process simulator is shown in Figure 3. The process simulation enables the evaluation of the process system in the actual production, which is not considered in the lab-scale production. For example, acetic anhydride (Ac_2O) can be recovered from the mixture of unreacted Ac_2O and acetic acid (AcOH) after the reactor and can be reused (DST-1 and -2). This is not normally done in the laboratory, but is a recycling process that is always considered in chemical plants. It is also possible to purify the byproduct, i.e., AcOH, to a level of purity that can be sold externally (DST-3). However, these unit operations are only optional and it is necessary to analyse the effect of introducing them.

LCA is carried out by combining the foreground data obtained by process simulation with the background data. In this study, the functional unit is the production of 1 kg of acetylated KP (AcKP), and the greenhouse gas emissions from the life cycle (LC-GHG) are calculated. The production scale was set at 100 tonnes/day of hydrous pulp as feedstock. Capital expenditure for equipment and other costs including staff costs and general administrative costs were calculated for each process alternative using the Aspen Process Economic Analyzer (APEA™) and costed in the same way as for LC-GHG.

3. Results and discussion

Figure 4 shows the assessment results on LC-GHG and production cost for unit amount of AcKP. Case 1 shows the results applying the inventories of lab-scale experiment. Cases 2 to 4 shows the results with CAPE tools. The dominant factor of the difference of case 1 against cases 2 to 4 was the recovery of Ac_2O, due to the relatively high cradle-to-gate LC-GHG of Ac_2O. In experiment, excess Ac_2O were consumed rather than raw materials.



(a) Cradle-to-gate LC-GHG [kg-CO_2eq/kg-AcKP] (b) Production cost [JPY/kg-AcKP]

Figure 4 Assessment results based on process simulation by CAPE tools. (110 JPY/USD)

Even for the results in Cases 2 to 4, the contribution of Ac_2O is significant rather than the other factors. As it is important to reduce the consumption of Ac_2O, it is effective to mitigate the reaction of Ac_2O with the water content in the KP, e.g., 27.8 wt% of water in this paper. However, it is essential to consider the effect of moisture in KP on the kneading mechanism of friction and disintegration. In addition, although there is a benefit from the recovered AcOH, Case 4 is not the best case because the increase in LC-GHG and production costs due to energy consumption is greater than the increase in AcOH recovery by building more distillation columns.

The results of Case 1 are based on the assumption that the production is carried out as it was in the laboratory. Such calculations have been adopted if environmental impacts and costs are needed for technologies with low technology readiness level (TRL), especially in the evaluation of emerging technologies. The CAPE tool particularly enables analyses of the additional energy input required to recover unreacted material, which is usually not done in the laboratory. The change from Case 1 to Case 2 can therefore be easily analysed using the CAPE tool, which also simulates the relationship between the recovery of by-products, i.e., AcOH, and the additional energy required as shown in the results of Cases 2 to 4. In Figure 1, the CAPE tool can be used for the activity: "Provide resource", which allows a certain analysis of the low TRL emerging technology.

4. Conclusions

In this study, we performed simulation-based life cycle inventory analysis to reveal the environmental and economic performance of AcCNF-RP considering the future scale-up of production processes. Through this case study, it was demonstrated that the application of CAPE tools into prospective LCA enables the strategic technology assessments for systems design. Especially in the proof of concept on technology implementation can be verified and validated with the ranged values of uncertainties in emerging technology under development.

CAPE tools have huge potentials for systems design and assessment adopting emerging technologies, which are necessitated towards carbon neutral society. Especially in chemical production, biomass-derived production can become one of the production routes with sustainable feedstocks. Not only conversion routes, but also the acquisitions of feedstocks from agriculture or forestry are now under development and construction. Before their huge installation, CAPE tools should be combined with prospective LCA to visualize the performances of such low TRL emerging technologies.

Acknowledgement

We thank Mr. Dai Kikuchi for the supports on data collection for LCA of cellulose nanofiber-reinforced plastics. This work was supported by MEXT/JSPS KAKENHI Grant Number JP21H03660. Activities of the Presidential Endowed Chair for “Platinum Society” at the University of Tokyo are supported by the KAITEKI Institute Incorporated, Mitsui Fudosan Corporation, Shin-Etsu Chemical Co., ORIX Corporation, Sekisui House, Ltd., the East Japan Railway Company, and Toyota Tsusho Corporation.

References

R. Arvidsson, A.M. Tillman, B.A. Sandén, M. Janssen, A. Nordelöf, D. Kushnir, S. Molander. 2018, Environmental assessment of emerging technologies: recommendations for prospective LCA. J Ind Ecol, 22(6), 1286-1294.

S.J. Eichhorn, A. Dufresne, M. Aranguren, N.E. Marcovich, J.R. Capadona, S.J. Rowan, C. Weder, W. Thielemans, M. Roman, S. Renneckar, W. Gindl, S. Veigel, J. Keckes, H. Yano, K. Abe, M. Nogi, A.N.. Nakagaito, A. Mangalam, J. Simonsen, A.S. Benight, A. Bismarck, L.A.. Berglund, T. Peijs. 2010. Review: Current International Research into Cellulose Nanofibres and Nanocomposites. J. Mater. Sci., 45, 1−33.

Y. Kanematsu, Y. Kikuchi, H. Yano. 2021. Life Cycle Greenhouse Gas Emissions of Acetylated Cellulose Nanofiber-reinforced Polylactic Acid Based on Scale-up from Lab-scale Experiments, ACS Sustainable Chem Eng, 9(31), 10444-10452

Y. Kikuchi, M. Hirao. 2009, Hierarchical Activity Model for Risk-Based Decision Making Integrating Life Cycle and Plant-Specific Risk Assessments, J Ind Ecol, 13(6) 945-964.

Y. Kikuchi, K. Mayumi, M. Hirao. 2010. Integration of CAPE and LCA Tools in Environmentally-Conscious Process Design: A Case Study on Biomass-Derived Resin, Computer-Aided Chemical Engineering, 27, 1051-1056.

Y. Kikuchi. 2014. Activity and Data Models for Process Assessment Considering Sustainability, Kagaku Kogaku Ronbunshu, 40(3) 211-223.

Kyoto University. 2019. Entrusted Work for Performance Evaluation Project of Cellulose Nanofiber (Proving, Evaluation, and Verification for Introduction of CNF Materials for Social Implementation: Automotive Field), 2020. http://www.env.go.jp/earth/mat49_kyoto-univH31.pdf (accessed November 2021).

S.M. Moni, R. Mahmud, K. High, M. Carbajales-Dale. 2020. Life cycle assessment of emerging technologies: A review. J Ind Ecol, 24, 52‑63.

D. Rotolo, D. Hicks, B.R. Martin. 2015. What is an emerging technology? Res. Policy, 44(10), 1827-1843.

N. Thonemann, A. Schulte, D. Maga. 2020. How to Conduct Prospective Life Cycle Assessment for Emerging Technologies? A Systematic Review and Methodological Guidance. Sustainability, 12(3), 1192.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50312-2

Climate Control in Controlled Environment Agriculture Using Nonlinear MPC

Wei-Han Chen[a*], Fengqi You[a]

[a]*Cornell University, Ithaca, New York, 14853, USA*
wc593@cornell.edu

Abstract

The climate in controlled environment agriculture (CEA) is a highly nonlinear complex system that contains nonlinearity. In addition, there are dependencies between each system state. In order to simultaneously control multiple system states in CEA climate, this paper develops a nonlinear model predictive control (NMPC) framework for the CEA climate control to minimize the total control cost and the constraint violation probability. The nonlinear dynamic model of the CEA climate, including temperature, humidity, CO_2 level, and lighting, will be first constructed. After constructing all dynamic models, historical weather data is gathered to identify the system parameters for the nonlinear CEA climate model. A nonlinear optimization problem can then be developed to obtain the optimal control inputs. A case is used to demonstrate the performance of the proposed NMPC framework.

Keywords: Controlled environment agriculture, temperature, humidity, CO_2, lighting.

1. Introduction

Because the controlled environment agriculture (CEA) climate is a multi-input multi-output system, model predictive control (MPC) has advantages over other classical control methods (e.g., On-Off control and proportional–integral–derivative (PID) control). In some past studies, linear MPC is adopted for CEA climate control (Piñón et al., 2005). However, CEA climate contains nonlinearity due to the complex system itself and the relationship between each system state (Chen and You, 2021), which makes nonlinear MPC (NMPC) a suitable approach to deal with the nonlinearity within CEA climate control (Ding et al., 2018). Several system states should be considered in CEA climate control. Among the studies that adopted NMPC, many only control one or two system states instead of simultaneously controlling all four system states mentioned (Blasco et al., 2007; Gruber et al., 2011; Liang et al., 2018; Lin et al., 2021). So far, there is still a lack of comprehensive studies that integrate temperature, humidity, CO_2 concentration, and lighting using the NMPC framework. Therefore, in this work, we propose a novel NMPC framework for CEA climate control to minimize the total control cost and the constraint violation probability.

2. Dynamic model formulation

Within an MPC framework for CEA climate control, a dynamic model is required in order to predict CEA climate (e.g., temperature, humidity, CO_2 concentration, and lighting) as a function of control inputs and certain disturbances, to minimize the control cost and avoid the CEA climate from becoming harmful to crop growth.

In this work, the states we consider are indoor air temperature, relative humidity, CO_2 concentration, and photosynthetically active radiation (PAR). Control actuators are fans, pad cooling, CO_2 injection, supplemental lightings, and blinds. The disturbances considered are ambient temperature, ambient relative humidity, ambient CO_2 concentration, and solar radiation. The structure of the proposed dynamic model for CEA climate control is shown in Figure 1.

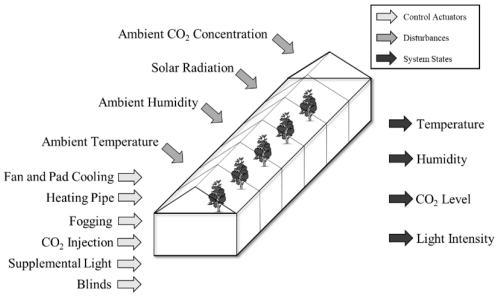


Figure 1: CEA structure that shows control actuators, disturbances, and system states.

The differential equation can be determined by analyzing the energy and mass balance between objects. The continuous-time CEA temperature model can be generally described as the following form:

$$\rho V C_a \frac{dT_i}{dt} = q_{sol} + q_{pipe} + q_{light} - q_{cover} - q_{vent} - q_{pad} \tag{1}$$

where ρ is the air density, V is the volume of the CEA, C_a is the air specific heat, T_i is the indoor temperature, q_{sol} is the net solar radiation, q_{pipe} is the heat flux from heating pipes, q_{light} is the heat flux from lighting, q_{cover} is the heat flux through the cover, q_{vent} is the heat flux from ventilation, and q_{pad} is the heat flux from the pad.

The humidity inside the indoor can also be modeled by differential equations. The absolute humidity is first modeled by using the mass balance equation. The relative humidity is then calculated from absolute humidity and indoor temperature (Chen and You, 2022). The mass balance equation of water, including the net flow from ventilation, evapotranspiration, and the fogging system, is shown as:

$$\rho V \frac{\mathrm{d}h_i}{\mathrm{d}t} = m_{vent} + m_{trans} + m_{fog} \tag{2}$$

where h_i is the absolute humidity, m_{vent} is the water net flow from ventilation, m_{trans} is the water net flow from transpiration of the plants, and m_{fog} is the water net flow from the fogging system.

In the CO_2 mass balance equation, the photosynthesis process consumes CO_2. The mass balance of CO_2 level is shown as:

$$\rho V \frac{\mathrm{d}X_i}{\mathrm{d}t} = X_v - X_{pho} + X_{inj} \tag{3}$$

where X_i is the indoor CO_2 concentration level, X_v is the CO_2 net flow from ventilation, X_{pho} is the net consumption from photosynthesis, and X_{inj} is the control input of CO_2 injection. The total light intensity in the CEA is the sum of the PAR from the sun and PAR provided by the supplemental lightings. The light intensity model can be shown as:

$$I = I_a K_a \left(1 - \tau_s u_{blind}\right) + \tau_l u_{light,\max} u_{light} \tag{4}$$

where I_a is the outdoor global radiation, K_a is the coefficient of the solar equation, τ_s is the shading percentage, u_{blind} is the control input of blind ranging from 0 to 1, τ_l is the energy to light conversion percentage, $u_{light,\max}$ is the maximum energy input of the supplemental light system, and u_{light} is the control input of the supplemental light ranging from 0 to 1.

3. Nonlinear model predictive control

NMPC is used for controlling the CEA climate in this work. The system dynamic models are discretized using the Euler method. Under a given length of prediction horizon *H*, a compact form of the dynamic CEA climate model can be expressed as:

$$\mathbf{x} = \mathbf{f}\left(x_0, \mathbf{u}, \mathbf{v}\right) \tag{5}$$

where **x**, **u**, and **v** are the system state, control input, and disturbance sequences vectors, respectively, and x_0 is the initial system state.

After constructing the dynamic model, we could then develop the nonlinear optimization problem to be solved at each time step. For the optimization problem, there are constraints on system states and control inputs. The constraints are defined for control inputs and system states throughout the entire prediction horizon *H*. The control inputs should be between the minimum and maximum values. A CEA environment should also be maintained within a specified range to facilitate the growth of plants and fruits and prevent them from being damaged by harsh climate conditions. The compact form of constraints for control inputs and system states can be represented as:

$$\mathbf{G}_x \mathbf{x} \le \mathbf{g}_x, \quad \mathbf{G}_u \mathbf{u} \le \mathbf{g}_u \tag{6}$$

where $\mathbf{G}_x$ and $\mathbf{g}_x$ are vectors that define the system state constraints in compact form. $\mathbf{G}_u$ and $\mathbf{g}_u$ are vectors that represent the control input constraints in compact form.

Once the dynamic model and the constraints are prepared, the optimization problem can then be written out. The objective function is to minimize the total control cost. **cc** represents the cost coefficient for different control actuators, and the coefficients are higher for actuators using energy with higher energy costs. A vector of slack variables $\boldsymbol{\varepsilon}$ is added to the objective function because there are limitations on control inputs which could cause the optimization problem to become infeasible. In order to penalize the constraint violation, the penalty weight **S** is added to the objective function. Since the slack variables are always positive, the state constraints are therefore softened (Lu et

al., 2020). In this control framework, the optimization problem is solved to get the optimal control inputs for each time step. The system states at the next time step can be updated by using the dynamic models.

$$\begin{aligned} \min_{u} \; & J = \mathbf{c}\mathbf{c}^T\mathbf{u} + \boldsymbol{\varepsilon}^T\mathbf{S}\boldsymbol{\varepsilon} \\ \text{s.t.} \; & \mathbf{x} = \mathbf{f}\left(x_0, \mathbf{u}, \mathbf{v}\right) \\ & \mathbf{G}_x\mathbf{x} \le \mathbf{g}_x + \boldsymbol{\varepsilon} \\ & \mathrm{G}_u\mathbf{u} \le \mathbf{g}_u \\ & \boldsymbol{\varepsilon} \ge 0 \end{aligned} \tag{7}$$

4. Case studies on simulated CEA

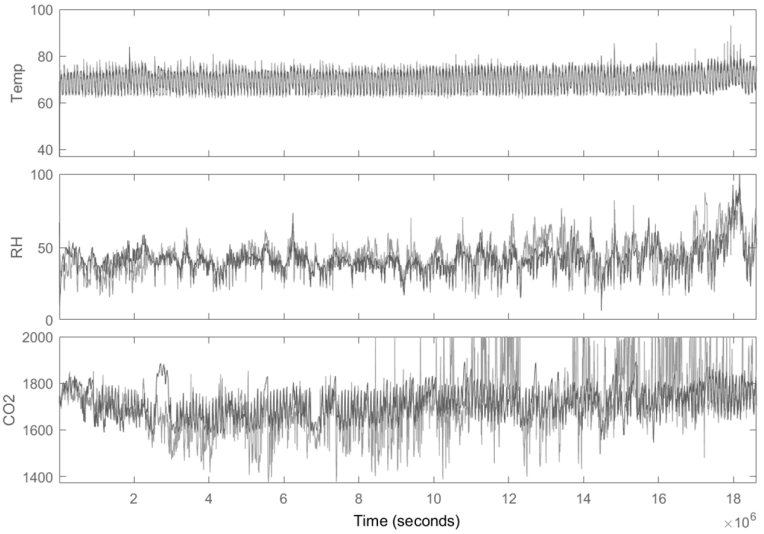


Figure 2: System identification results for greenhouse climate. The darker lines are the predicted trajectories by system identification; the lighter ones are measurement data.

In this work, a CEA located in Ithaca, New York, USA, for tomato production is simulated for closed-loop temperature, humidity, CO_2 concentration level, and lighting control under the NMPC control framework. The system states controlled in this work are indoor temperature, relative humidity, CO_2 level, and PAR. System identification is first conducted to obtain the undetermined parameters in the CEA climate model. The weather data and CEA indoor climate data from November 1, 2019, to May 31, 2020, are gathered to conduct the system identification. The simulation is performed for one week in winter during December 17-23, 2019. The weather forecast data from December 17-23 are collected for the optimization problem. The actual measurement data at the same period are also collected to reveal the system states at the next time step. The sampling interval is 15 min, and the control horizon is 5 hours. The average CPU time for solving the optimization problem is 2.53 seconds on a computer with an Intel Core i7-6700 processor at 3.40 GHz and 32 GB of RAM.

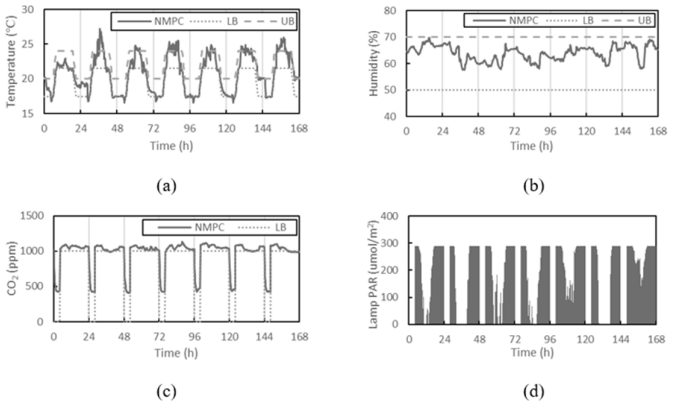


Figure 3: The profiles in winter (December 17-23, 2019), (a) temperature profile, (b) humidity profile, (c) CO_2 level profile, (d) lamp PAR profile.

Figure 2 shows the system identification results of indoor temperature, relative humidity, and CO_2 concentration. The darker lines are the trajectories predicted by system identification, and the lighter lines are the trajectories of CEA measurement data. The figure shows that the temperature model and relative humidity model predict better than the CO_2 model with around 78% and 65% accuracy compared to the 43% accuracy of the CO_2 model. The CO2 model is not as accurate as of the other two models due to the lack of CO_2 injection data. The way to overcome the lack of CO_2 input data is by estimating data using back-calculation. The CO_2 model can still be used to minimize the total control cost in practice because the system states would be updated in each time step. The effect of model error will then be reduced.

Figure 3(a) shows the temperature profile in winter during December 17-23, 2019. The lower and upper bound are set differently throughout the day for the light period and dark period of the photosynthesis. The light period starts from 4 am to midnight and the dark period is between midnight and 4 am. When the CEA changes from light period to dark period or vice versa, the lower and upper bound are set to be gradually increased or decreased to avoid the abrupt changes of the indoor temperature. There is a clear diurnal pattern, and the profile could be maintained within the region between lower and upper bounds. However, the constraint violation still occurs sometimes due to forecast errors. The relative humidity profile in winter during December 17-23, 2019, is presented in Figure 3(b). The lower bound and upper bound are set as 50% and 70%. The humidity profile is better maintained within the region between lower and upper bounds compared to the temperature profile in winter. The cold outdoor air is drier than indoor air so that the ventilation system could help draw in the outdoor air when the humidity level is about to surpass the upper bound. Figure 3(c) depicts the CO_2 level profile in winter during December 17-23, 2019. The CO_2 injection occurs from 4 am to midnight for the light period. During the light period, the CO_2 level is maintained above 1000 ppm to stimulate tomato growth. CO_2 is not necessary during the dark period, so the CO_2 injection is set to zero. The CO_2 level would gradually drop to the ambient level because of the ventilation. The lamp PAR result in winter during December 17-23, 2019,

can be found in Figure 3(d). The supplemental lightings are turned off during the dark period from midnight to 4 am and are turned on during the light period to ensure the plants receive sufficient PAR. When the sunlight PAR is adequate during the day, lamps are turned off to reduce energy consumption. Yet, supplemental lightings are required even at midday when the sunlight PAR is insufficient, which is the case on the last day.

5. Conclusions

In this work, we developed a nonlinear MPC framework for a CEA that could simultaneously control multiple system states of the CEA climate. Energy and mass balance equations followed by system identification were utilized to generate nonlinear dynamic models for temperature, relative humidity, CO_2 concentration level, and light intensity. The nonlinear optimization problem was then solved at each time step to obtain the optimal control inputs for the CEA climate. A case study of a CEA located in Ithaca, New York was conducted. The results showed the NMPC framework could efficiently minimize total control cost and constraint violation. Future extensions of this work could include irrigation control (Shang et al., 2020), sensor integration (Chen et al., 2021), and accounting for uncertainty in weather forecast (Shang et al., 2019).

References

X. Blasco, M. Martínez, J. M. Herrero, C. Ramos, and J. Sanchis, 2007, Model-based predictive control of greenhouse climate for reducing energy and water consumption, Computers and Electronics in Agriculture, 55(1), 49-70.

W.-H. Chen, F. You, 2021, Smart greenhouse control under harsh climate conditions based on data-driven robust model predictive control with principal component analysis and kernel density estimation. Journal of Process Control, 107, 103-113.

W.-H. Chen, C. Shang, S. Zhu, et al., 2021, Data-driven robust model predictive control framework for stem water potential regulation and irrigation in water management. Control Engineering Practice, 113, 104841.

W.-H. Chen, F. You, 2022, Semiclosed Greenhouse Climate Control Under Uncertainty via Machine Learning and Data-Driven Robust Model Predictive Control. IEEE Transactions on Control Systems Technology, DOI:10.1109/tcst.2021.3094999.

Y. Ding, L. Wang, Y. Li, and D. Li, 2018, Model predictive control and its application in agriculture: A review, Computers and Electronics in Agriculture, 151, 104-117.

J. K. Gruber, J. L. Guzman, F. Rodriguez, C. Bordons, M. Berenguel, and J. A. Sanchez, 2011, Nonlinear MPC based on a Volterra series model for greenhouse temperature control using natural ventilation, Control Engineering Practice, 19(4), 354-366.

M.-H. Liang, L.-J. Chen, Y.-F. He, and S.-F. Du, 2018, Greenhouse temperature predictive control for energy saving using switch actuators, IFAC-PapersOnLine, 51(17), 747-751.

D. Lin, L. Zhang, and X. Xia, 2021, Model predictive control of a Venlo-type greenhouse system considering electrical energy, water and carbon dioxide consumption, Applied Energy, 298, 117163.

S. Lu, J.H. Lee, F. You, 2020, Soft-constrained model predictive control based on data-driven distributionally robust optimization. AIChE Journal, 66, e16546.

S. Piñón, E. F. Camacho, B. Kuchen, and M. Peña, 2005, Constrained predictive control of a greenhouse, Computers and Electronics in Agriculture, 49(3), 317-329.

C. Shang, W.-H. Chen, A.D. Stroock, F. You, 2020, Robust Model Predictive Control of Irrigation Systems With Active Uncertainty Learning and Data Analytics. IEEE Transactions on Control Systems Technology, 28, 1493-1504.

C. Shang, F. You, 2019, A data-driven robust optimization approach to scenario-based stochastic model predictive control. Journal of Process Control, 75, 24-39.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50313-4

Thermodynamic Analysis of an Integrated Renewable Energy Driven EWF Nexus: Trade-off Analysis of Combined Systems

Jamileh Fouladi, Ahmed AlNouss, Yusuf Bicer, Tareq Al-Ansari*

College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar

**talansari@hbku.edu.qa*

Abstract

The utilization of energy, water and food resources across multiple technologically driven sub-systems has attracted much attention in the literature. In this work, a solar energy-based system integrated with a utility unit is proposed to generate fresh water from seawater by desalination, power, ammonia/urea, and syngas from biomass utilisation. The main objective is to design a combined integrated system (solar/biomass system) based on solar energy, and to study the dependence of the Energy-Water-Food nexus performance on the solar capacity. The core components of the developed system include solar thermal collectors, Rankine cycle, reverse osmosis (RO) desalination unit, food/agriculture sector, biomass gasification process, ammonia and urea production units. The beneficial uses of reverse osmosis brine streams in agriculture sector is considered. The syngas produced from gasification process is used for Ammonia/Urea production as well as in the Rankine cycle to generate electricity. A comphrensive thermodynamic model and energy-exergy balances are used to assess the performance of the proposed system using the Engineering Equation Solver (EES). Different scenarios are solved to capture the trade offs amongst different technologies and explore the optimum EWF interlinkages. Furthermore, the effects of different load changes such as solar radiation and ambient temperature on some of the outputs of the system are investigated. In addition, the energy and exergy efficiencies of the system are calculated and compared.

Keywords: Multi-generation system, EWF Nexus, Solar, Biomass, Exergy.

1. Introduction

As the global population increases towards 9 billion in 2050, the need for energy, water and food (EWF) resources will increase accordingly. It is expected that the demand for food and water resources will rise by 50 % (Karan et al., 2018). Incidentally, there are inherent inter-linkages between EWF resources. As such, the EWF nexus concept was developed at the Bonn Nexus Conference in 2011 as a consequence to the realization of these inter-dependencies. The underlying analysis within the EWF Nexus concept enables the identification of the inter-linkages amongst different resources, and as such supports the identification of synergies and trade-offs (Al-Ansari et al., 2017; 2015). Multi-generation systems, which are integrated resource systems can support the

development of EWF nexus systems. For instance, Luqman et al. (2020) presented a multigeneration system that explores the thermodynamics of oxy-hydrogen combustor based on wind and solar energy. The useful products of the developed system are freshwater, power, hydrogen, cooling, and domestic water heating. Energy and exergy balances are used to evaluate the performance, where the overall respective efficiencies of the system were 50 % and 34 %, respectively. Nazari and Porkhial (2020) developed a multi-generation system that integrates solar energy and a biomass utilization unit to generate heating, cooling, freshwater, and electricity. The thermodynamic and economic model of the system was conducted, which demonstrates that through biomass integration, an exergy efficiency of 21.48 % can be achieved. Moreover, a multi-objective optimization problem considering second law efficiency maximization and total product cost minimization were solved. The cost of the optimum solution is decreased by 10 %, while the exergy efficiency was 0.2% higher than the base case results. Ghasemi et al.(2018) analyzed a multi-generation system operating through solar-biomass energy using a thermodynamic and thermo-economic approach. The energy system includes both the desalination process and liquefaction of natural gas (LNG). Results indicate that the proposed system has energy and exergy efficiencies of approximately 46 % and 11 % respectively.

Evidently, there are numerous multi-generation systems, which study the utilization of waste streams, demonstrating that integrating renewable resources can enhance sustainability, decentralization and resilience of integrated systems. Incidentally, in most previous studies, biomass utilization was studied without considering the wider EWF nexus elements. The hybrid EWF nexus and multigeneration concepts can be further expanded to include biomass utilization, associated with food production units and the energy-water nexus in an integrated manner (Fouladi et al., 2021). The novelty of this work is the trade-off analysis of an integrated solar-biomass energy system driven by the EWF nexus combined with a fertilizer production unit. The main objectives of this study are to design an integrated renewable EWF nexus system, and to capture the trade-offs between multiple resources to optimize the EWF interlinkages.

2. System Description

Figure 1 illustrates a representation of the integrated EWF multigeneration system. In this multigeneration system, the main sources of energy are from the syngas generated from biomass gasification process and solar energy. A Rankine cycle is used within the system to produce power and to be utilized within the existing energy sinks. Parabolic Trough Collectors (PTCs) are integrated to utilize solar energy to satisfy the required demand. The main components of the system include an agriculture unit, a reverse osmosis desalination plant, a biomass utilization process, and an ammonia/urea plant for fertilizer production. Moreover, the brine stream from the desalination plant is integrated to enhance the system and reduce. Finally, to improve the environmental emissions of the overall system, the potential for CO_2 reduction via capture technology from the Rankine cycle for reuse within urea production is considered.

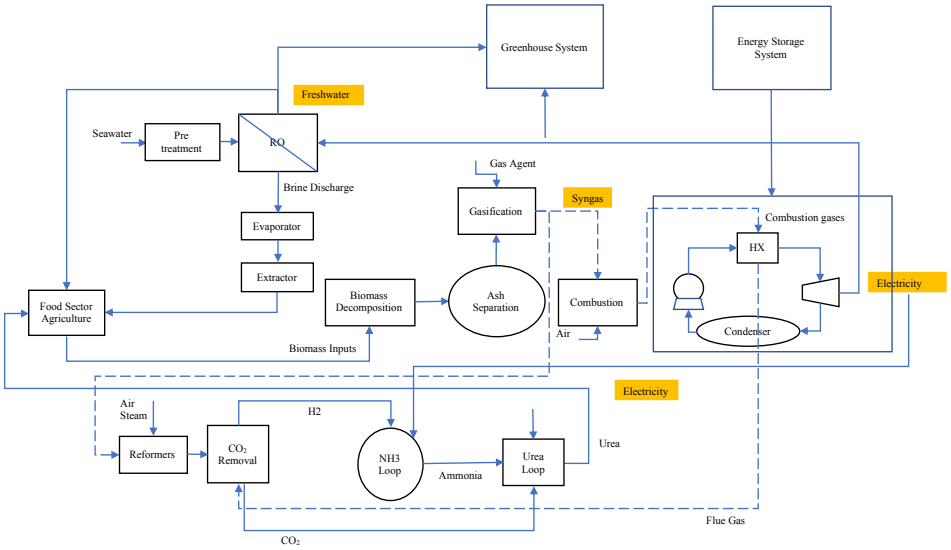


Figure 1: Proposed Integrated System.

Multiple assumptions and input data have been used to simulate the integrated system using EES. Table 1 indicates the main parameters used in this study.

Table 1: Input data for the proposed system.

Parameters	Value
Reference Temperature, T_0 and Pressure, P_0	25 °C and 101 kPa
Biomass gasifier operating conditions	888 °C and 101 kPa
Reverse osmosis recovery ratio	0.4
Seawater salinity	35000 ppm
Fresh water salinity	450 ppm
Isentropic efficiencies of pump and turbine	85 %
Rankine cycle pressure ratio	100
Surface temperature of the Sun	5500 °C

3. Thermodynamic Analysis & Results

The thermodynamic analysis of the integrated system is performed using mass, energy, entropy, and exergy balances (using the first and second laws of thermodynamics). Therefore, to evaluate the performance of every component and the overall system, the energy and exergy efficiencies of all units in the proposed system are calculated. Figure 2 illustrates the obtained efficiency values. Overall, the energy and exergy efficiencies of the system is found to be approximately 46 % and 19 %, respectively. Furthermore, the maximum exergy destruction rates correspond to the gasifier and ammonia process.

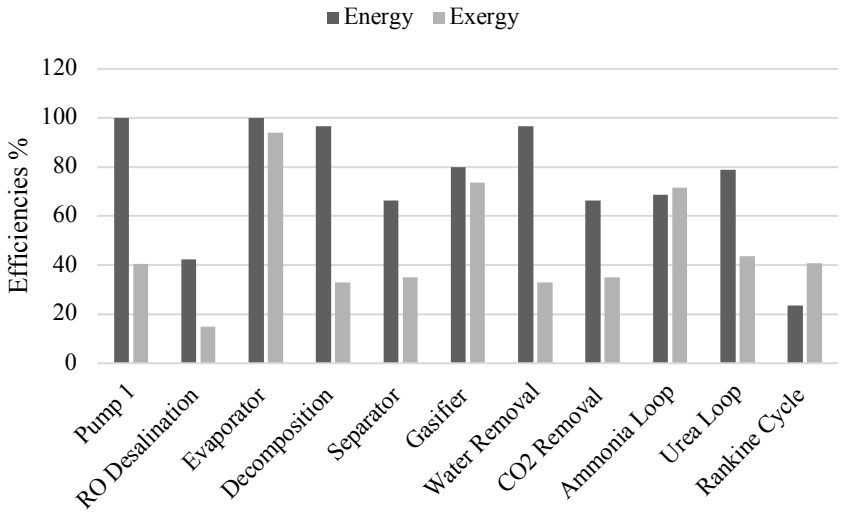


Figure 2: Energy and Exergy efficiencies of all components.

The trade-offs analysis of the integrated system demonstrates that the syngas produced from the biomass utilization process offsets the natural gas requirement as input to other processes, which results in reduced energy consumption. Figure 3 indicates that by optimizing the energy-food nexus segment and increasing the biomass flow rate, the overall energy and exergy efficiencies of the system reduce. The large exergy destruction rates of these processes are the main reason of this behaviour. The most common sources of irreversibility in the gasifier and ammonia unit are the chemical reactions occurring within the process, which lead to destruction of chemical exergies. Moreover, physical exergy destruction are also associated with the expansion and compression unit within the plants.

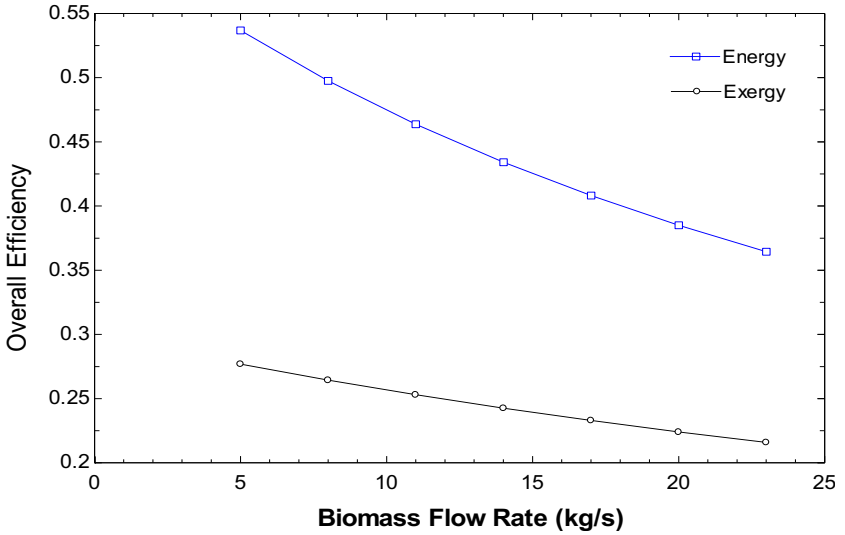


Figure 3: Effect of input biomass rate variation on overall system efficiencies.

Furthermore, the impact of changing the solar radiation on the performance of the turbine and seawater supply flow are illustrated in Figure 4 and Figure 5. The solar radiation linearly affects the outputs of the combined system. By increasing the solar radiation value, the net power generated by the Rankine cycle turbine increases linearly. This is due to the higher outlet temperature of the PTCs, which causes a higher temperature at the inlet of the Rankine cycle as well.

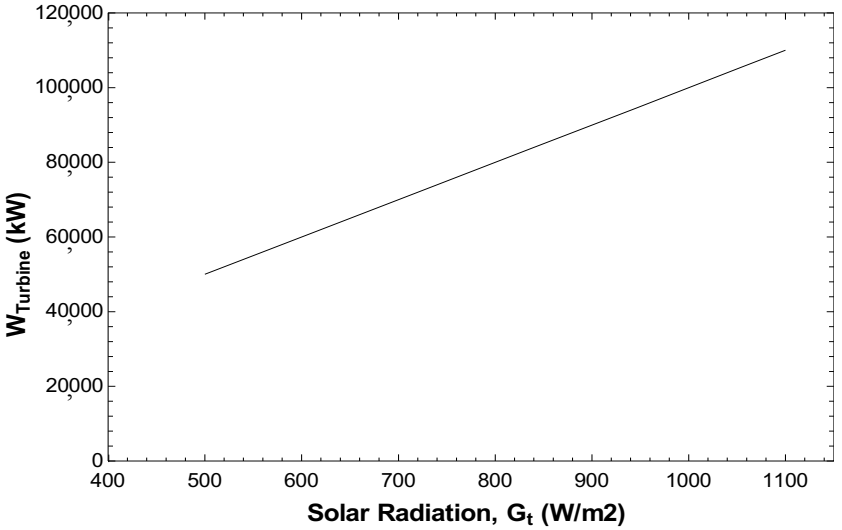


Figure 4: Effect of solar radiation variation on work generated by turbine.

From Figure 5, it is observed that the mass flow rate of seawater in the desalination unit increases as the solar radiation rate increases. The parametric studies demonstrate that the variation of some parameters has direct impact on the overall efficiencies of the system, which further integrates the system by optimizing the operational conditions.

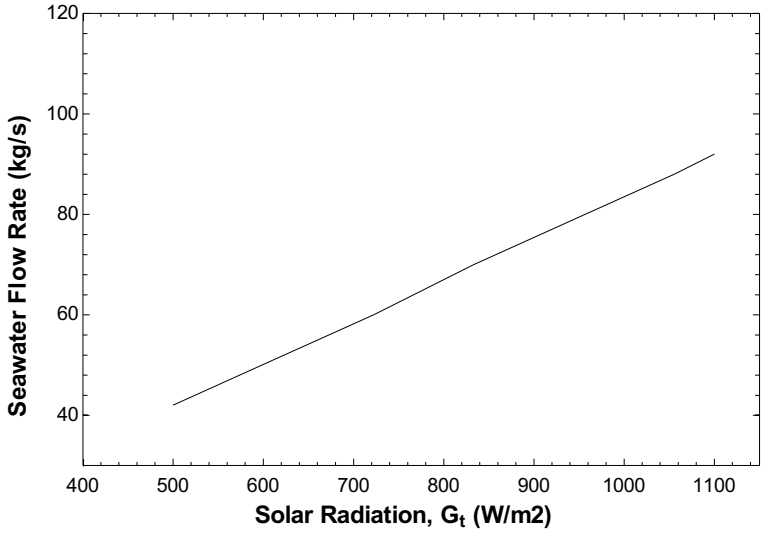


Figure 5: Effect of solar radiation variation on seawater flowrate

4. Conclusion

This study developed an integrated renewable energy driven EWF nexus system by combining various units, such as those within food systems, Rankine cycle, biomass utilization, ammonia/urea process and desalination. Electricity, freshwater, and fertilizer are the main outputs of the proposed system. Thermodynamic analysis including mass, energy, entropy, and exergy balances are implemented using Engineering Equation Solver (EES). Exergy destruction rates are calculated, which indicates the units that can be further integrated and optimized to enhance the exergy efficiencies. Using parametric studies, variation in certain factors are studied. The system generated net electrical power of almost 78 MW, 28 kg/s of freshwater, and approximately 7 kg/s of fertilizer. The maximum exergy destruction rates are within the biomass utilization unit and Ammonia production unit. The feed biomass flow rate and solar radiation value affects the different outputs and the overall performance of the proposed system. For future studies, cooling/heating loads can be integrated further into the system to enhance the resilience.

5. Acknowledgment

The authors acknowledge the support of Qatar National Research Fund (a member of Qatar Foundation) provided by GSRA grants no. GSRA6-1-0416-19014.

References

A. Ghasemi, P. Heidarnejad, A. Noorpoor, 2018, A novel solar-biomass based multi-generation energy system including water desalination and liquefaction of natural gas system: Thermodynamic and thermoeconomic optimization, J. Clean. Prod, 196, 424–437.

E. Karan, S. Asadi, R. Mohtar, M. Baawain, 2018, Towards the optimization of sustainable food-energy-water systems: A stochastic approach, J. Clean. Prod, 171, 662–674.

J. Fouladi, A. AlNouss, T. Al-Ansari, 2021, Sustainable energy-water-food nexus integration and optimisation in eco-industrial parks. Computers & Chemical Engineering, 146, 107229.

M. Luqman, Y. Bicer, T. Al-Ansari, 2020, Thermodynamic analysis of an oxy-hydrogen combustor supported solar and wind energy-based sustainable polygeneration system for remote locations, Int. J. Hydrogen Energy, 45, 3470–3483.

N. Nazari, S. Porkhial, 2020, Multi-objective optimization and exergo-economic assessment of a solar-biomass multi-generation system based on externally-fired gas turbine, steam and organic Rankine cycle, absorption chiller and multi-effect desalination, Appl. Therm. Eng, 179, 115521.

T. Al-Ansari, A. Korre, Z. Nie, N. Shah, 2017, Integration of greenhouse gas control technologies within the energy, water and food nexus to enhance the environmental performance of food production systems, J. Clean. Prod. 162, 1592–1606.

T. Al-Ansari, A. Korre, Z. Nie, N. Shah, 2015, Development of a life cycle assess- ment tool for the assessment of food production systems within the energy, water and food nexus, Sustainable Production and Consumption, 2, 52–66.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50314-6

Low-Carbon Hydrogen Production in Industrial Clusters

Yasir Ibrahim, Mohammad Lameh, Patrick Linke, Dhabia M. Al-Mohannadi*

Departement of Chemical Engineering, Texas A&M University at Qatar, Education City, PO Box 23874, Doha, Qatar
dhabia.al-mohannadi@qatar.tamu.edu

Abstract

Hydrogen has gained a huge hype as future fuel that might aid in the transformation to a zero-carbon energy system. In this work, a dynamic model of green hydrogen production is proposed to assess the trade-offs between the economic and environmental impacts of incorporated green hydrogen in energy-intensive chemical processes. The model considers the intermittency of renewable energy sources and their effect on the sizing of the renewable energy and hydrogen production units, as well as on the size required for H_2 storage. The validity of the model was tested in the decarbonization of two energy-intensive processes namely, ammonia and methanol. The results outputs clearly show the trade-offs between the economic and environmental performances at various decarbonization targets based on hourly solar availability from a time horizon of one year.

Keywords: Renewable Hydrogen, Decarbonized Economy, Solar Electrolysis

1. Introduction

In light of climate change, the Intergovernmental Panel on Climate Change (IPCC) establishes a protocol to cut down the Greenhouse Gas (GHG) emissions in order to achieve so-called carbon neutrality by 2050 (IPCC 2018). Unfortunately, the majority of the global energy is generated by carbon-based resources, which leads to the continuous release of CO_2 emissions. Most of these emissions are driven from hard-to-abate sectors, such as petrochemical production, steel, oil refining, etc.In this context, Hydrogen (H_2) would play a substantial role to decarbonize the global energy system since it is a clean-burning molecule. Many factors are reinforcing why hydrogen is the key block in energy transition, which are, H_2 can solve the renewable energy intermittency issue by utilizing H_2 as a cleaner, affordable, and available storage. In addition, H_2 can deliver a deep reduction in CO_2 emissions specifically in the hard-to-abate sectors. In these sectors, renewable energy may not contribute significantly to decarbonize of these sectors as much as it can contribute to utility and power sectors. Although there many factors favoring a sustainable uprising in the investment of H_2 , which are substantially stronger than any period, significant challenges are yet to be overcome. The major challenge to be addressed is the enormous emissions associated with current H_2 production. The majority of H_2 production accounting for 95% of the world capacity is produced from fossil fuels via Steam Methane Reforming (SMR), resulting in so-called grey H_2 (Renewable and Agency 2019). One promising solution is to produce the H_2 from a wide variety of renewable resources, which is labelled as green H_2. The main technology to produce green H_2 is water electrolysis in where water decomposed into H_2 and Oxygen (O_2) molecules in presence of electric current. For green H_2 to become a major energy carrier, production scale will need to be increased and in so partly address the production cost.

Notwithstanding, there has been little quantitative analysis of green H_2 from variable renewable energy in off-grid connection. (Schnuelle et al. 2020) developed a simulation model to determine the economic performance based on determined operating characteristics of onshore and offshore wind, as well as Photovoltaics (PV) plants. (Decker et al. 2019) evaluated the cost breakdown of a baseline case for PEM in on-grid settings. (Glenk and Reichelstein 2019) proposed a techno-economic model to investigate the H_2 produced from variable renewable energy considering the scalability of the electrolyzer in their analysis. (Mallapragada et al. 2020) developed a framework to decide the plant size and operating condition through the optimization of the size of the components taking into account hourly solar availability, and production requirement. However, it did not consider the utilization of the H_2 produced. (Koleva et al. 2021) established a mathematical model to evaluate PV-powered water electrolysis from an economic point of view. The model was tested under different weather conditions. The majority of the previous work has been focused on the economics of green H_2; nevertheless, there is no detailed investigation of dynamic modelling of PV-powered electrolysis. Therefore, a comprehensive dynamic model is developed to assess the trade-offs between the economic and environmental impacts of the green H_2. Furthermore, the dynamic model was formulated based on the intermittency of renewable energy that can affect the size of renewable energy units, the H_2 production and utilization facilities, as well as the H_2 storage. The mass and energy balance is calculated based on the hourly solar availability. The economic and environmental impacts are characterized by two metrics, which are total cost and CO_2 emissions saved respectively. A detailed discussion on the mechanism of the proposed dynamic model is present in the following section.

2. Methodology

The aim of this study is to analyze the economic and environmental impacts of utilizing green hydrogen on the performance of chemical processes that act as H_2 sinks. The study considers the intermittency of renewable energy sources and their effect on the sizing of the renewable energy and hydrogen production units, as well as on the size required for H_2 storage. The economic impact of introducing green H_2 is characterized by the total cost of establishing and operating the different units. Figure 1 shows the flow diagram with the different components considered in the evaluation. H_2 can be supplied to the chemical process either through green hydrogen production, or through grey hydrogen production. The production of grey hydrogen is accompanied with a relatively high level of CO_2 emissions that can be avoided by using green hydrogen with a low environmental footprint. The dynamics of the varying renewable energy source are considered by discretizing the annual operation into hourly time steps; in each time step, mass and energy balance calculations determine the renewable power generated, the rate of production of green hydrogen, and the flowrate of H_2 deliver to and from storage. These outcomes depend on the size of the different units, and on different technical parameters that are used as inputs. The hourly capacity factor of the renewable energy production unit ($CF_{RE}(t)$) can be determined based on the radiation data of the geographic location chosen, and the expected losses in energy transformation. The green hydrogen production unit is characterized by its efficiency-ε (the amount of electricity required to produce H_2 – kWh/kgH_2) and cost-C_{green} ($/kW). The hourly production of green H_2 is determined as shown in Eq.(1)

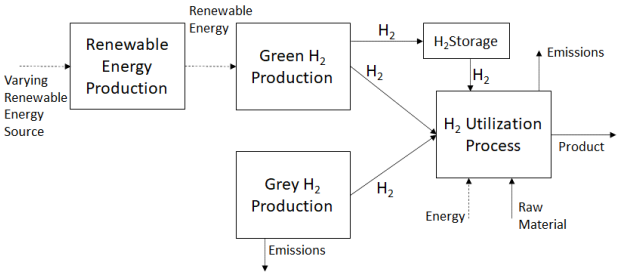


Figure 1 Flow diagram of the considered process

$$F_{H_2}^{Green}(t) = \frac{Size_{RE} \times \mathrm{CF_{RE}(t)} \times \Delta \mathrm{t}}{\varepsilon} \tag{1}$$

The size of the renewable energy production unit is selected so that the total annual production of green hydrogen meets the annual demand for green hydrogen, which is set based on the H_2 intake of the utilization unit (kgH_2/kgProduct), and the contribution of the green hydrogen to the total demand (%Green). This is shown in Eq.(2). Two different cases for addressing the temporal variation in green hydrogen production are investigated in this work: introducing H_2 storage and oversizing the H_2 utilization unit.

$$\sum_{t=1}^{t=t_f} F_{H_2}^{Green}(t) = intake_{H_2} \times Annual\ Product\ Rate \times \%\mathrm{Green} \tag{2}$$

Note that the rate of H_2 production from the grey H_2 unit is equal to the difference between the demand for H_2 and the supply of green H_2. The introduction of H_2 storage allows the utilization of the excess H_2 when the rate of green H_2 production is higher than the demand for H_2 by the sink, while maintaining a consistent production rate by the sink. In this case, the production rate of the sink in each timestep is equal to the annual production rate divided by the number of timesteps. The storage capacity is tracked to determine the initial H_2 storage requirement and the total size of the storage. Eq (3), (4), and (5) show the equations that describe the storage.

$$H_2 toStr\ (t) = (F_{H_2}^{Green}(t) - H_2 demand(t)) \times x(t) \tag{3}$$

$$H_2 fromStr\ (t) = (H_2 demand(t) - F_{H_2}^{Green}(t)) \times (1 - x(t)) \tag{4}$$

$$H_2 inStor\ (t) = H_2 inStr\ (t-1) - H_2 fromStr\ (t) + H_2 toStr\ (t) \tag{5}$$

Note that x(t) is a binary variable that is equal to one when there is excess H_2, and it is zero otherwise. The initial mass of H_2 in the storage ($H_2 inStor\ (0)$) is selected such that the minimum content of the storage is zero. The size of the storage is set equal to the maximum $H_2 inStor\ (t)$ achieved throughout the year. The other option of addressing the dynamic variations in the hydrogen production is to allow the dynamic variation in the production of the H_2 utilization unit. This will result in varying product flowrate, and the size of the sink to allow the utilization of all the hydrogen produced. Hence, for each time step, the produced hydrogen ($F_{H_2}^{Green}(t)$) is determined (equation (1)), and the

corresponding production rate is determined based on H_2 intake, and the contribution of the green hydrogen to the total hydrogen demand (eq.(6)).

$$Product\ Rate\ (t) = \frac{F_{H_2}^{Green}(t)}{intake_{H_2} \times \%\text{Green}} \tag{6}$$

The total cost of the process is the summation of the annualized capital and operating costs of all the considered units. The total environmental impact is characterized by the emissions flowrate from the grey hydrogen production unit and from the sink. The assessment is conducted for different H_2 utilization technologies to investigate the variation in the impact of introducing green hydrogen to different processes on their environmental and economic performance. This can be reflected in the marginal abatement cost (MAC) of introducing green hydrogen, which is defined as below

$$MAC = \frac{Total\ cost\ with\ green\ H_2 - Cost\ of\ Base\ Case}{Emissions\ in\ the\ Base\ Case - Emissions\ with\ green\ H_2} \tag{7}$$

3. Results & Discussion

In this section, the versatility of the dynamic model was tested to assess the trade-offs between the economic and environmental impact of decarbonizing specific energy-intensive chemical processes. Two chemical processes were analyzed namely, ammonia and methanol. The decarbonization is achieved through incorporating direct green H_2. The proposed dynamic model will provide insights into these trade-offs at various decarbonization targets. These targets are varied from 50% to 100% in 10% intervals. In each interval, the dynamic model will determine the size of the production and utilization of the green H_2, and the flowrate of H_2 deliver to and from storage, adhering to the hourly solar availability from a time horizon of one year. For the environmental impacts, the green H_2 contribution is compared with blue H_2, which is grey H_2 accompanied with Carbon Capture and Storage (CCS) based on their MAC. Integrating CCS with grey hydrogen can cut down emissions up 90% (IEA 2019a). The MAC of blue H_2 is not dependent on H_2 utilization process, and it depends only on grey H_2 and H_2 storage. It should be noted that this work uses Proton Exchange Membrane PEM electrolysis for green hydrogen production for its flexibility with the variation in renewable energy and SMR for grey and blue hydrogen. The techno-economic data assumptions are summarized in Table 1. The results of the analysis described in the methods section are shown in figuresFigure 2 andFigure 3.

Table 1 Techno-economic data assumptions

Parameter	Value	Reference
Ammonia Cost ($/ton)	901	(IEA 2019b)
Methanol Cost ($/ton)	392	(Al-Mohannadi et al. 2017)
PV Installation cost ($/kW)	714	(Agency 2020)
PEM Electrolysis Cost ($/kW)	1,100	(IEA 2019b)
Grey H_2 Cost ($/kg H_2)	348	(IEA 2019b)
H_2 Storage Cost ($/kg H_2)	615	(Nordin and Rahman 2019)
OPEX of % CAPEX	2.5	(IEA 2019b)
CCS Cost ($/ton H_2)	80.2	(Ahmed et al. 2020)

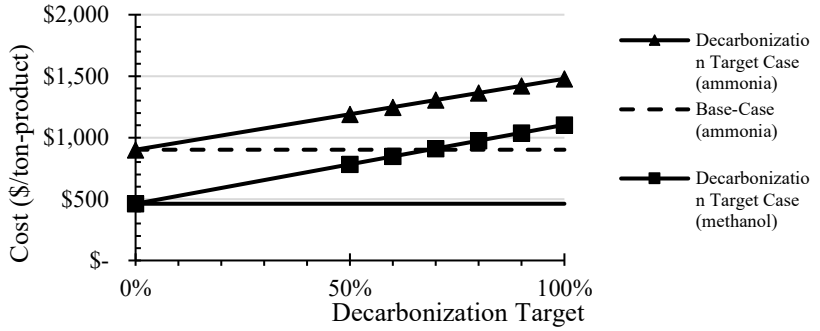


Figure 2 The effect of the decarbonization target on the total cost of utilization plant

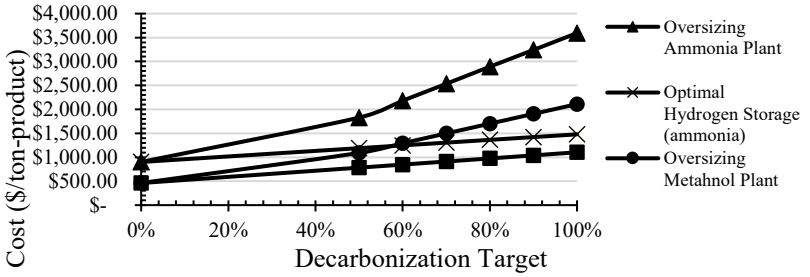


Figure 3 Options comparison between oversizing the utilization plant and the optimal H_2 storage

It can be seen clearly from Figure 2 that the decarbonization target significantly affects the total cost since the cost of ammonia and methanol produced. At 100% green H_2 is 1.6 and 2.4 times that of ammonia and methanol produced at 0% green H_2 respectively. It should be emphasized that the option of oversizing the utilization plant to address the variation of the production of the green H_2 is not attractive from an economic point of view. As shown in Figure 3, there is less variation in the cost for the two options (optimal H_2 storage and oversizing the utilization plant) at a lower decarbonization target. However, at a higher decarbonization target, the variation in the cost is significant since the total cost of oversizing the utilization plant is increased by 143% and 91% for ammonia and methanol respectively. For the environmental impacts, as mentioned in the methodology section MAC metric is selected to assess the environmental performance of incorporating green H_2. The MAC of ammonia and methanol is \$304, and \$802/ton of CO_2 saved respectively. It can be noted that MAC of methanol is higher than MAC of ammonia due to the high emissions associated with methanol production. As expected, the MAC of blue H_2 is lower than the MAC of green H_2 as it was estimated at \$110/ton of CO_2 saved, considering 90% can be captured from both production process and energy emission streams.

4. Conclusions

A dynamic model to evaluate the economic and environmental impacts of incorporating green H_2 for decarbonizing energy-intensive chemical processes. The proposed dynamic model takes into account the intermittency of renewable energy sources and their effect

on the sizing of the renewable energy and hydrogen production and utilization units, as well as on the size required for H_2 storage. This analysis assesses the trade-offs between the economic and environmental impacts at various decarbonization targets based on hourly solar availability from a time horizon of one year. The model was examined to decarbonize two energy-intensive industrial processes at various CO_2 reduction targets. The total cost and MAC measurements are selected as economic and environmental performances. The dynamic model is implemented on a spreadsheet to enable a user-friendly interface so that the results output can be obtained within a matter of seconds. Nevertheless, further aspects still need attention to improve and add resilience to the model. These aspects are investigation of centralized vs decentralized H_2 production facility in terms of economic and environmental impacts. In addition, further modelling work is needed to determine the optimum decisions related to capacity of H_2 production and utilization plants, size of the H_2 storage, and size of PV. Moreover, further research might explore well-established as well as emerging H_2 production technologies (e.g., Photoelectrocatalysis).

References

Agency, International Renewable Energy. 2020. *Renewable Power Generation Costs in 2019*.

R. Ahmed, S. Shehab, D. M. Al-Mohannadi, and P. Linke. 2020, Synthesis of Integrated Processing Clusters, *Chemical Engineering Science* 227, 115922.

D. M. Dhabia M., K. Abdulaziz, S. Y. Alnouri, and P. Linke. 2017. On the Synthesis of Carbon Constrained Natural Gas Monetization Networks. *Journal of Cleaner Production,* 168, 735–45.

M. Decker, F. Schorn, R. Can Samsun, R. Peters, and D. Stolten. 2019,Off-Grid Power-to-Fuel Systems for a Market Launch Scenario – A Techno-Economic Assessment." *Applied Energy* 250, 1099–1109

G. Glenk, and S. Reichelstein. 2019, Economics of Converting Renewable Power to Hydrogen, *Nature Energy* 4(3), 216–22.

IEA. 2019. *World Energy Outlook 2019*.

IPCC. 2018. *Global Warming of 1.5°C*.

M. Koleva, O. J. Guerra, J, Eichman, B,Hodge, and J. r Kurtz. 2021,Optimal Design of Solar-Driven Electrolytic Hydrogen Production Systems within Electricity Markets, *Journal of Power Sources* 483:229183.

D. Mallapragada, S. Emre Gençer, P. Insinger, D. William Keith, and F. Martin O'Sullivan, 2020, ,Can Industrial-Scale Solar Hydrogen Supplied from Commodity Technologies Be Cost Competitive by 2030?, *Cell Reports Physical Science* 1(9)

Renewable, International, and Energy Agency, 2019, *HYDROGEN : A RENEWABLE*.

Schnuelle, Christian, T. Wassermann, D. Fuhrlaender, and E. Zondervan. 2020,Dynamic Hydrogen Production from PV & Wind Direct Electricity Supply – Modeling and Techno-Economic Assessment, *International Journal of Hydrogen Energy* 45(55),29938–52.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50315-8

Thermoacoustic Flow-Through Cooler for Cryogenic Hydrogen

Konstantin I. Matveev*, Jacob W. Leachman

School of Mechanical and Materials Engineering, Washington State University, Pullman, WA, 99164, USA
matveev@wsu.edu

Abstract

As hydrogen is becoming an increasingly important energy carrier for renewable energy systems, a need for efficiency improvements of hydrogen cooling and liquefaction rises as well. An additional challenge associated with these processes in cryogenic conditions is the exothermic conversion between ortho and para isomers of hydrogen, which requires removing extra heat and accelerating this reaction with catalysis. In this paper, a new system with flowing-through hydrogen and elegantly combined thermoacoustic heat pump and catalytic regenerator is analyzed using thermoacoustic theory. Calculations with variable channel sizes in a regenerator and acoustic impedances are conducted for standing-wave and travelling-wave variants of this system. The corresponding optimal second-law efficiencies of these setups are estimated to be about 0.3 and 0.6. The throughput of cooled hydrogen is assessed for systems with catalyzed and non-catalyzed regenerators.

Keywords: Hydrogen; Thermoacoustics; Cryogenics; Cooling.

1. Introduction

Hydrogen is one of the most promising fuels for the future "green" economy, as it does not produce harmful emissions when reacting with oxygen. As a liquid, hydrogen is one of the most energy-dense carriers. However, the process of cooling hydrogen at cryogenic temperatures either for subsequent liquefaction or to reduce boil-off during storage and transportation is a challenging problem, especially in small- and medium-scale systems (Rivard et al., 2019), which are needed to broaden applications of hydrogen fuel.

In this study, a novel system for cooling cryogenic hydrogen is analyzed. This system involves hydrogen flowing through a porous medium where both thermoacoustic heat pumping and conversion between hydrogen isomers are taking place (Figure 1). Acoustic oscillations can produce heat flux in fluids near solid surfaces that can be utilized for refrigeration (Swift, 2002). To achieve significant thermoacoustic cooling, porous materials with large surface-to-volume ratio are required.

A peculiar feature of cooling cryogenic hydrogen is a spin conversion of hydrogen molecules from predominantly ortho- to para-state that happens primarily below 100 K prior to condensation (Pedrow et al., 2021). When unassisted, this conversion is a very slow process. Moreover, this reaction is exothermic, thus requiring removal of additional heat. If this conversion does not happen during cooling processes, then the liquefied hydrogen evaporates more intensely during storage or transport, and thus, it will be lost. To accelerate the ortho-to-para conversion, surface catalysis can be employed, which requires porous materials to be effective. The main idea explored in this study is to utilize

a single porous matrix for both thermoacoustic heat pumping and as a catalytic bed. Such a system can be more efficient, compact and reliable than competing technologies.

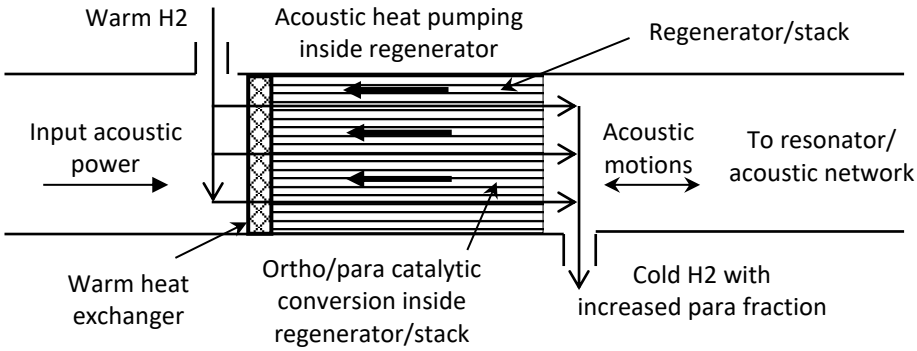


Figure 1 Schematic of thermoacoustic-catalytic system for cooling cryogenic hydrogen.

A device that accommodates both thermoacoustic and catalytic processes is shown in Figure 1. Acoustic power comes from either a linear motor or thermoacoustic engine. At appropriate phasing between acoustic pressure and velocity, heat is pumped from one side of the porous material (also called regenerator or stack) to the other side. Hydrogen entering at the warmer end cools down when flowing toward the cold end and exits the system at lower temperature. If a regenerator matrix is covered with a catalyst boosting ortho-to-para conversion, some hydrogen molecules will not only cool down but will also change their spin orientation. Hence, both a catalytic converter and a cooling device are efficiently combined in a single setup. Additional advantages include elimination of the cold heat exchanger, minimization of moving parts, and intrinsically efficient continuous heat transfer at smaller temperature differences along the stack. A few exploratory studies with flow-through thermoacoustic systems were done in the past, but not in cryogenic conditions or spin-transforming fluids (Hiller and Swift, 2000; Reid and Swift, 2000).

This paper outlines a simplified thermoacoustic model that can be used for initial assessment of the system operation in standing- and travelling-wave configurations. Amounts of hydrogen that can be cooled down are estimated in setups with catalyzed and non-catalyzed stacks. The presented model and results can benefit practitioners working on hydrogen systems and implementation of sustainable energy concepts.

2. Mathematical Model

A simplified approach to evaluate thermoacoustic heat transport in a stack or regenerator follows a theory developed for thermoacoustic devices (Swift, 2002). It is assumed that a stack contains a number of narrow channels between hot and cold ends. Other stack geometries, including random porous materials, can also be considered (Matveev, 2010), and basic thermoacoustic effects will remain similar.

Fluid inside the stack performs oscillations with the primary motions along the channels, while acoustic power comes from an external source, such as a motor or thermoacoustic engine. Due to thermal and acoustic interactions, heat can be transported along the channels. In case of large temperature gradients imposed in the stack, acoustic power can be generated. In case of relatively low temperature gradients, heat can be pumped from colder to warmer space, while acoustic power will be consumed in the process.

At the starting point of computing the heat flow and acoustic dissipation in stacks of thermoacoustic systems, one need to define acoustic pressure fluctuation p' and volumetric velocity fluctuation U'. The common notation involves complex numbers,

$$p'(x,t) = Re[p_1(x)e^{i\omega t}] \tag{1}$$

$$U'(x,t) = Re[U_1(x)e^{i\omega t}] \tag{2}$$

where p_1 and U_1 are the complex amplitudes of acoustic pressure and volumetric velocity fluctuations, i is the imaginary unity, ω is the angular frequency of oscillations, and t is the time. The x-axis is directed along the primary orientation of gas particle motions in the acoustic wave (along the stack). A relation between acoustic pressure and volumetric velocity at the stack depends on the entire system (not just stack), so that the system can be designed to achieve desirable impedances at the stack location (Matveev et al., 2006).

The heat transport rate produced by this thermoacoustic mechanism with a correction for ordinary heat conduction can be calculated as follows (Rott, 1975),

$$Q = \frac{1}{2}Re\left[p_1\tilde{U}_1\frac{\tilde{f}_v - f_k}{(1+\sigma)(1-\tilde{f}_v)}\right] + \left[\frac{\rho_m c_p|U_1|^2}{2\omega A(1-\sigma^2)|1-f_v|^2}Im(f_k + \sigma\tilde{f}_v) - (Ak + A_{sol}k_{sol})\right]\frac{\Delta T_m}{\Delta x} \tag{3}$$

where f_k and f_v are the thermoacoustic functions that depend on viscous and thermal penetration depths, $\delta_v = \sqrt{2\mu/(\omega\rho_m)}$ and $\delta_k = \sqrt{2k/(\omega\rho_m c_p)}$, and the channel thickness h; A and A_{sol} are the cross-sectional areas of the stack occupied by gas and solid, respectively; tilde indicates complex conjugate; σ is the Prandtl number; μ, ρ_m and c_p are the gas viscosity, mean density and specific heat, respectively; k and k_{sol} are the heat conductivities of gas and solid, respectively, and ΔT_m is the variation of temperature in the x-direction over distance Δx (stack length).

The pumped heat already accounts for heat conduction along the stack in Eq. (3). However, besides providing cooling power Q_C, this pumped heat must also remove acoustic power W_a dissipated in the stack, which can be estimated as follows,

$$W_d = \frac{1}{2}Re\left[\tilde{U}_1\Delta p_1 + \tilde{p}_1\Delta U_1\right] \tag{4}$$

where Δp_1 and ΔU_1 are relatively small changes of acoustic pressure and velocity amplitudes over the stack. As additional dissipation will occur in the other parts of the system (outside stack), the total acoustic power that need to be supplied is estimated as $W_a = bW_d$, where b is a given resonator loss correction. Then, the coefficient of performance and the second-law efficiency can be calculated as follows,

$$COP = \frac{Q_C}{W_a} = \frac{|Q| - W_d}{bW_d} \tag{5}$$

$$\eta_{II} = \frac{Q_C}{W_a}\frac{T_H - T_C}{T_C} \tag{6}$$

where T_H and T_C are the temperatures of the hot and cold ends of the stack, and ΔT_m in Eq. (3) is the difference between these temperatures, $T_C - T_H$.

3. Sample Results

Two examples of thermoacoustic coolers are presented below. The first setup involves a standing-wave phasing, whereas the second system is of the traveling-wave type. Their selected parameters, corresponding to typical thermoacoustic and hydrogen systems (aiming at H_2 flow rates of the order of 1 g/s), are listed in Table 1. The desired temperature drop is initially specified as 15 K. The inlet temperature of 77 K is chosen as the boiling temperature of liquid nitrogen, which permits economical cooling of hydrogen down to this temperature. The variable parameters used for optimization included a spacing distance h between plates in the stack and a magnitude of normalized acoustic impedance $|z|$ at the stack location. The normalized impedance is defined as follows,

$$|z| = \frac{p_1 A}{U_1 \rho a} \tag{7}$$

where a is the speed of sound. Additional requirements imposed in this optimization study are the cooling capacity of at least 200 W (needed for intended applications) and acoustic power input within 300 W, whereas the optimized parameter is the second-law efficiency.

Table 1. Selected system parameters.

Mean pressure	10^6 Pa	Stack length	2 cm
Warm temperature	77 K	Porosity	0.8
Cold temperature	62 K	Acoustic frequency	350 Hz
Stack plate material	steel	Acoustic pressure amplitude	$2 \cdot 10^5$ Pa
Resonator diameter	5 cm	Resonator loss correction	1.5

The calculated performance metrics for the standing-wave system, with velocity lagging pressure by about 90°, are shown in Figure 3. These metrics include the cooling power Q_C, coefficient of performance COP, and the 2nd-law efficiency η_{II}. This efficiency peaks at a certain value of the channel thickness normalized by the thermal penetration depth, h/δ_k. Among the considered variations of the plate spacing and acoustic impedance, the configuration with $h/\delta_k \approx 2.9$ and $|z| = 8$ produces the highest $\eta_{II} \approx 0.340$, while providing 207 W of cooling power and requiring 147 W of power input, thus satisfying criteria for the minimum cooling capacity and maximum acoustic power.

For the found optimal values of the stack-plate spacing and acoustic impedance, calculations have been also conducted to determine the amount of hydrogen flowing through the stack that can be cooled down to 65 K, 62 K, and 59 K, which correspond to temperature drops in the stack of 12 K, 15 K, and 18 K. Two situations were considered. First, no ortho-para conversion of hydrogen was assumed, implying a non-catalyzed stack. In the second scenario, a catalytic stack was used, and the complete conversion down to the equilibrium ortho-para ratio at the exit temperature was assumed. The results are shown in Figure 4. The dependence of the input acoustic power on flow rate of hydrogen is linear, as higher-order phenomena were neglected in this study. The larger the required temperature differential, the less hydrogen can be cooled.

The equilibrium ortho-para fraction ratio at the entrance temperature of 77 K is about 0.49, whereas this ratio decreases down to 0.39, 0.36 and 0.33 at the three exit temperatures considered here. When the maximum possible ortho-para conversion is achieved, the amounts of cooled hydrogen will be smaller (thin lines in Figure 4), since some of the cooling power has to compensate for the heat released during this conversion.

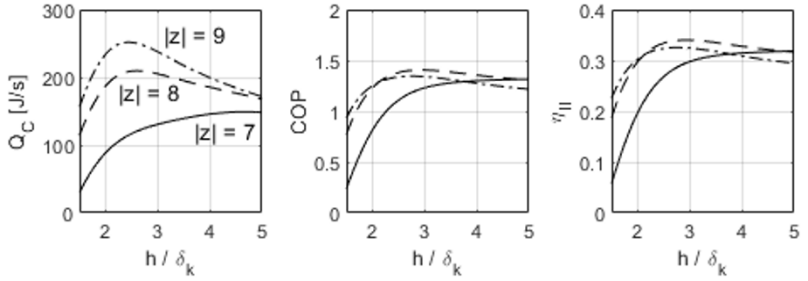


Figure 3 Characteristics of standing-wave system: cooling power Q_C, COP, and 2nd-law efficiency η_{II}. Normalized impedance $|z|$: 7, solid line; 8, dashed, 9 dash-dotted line.

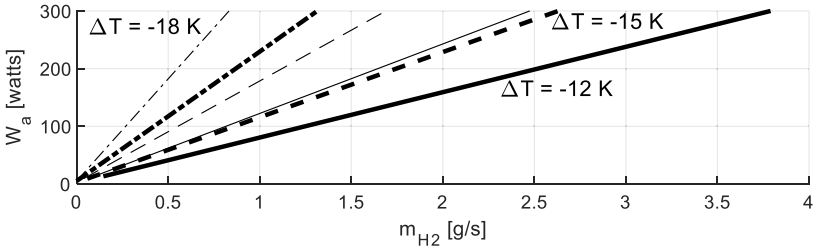


Figure 4 Acoustic power required to cool down flowing hydrogen in standing-wave system. Bold lines correspond to no para-to-ortho conversion, and thin lines to complete conversion. Temperature drops: 12 K, solid lines; 15 K, dashed; 18 K, dash-dotted lines.

Results of the optimization study conducted for the travelling-wave system, where pressure acoustic velocity and pressure fluctuations are in phase, are shown in Figure 5. Narrower channels and higher impedances are needed in such systems in comparison with standing-wave setups (Swift, 2002). The second-law efficiency for the travelling-wave reaches 0.574 for the relative plate spacing $h/\delta_k \approx 0.4$ and acoustic impedance $|z| = 40$, although the difference between peaks of the efficiency curves is small (Figure 5). The cooling power in this state is 538 W, while acoustic power of 227 W is required. One can note significantly higher efficiency is attained in the travelling-wave setup.

Results of calculations for the amount of flowing-through hydrogen in the optimized travelling-wave system are given in Figure 6 for the same temperature drops. Due to higher efficiency of this configuration, roughly twice larger flow rates of the cooled hydrogen are possible. Again, in case of catalyzed regenerators enabling ortho-para conversion, the amounts of cooled hydrogen are lower (thin lines in Figure 6).

4. Conclusions

A novel approach to cool cryogenic hydrogen, involving a flow-through thermoacoustic system, has been analyzed. Using thermoacoustic theory, high-performing geometrical and acoustic parameters of this device were determined under given operational conditions. The second-law efficiencies around 0.3 and 0.6 for standing-wave and travelling-wave setups were estimated. Mass flow rates of hydrogen flowing through the system and undergoing ortho-para conversion were evaluated. The possible theoretical extensions of this study can include modeling of the entire apparatus, accounting for finite convection heat transfer rate and ortho-para conversion rate, and modification of the

model to account for condensation process to consider a possibility of hydrogen liquefaction inside a thermoacoustic system.

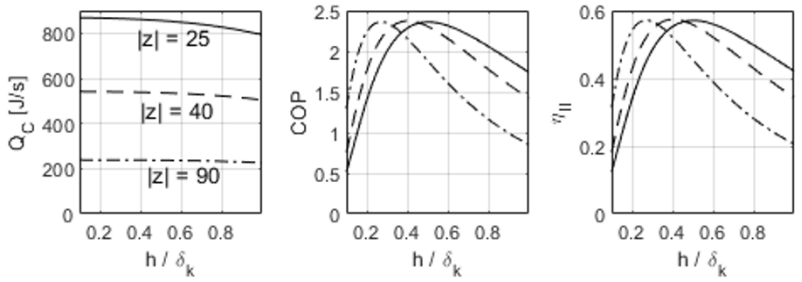


Figure 5 Characteristics of travelling-wave system: cooling power Q_C, COP, and 2nd-law efficiency η_{II}. Normalized impedance $|z|$: 25, solid line; 40, dashed; 90 dash-dotted lines.

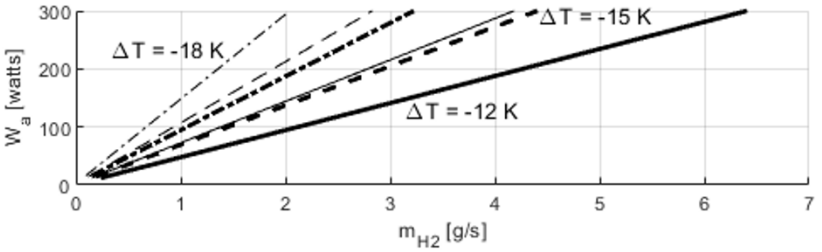


Figure 6 Acoustic power required to cool down flowing hydrogen in traveling-wave system. Bold lines correspond to no para-to-ortho conversion, and thin lines to complete conversion. Temperature drops: 12 K, solid lines; 15 K, dashed; 18 K, dash-dotted lines.

References

R.A., Hiller, G.W. Swift, 2000, Condensation in a steady-flow thermoacosutic refrigerator, Journal of the Acoustical Society of America, 108, 1521-1527.

K.I. Matveev, 2010, Thermoacoustic energy analysis of transverse-pin and tortuous stacks at large acoustic displacements, International Journal of Thermal Sciences, 49, 1019-1025.

K.I. Matveev, G.W. Swift, S. Backhaus, 2006, Temperatures near the interface between an ideal heat exchanger and a thermal buffer tube or pulse tube, International Journal of Heat and Mass Transfer, 40, 868-878.

B. P. Pedrow, S.K. Muniyal Krishna, E.D. Shoemake, J.W. Leachman, K.I. Matveev, 2021, Parahydrogen-orthohydrogen conversion on catalyst-loaded scrim for vapor-cooled shielding of cryogenic storage vessels, Journal of Thermophysics and Heat Transfer, 35, 142-151.

R.S. Reid, G.W. Swift, 2000, Experiments with a flow-through thermoacosutic refrigerator, Journal of the Acoustical Society of America, 108, 2835-2842.

E. Rivard, M. Trudeau, K. Zaghib, 2019, Hydrogen storage for mobility: a review, Materials, 12, 1973.

N. Rott, 1975, Thermally driven acoustic oscillations, part III: second-order heat flux, Z. Angew. Math. Phys., 26, 43-49.

G.W. Swift, 2002, Thermoacoustics: a Unifying Perspective for Some Engines and Refrigerators, Acoustical Society of America, Melville, NY, USA.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50316-X

Life Cycle Assessment of Green Hydrogen Transportation and Distribution Pathways

Malik Sajawal Akhtar, J. Jay Liu*

Department of Chemical Engineering, Pukyong National University, 48513 Busan, South Korea
jayliu@pknu.ac.kr

Abstract

The production of green hydrogen from renewable energy produced from wind and solar resources is deemed a more promising solution due to high energy quality, comparatively easy storage compared to electricity, and the prospect of using it at the time of use. Hydrogen has increasingly emerged as a potential energy carrier, making a global hydrogen mobility infrastructure essential to accelerating the transition to a hydrogen economy. Therefore, this work presents a cradle-to-gate life cycle assessment (LCA) for four hydrogen delivery pathways: compressed gas via tube trailers (CGH_2-TT), liquid hydrogen (LH_2), liquid organic hydrogen carrier (LOHC), liquid ammonia (LNH_3). The LCA results depict that for short distance of 100 km CGH_2-TT is the most eco-friendly option with the lowest global warming potential (GWP) of 1.81 $kgCO_2$-eq/kgH_2. Whereas, the LOHC pathway has shown the worst results with the highest GWP of 3.58 $kgCO_2$-eq/kgH_2. Likely, delivery via LNH_3 also showed significant emissions of 3.14 $kgCO_2$-eq/kgH_2 and remained the second worst candidate or hydrogen delivery.

Keywords: Green hydrogen, Life cycle assessment, Hydrogen transportation and distribution, Liquid organic hydrogen carriers, Green ammonia.

1. Introduction

The intermittency of solar and wind power results in the overproduction of electricity at times or in less production than needed at other times. Therefore, it is deemed important to store the overproduced electricity and use it when needed. Large scale energy mobility infrastructure is other challenge that needs to be addressed by finding the most sustainable energy carrier. Hydrogen owing to its high gravimetric energy density (120 MJ/kg) is being considered as a game changer. However, its low volumetric energy density hinders its mobility on a large scale and long distances (Akhtar and Liu, 2021a). Hydrogen can be stored as gaseous state as a compressed gas in high-pressure vessels or medium pressure pipelines, as a liquid state in cryogenic tanks, and as a liquid state in material based hydrogen storage (e.g., chemical hydrides and metal hydrides) (Abdin et al., 2020; Niermann et al., 2019; Wan et al., 2021).

So far, high pressure compressed hydrogen gas (CGH_2) and liquid hydrogen (LH_2) are the most hydrogen storage forms. However, both methods are not efficient on economic perspective depending on the current technological conditions, as CGH_2 storage uses around 15% of the stored energy of hydrogen to achieve 700 bar compression and LH_2 uses as much as 30% for the process of liquefaction (based on lower heating value of 120 MJ/kg) (Felderhoff et al., 2007). Liquid organic hydrogen carriers (LOHCs), a material-based hydrogen storage, are gaining much importance owing to easy handling,

transportation, and no CO_2 emissions during hydrogenation or dehydrogenation.(Aakko-Saksa et al., 2018; Niermann et al., 2019; Preuster et al., 2017). LOHCs have hydrogen storage capacities of 6%–8% and can generally store hydrogen at atmospheric pressure and temperature.(Aakko-Saksa et al., 2018). Hydrogen storage and transportation using LOHCs can therefore be more cost-effective and environmentally friendly than conventional methods, which require high-pressure vessels or cryogenic tanks. A second advantage of the LOHCs is that they are chemically similar to gasoline and diesel, therefore can be transported at a larger scale while using the existing infrastructure for petroleum processing and transport. Recently, hydrogen transportation via liquid carriers was studied and LOHC was declared as a favorable solution from an economic perspective (Wulf and Zapp, 2018). Moreover, in another study, LOHCs were compared with CGH_2 and LH_2 with regard to hydrogen transportation and were declared as feasible solution for hydrogen transportation compared with CGH_2 and LH_2 (Reuß et al., 2017).

Ammonia (NH_3) is also considered a potential candidate for hydrogen storage owing to its high hydrogen content of 17.8% and existing transportation infrastructure. (Akhtar and Liu, 2021b) In a recent study, Aziz et al. studied the transportation of hydrogen from Australia to Japan in the form of LOHC, LH_2, and NH_3 and concluded that NH_3 is the most cost effective solution for hydrogen transportation over long distances (Aziz et al., 2019). In another study, Akhtar and Liu presented a comparative feasibility study on NH_3 as a hydrogen carrier via a techno-economic analysis of transporting NH_3 from Australia to Korea and concluded that using imported green NH_3 is an economically viable alternative compared with the domestic production and transportation of hydrogen (Akhtar and Liu, 2021b).

It is important to note that not all economically optimal solutions are environmentally sustainable. Hydrogen supply chain has been thoroughly studied on economic perspective but a comprehensive analysis on environment sustainability perspective is still needed. Recent studies have exclusively focused on the gaseous transportation of hydrogen, while others have looked at the liquid transport, concluding the best and worst methods. However, afterward, the goal of what to do next was left as a question for research. Therefore, in this work, a comprehensive life cycle assessment (LCA) on a cradle-to-gate approach has been performed for four pathways:

1. Hydrogen delivery as highly compressed (500 bar) gas via tube trailers (CGH_2-TT).
2. Hydrogen delivery as liquid in cryogenic tanks via liquid trucks (LH_2).
3. Hydrogen delivery in dibenzyl toluene (DBT) with the natural gas-assisted dehydrogenation process (LOHC).
4. Hydrogen delivery by liquid NH_3 (LNH_3).

2. Methodology

Life cycle assessment (LCA) is a universal tool to access or quantify the environmental impacts associated with a product throughout its life cycle based on ISO 14040 and 14044 (Akhtar and Liu, 2021a). In LCA, all energy and material flow that occur during upstream, midstream, and downstream stages including recycling or disposal of the analyzed products are quantified and evaluated. Simapro 9.1.1.1 is used to evaluate the environmental impacts of the entire lifecycle of the above-mentioned hydrogen delivery pathways.1kg of hydrogen gas is used as a functional unit and CML-IA baseline V3.06

method is used for the evaluation of the environmental impacts. A case study for the city of Perth, which is located in Western Australia and has an urban market and a population of 2.1 million is presented. The four hydrogen delivery pathways as mentioned above are shown in **Fig. 1**. The hydrogen delivery network consists of (1) hydrogen production, (2) pre-treatment and storage of the hydrogen gas, (3) hydrogen gas/carrier transportation from production facility to hydrogen refueling station (HRS), (4) post-treatment of hydrogen gas, and (5) dispensing of the hydrogen gas to fuel cell vehicle. In all pathways, hydrogen is produced via alkaline water electrolysis using the electricity from an on-site wind power plant.

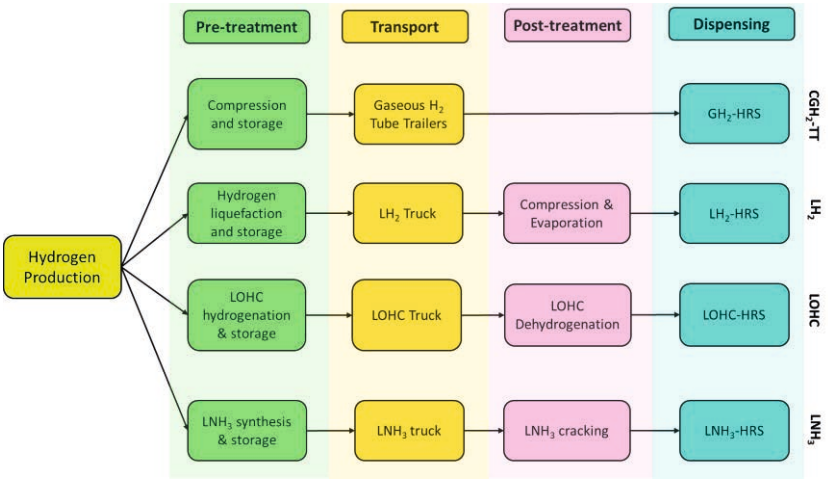


Fig. 1. Hydrogen delivery pathways.

Table 1. Main process conditions and life cycle inventory.

Process	Parameter	Units	CGH_2	LH_2	LOHC	LNH_3
Pre-treatment	Pressure	bar	500	1	1	150
	Temperature	°C	-	-253	150	450
	Electricity	kWh/kgH_2	2.2	10	0.7	4.2
Transport	Capacity per truck per trip	kgH_2	1100	4300	1800	7200
Post-treatment	Pressure	bar	-	-	350	1
	Temperature	°C	-	-	-	400
	Electricity	kWh/kgH_2	-	0.5	0.4	0.2
	Heat	kWh/kgH_2	-	-	10.5	14.3
Dispensing	Pressure	bar	700	700	700	700
	Electricity compression	kWh/kgH_2	1.6	0.5	3.96	3.45
	Electricity pre-cooling	kWh/kgH_2	4	-	4	4

At the pre-treatment stage, different processes such as compression, liquefaction, catalytic hydrogenation of LOHC, or NH_3 synthesis occur to increase the volumetric density of the hydrogen gas. Whereas, compression, pumping and evaporation, dehydrogenation of LOHC, and NH_3 cracking are the post-treatment processes occurring at HRS. For the analysis of hydrogen delivery by LOHC, dibenzyl toluene (DBT) is used, since it is the most promising hydrogen carrier.(Teichmann et al., 2012) The distance for transportation is taken as 100 km for the base case. However, in order to see the impact of distance variation, the distance is varied from 100 km to 500 km. The LCI and details of pre-treatment and post-treatment processes for all pathways is presented in **Table 1**.The annual operating hours are 8300 and the capacity of HRS is 850 kg/d (Reuß et al., 2017).

3. Results and discussion

The entire supply chain is divided into three stages: (a) *production* (hydrogen production using alkaline water electrolysis), (b) *delivery* (pre-treatment: compression, liquefaction or hydrogenation, and storage in underground salt caverns or liquid tanks; transport: via liquid carriers in trucks, pressurized transports in trucks, or by pipelines; post-treatment: evaporation or dehydrogenation at HRS), and (c) *dispensing* (operational activities like compression to bring hydrogen/NH_3 at FCV pressure, followed by filling into vehicles). **Fig. 2** illustrates that hydrogen delivery via CGH_2-TT is the most environmentally friendly option, as it resulted in the lowest contribution to global warming potential (GWP) of 1.81 $kgCO_2$-eq/kgH_2. On the contrary, hydrogen delivery via LOHC is the least environmentally friendly option with a GWP of 3.57 $kgCO_2$-eq/kgH_2. **Fig. 2** further shows that hydrogen production and hydrogen delivery are the two main stages contributing to emissions in all the pathways. For hydrogen production, the key driver of environmental emissions is the windmill-derived electricity used for electrolysis.

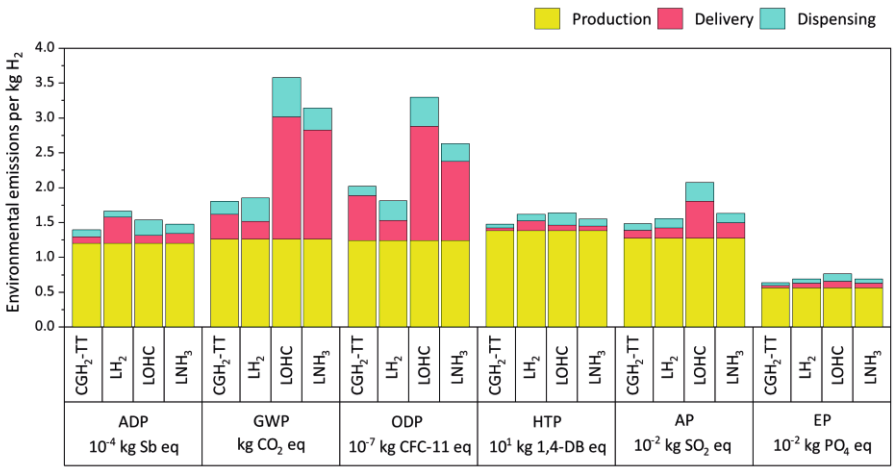


Fig. 2. Breakdown of LCA results for a transport distance of 100 km.

The delivery through LOHC presented the highest number of emissions because the post-treatment process, i.e., dehydrogenation at the HRS, is very energy-intensive, as shown in **Fig. 3(c).** Following LOHC, the delivery via LNH_3 results in the highest CO_2 emissions. A temperature of over 500°C is required for releasing hydrogen from NH_3. More than 80% of the CO_2 delivery emissions through LNH_3 are caused by the post-treatment

process of NH_3 cracking, and 99% of the 86% are produced by the energy required to achieve the required reaction temperature, whereas NH_3 synthesis only contributes 11% to the CO_2 emissions, as shown in **Fig. 3(d).**

With the increase in the distance of transportation from 100km to 400km, the highest increase of 75% is observed for transporting hydrogen via CGH2-TT with a distance increase from 100 to 400 km as shown in **Fig. 4**. With the distance increase from 100 to 400 km, hydrogen delivery via LNH_3 showed the lowest increase for transporting hydrogen by trucks compared with hydrogen delivery via LOHC since the amount of hydrogen transported per trip is much higher (7200 kg for LNH_3 trucks compared with LOHC (1800 kg)) owing to the high volumetric densities of 682 kg/m^3 for NH_3 and 57 kg/m^3 for DBT.

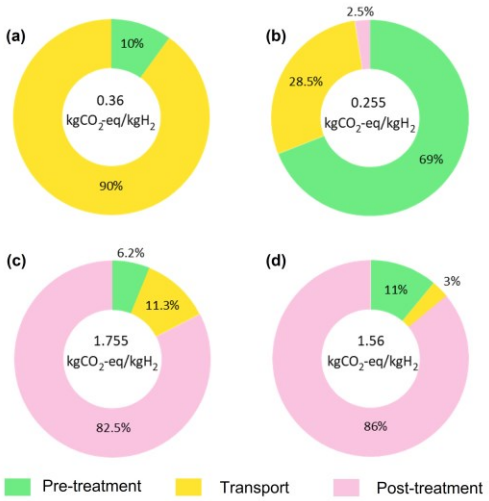


Fig. 3. Breakdown of LCA results for the impact category GWP for the delivery of hydrogen via (a) CGH_2-TT, (b) LH_2, (c) LOHC, and (d) LNH_3.

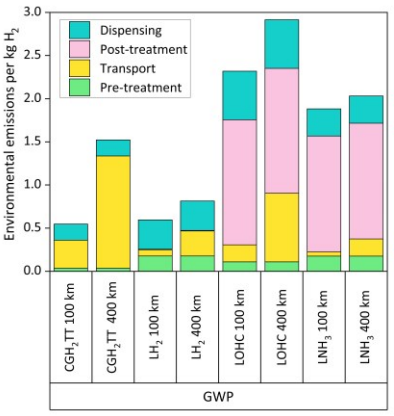


Fig. 4. Comparison of LCA results for GWP of all pathways.

4. Conclusions

In order to use hydrogen as an energy carrier for intermittent power sources, such as wind and solar, a sustainable, safe, and efficient method for storing and delivering hydrogen is needed. This paper presented a LCA for four hydrogen delivery pathways for short and long transport distances for the city of Perth, located in Western Australia. The results show that for short distances CGH_2-TT is the most responsible candidate on environmental perspective. However, with the increase in distance of transportation from 100km to 400km the highest increase in GWP is observed for CGH_2-TT. On the contrary the lowest increased is observed for the case of LHN_3 when the distance is increased to 400km. Therefore, for long distances of transportation at a larger scale, the delivery via CGH_2-TT would not be an ecologically responsible option. On the contrary, LNH_3 can play a vital role in development of a large-scale hydrogen delivery infrastructure if in future NH_3 can be directly utilized in fuel cells in order to mitigate the significant impact related to NH_3 cracking.

References

Aakko-Saksa, P.T., Cook, C., Kiviaho, J., Repo, T., 2018. Liquid organic hydrogen carriers for transportation and storing of renewable energy – Review and discussion. J. Power Sources 396, 803–823. https://doi.org/10.1016/J.JPOWSOUR.2018.04.011

Abdin, Z., Zafaranloo, A., Rafiee, A., Mérida, W., Lipiński, W., Khalilpour, K.R., 2020. Hydrogen as an energy vector. Renew. Sustain. Energy Rev. 120, 109620. https://doi.org/10.1016/J.RSER.2019.109620

Akhtar, M.S., Liu, J., 2021a. Life Cycle Assessment of Hydrogen Production from Imported Green Ammonia: A Korea Case Study. Comput. Aided Chem. Eng. 50, 147–152.

Akhtar, M.S., Liu, J., 2021b. Process Design and Techno-economic analysis of Hydrogen Production using Green Ammonia Imported from Australia- A Korea Case Study. Comput. Aided Chem. Eng. 50, 141–146. https://doi.org/10.1016/B978-0-323-88506-5.50023-1

Aziz, M., Oda, T., Kashiwagi, T., 2019. Comparison of liquid hydrogen, methylcyclohexane and ammonia on energy efficiency and economy. Energy Procedia 158, 4086–4091. https://doi.org/10.1016/j.egypro.2019.01.827

Felderhoff, M., Weidenthaler, C., Von Helmolt, R., Eberle, U., 2007. Hydrogen storage: The remaining scientific and technological challenges. Phys. Chem. Chem. Phys. 9, 2643–2653. https://doi.org/10.1039/b701563c

Niermann, M., Beckendorff, A., Kaltschmitt, M., Bonhoff, K., 2019. Liquid Organic Hydrogen Carrier (LOHC) – Assessment based on chemical and economic properties. Int. J. Hydrogen Energy 44, 6631–6654. https://doi.org/10.1016/J.IJHYDENE.2019.01.199

Preuster, P., Papp, C., Wasserscheid, P., 2017. Liquid organic hydrogen carriers (LOHCs): Toward a hydrogen-free hydrogen economy. Acc. Chem. Res. 50, 74–85. https://doi.org/10.1021/acs.accounts.6b00474

Reuß, M., Grube, T., Robinius, M., Preuster, P., Wasserscheid, P., Stolten, D., 2017. Seasonal storage and alternative carriers: A flexible hydrogen supply chain model. Appl. Energy 200, 290–302. https://doi.org/10.1016/J.APENERGY.2017.05.050

Teichmann, D., Arlt, W., Wasserscheid, P., 2012. Liquid Organic Hydrogen Carriers as an efficient vector for the transport and storage of renewable energy. Int. J. Hydrogen Energy 37, 18118–18132. https://doi.org/10.1016/J.IJHYDENE.2012.08.066

Wan, Z., Tao, Y., Shao, J., Zhang, Y., You, H., 2021. Ammonia as an effective hydrogen carrier and a clean fuel for solid oxide fuel cells. Energy Convers. Manag. 228, 113729. https://doi.org/10.1016/J.ENCONMAN.2020.113729

Wulf, C., Zapp, P., 2018. Assessment of system variations for hydrogen transport by liquid organic hydrogen carriers. Int. J. Hydrogen Energy 43, 11884–11895. https://doi.org/10.1016/j.ijhydene.2018.01.198

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50317-1

Sector coupling of green ammonia production to Australia's electricity grid

Nicholas Salmon[a], René Bañares-Alcántara[a*]

[a]*Department of Engineering, University of Oxford, Parks Road, Oxford, OX1 3PJ, UK*
[*]*rene.banares@eng.ox.ac.uk*

Abstract

'Sector coupling' of large industries to national power networks has been identified as a technique to stabilise energy systems as they electrify, and hence to reduce the cost of decarbonisation. This work explores sector coupling of green ammonia production, for which the grid can provide a stable backup power supply to reduce the costs of energy and hydrogen storage; in turn, green ammonia plants can provide renewable electricity to displace fossil fuels in the grid. We present a model which minimises the levelised cost of ammonia at 701 locations across Australia; it finds that in almost 50% of cases, paying for a grid connection reduces the cost of ammonia production, with savings of more than 10% realisable in some locations. We further show that a grid connection creates a relationship between LCOA and production scale, and improves operational stability.

Keywords: Green Fuels, Levelised Cost of Ammonia, Electricity Arbitrage, MILP

1. Introduction

Green ammonia is a derivative of green hydrogen, produced from renewable electricity, water and air. It is carbon-free, and has applications as an energy transport and storage vector, as a shipping fuel and as a fertiliser (Nayak-Luke and Bañares-Alcántara, 2020). Because of Australia's reliable renewable energy resource, large land availability, and its proximity to large future markets in East Asia, green ammonia production represents a significant economic opportunity (Srinivasan et al., 2019).

In this work, we explore how a further strategic advantage – namely, Australia's high reliability electricity network – can further improve the economic potential of green ammonia production. This coupling of a large export industry to the national electricity system has been identified as a promising opportunity to decarbonise affordably without threatening the stability of energy systems (Bloomberg New Energy Finance, 2020).

We extend on the work of previous authors who have optimised green ammonia plant designs (Fasihi et al., 2021, Nayak-Luke and Bañares-Alcántara, 2020) by developing a MILP model which includes the opportunity for a grid connection (if that connection reduces overall costs), which we solve using the Gurobi optimisation solver. Some non-linear optimisation techniques, such as a brute force calculation and genetic algorithms, are considered in the literature; however, because they take much longer to converge, and do not significantly improve solution accuracy, these techniques are not appropriate for this application.

This work further considers the relationship between a grid connection and the optimum plant scale, as well as exploring the modifications which might be required for plant design in order to achieve stable operation with and without grid connectivity.

2. Methodology

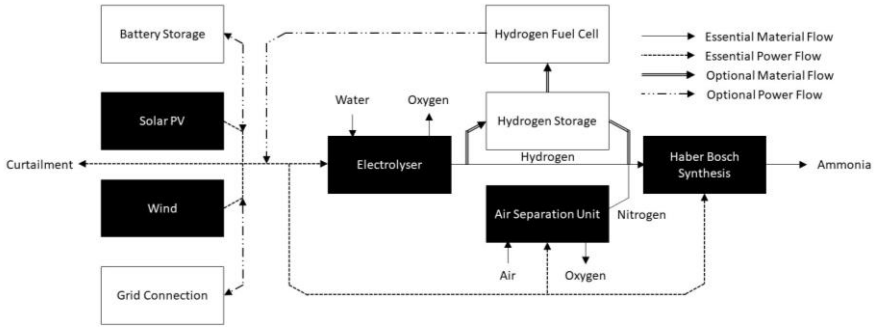


Figure 1 - Green ammonia production flowchart.

Figure 1 shows the green ammonia production process from renewable electricity. As described in detail in other works, green ammonia production ceases to benefit from economies of scale at a production rate of around 1 MMTPA (Salmon et al., 2021). This eliminates nonlinearities, meaning MILP is suitable to solve this problem.

In design mode, the solver uses location-specific weather and grid data to minimise the levelized cost of ammonia (LCOA) production by deciding (i) whether the plant should invest in a grid connection; (ii) the size in MW (or MWh for storage equipment) of the non-grid units shown on Figure 1; and (iii) the plant's operating behaviour. The design is subject to physical (i.e. mass and energy balances over each unit) and technical (i.e. minimum operating rates and maximum ramp rates of the HB plant) constraints.

The capital cost of grid connection is estimated based on the distance between the site and the nearest electricity transmission line. AC connections are used for distances of <40 km; HVDC connections are used for larger distances. The plant can import electricity using the live power price for the state in which it connects to the grid, and export excess electricity back to the grid. Imported electricity carries a transmission usage cost estimated to be 10 AUD/MWh. Costs of Australian equipment were estimated using the IRENA database (for renewable energy/electrolysers); CSIRO estimates (for grid connections/battery storage) and data from Nayak-Luke and Bañares-Alcántara (2020) (for the ammonia plant/air separation unit). Operating costs are calculated from the power withdrawn from the electricity grid, water consumption (estimated at 2 USD/kL), and an operations and maintenance (O&M) fraction of 2%. The LCOA is estimated from the CAPEX and OPEX using a discount rate of 7%.

In operation mode, the model takes as inputs both the plant design and a different year of weather/grid electricity data, and selects operating conditions which maximise cash flow. As an approximation, the costs of water and O&M are neglected, and the cash flow is calculated simply as ammonia sales minus net electricity cost. The sale price of ammonia is estimated as 500 USD/t, which is at the higher end of spot prices from the last decade.

The design problems have 157,684 constraints and 122,650 variables; all but one of these variables are continuous, the binary variable being used to decide if a grid connection is suitable. The root relaxation and node relaxation are solved using the concurrent and barrier algorithms respectively, since these were found to converge the fastest.

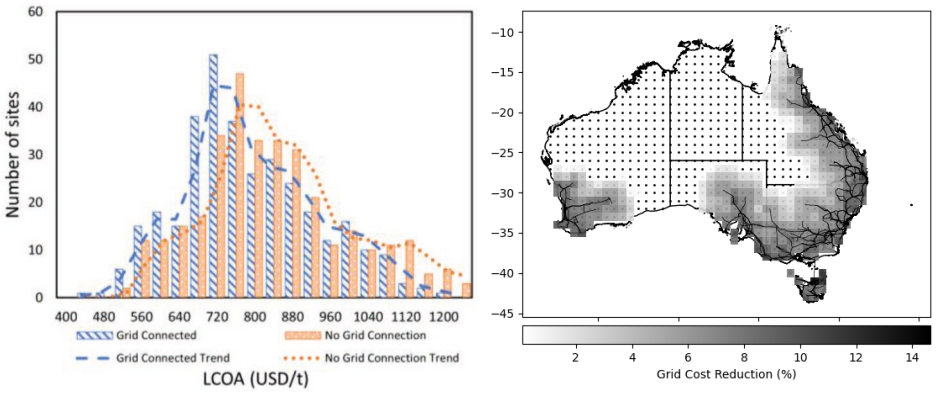


Figure 2 - Impacts of grid connection compared to a no grid case. (a) - Left: Distribution of LCOAs with grid connection compared to the same sites without grid connection. (b) - Cost reduction achieved from grid connection (Sites with a black dot and no shading did not connect to the grid).

3. Grid Connectivity Results

Over 701 cases, the average time taken for the model to solve on an i7 processor with 8 cores and 16 GB of RAM was 70 s/problem; it solves almost twice as quickly if the binary variable is fixed to a value of 0 (i.e. the grid connection is disallowed). In all cases, the model solved to within a tolerance of 10^{-4}.

The benefits of connecting to the grid are meaningful in many locations. Figure 2 (a) shows the cost distribution shifts to the left when a grid connection is included; on average, this reduction was 6%; however, as Figure 2 (b) demonstrates, there is a correlation between the distance of the site from the grid and the cost benefit. Less than 50 km from the grid, the cost reduction averages more than 8%. The best site, in Tasmania, is cheapest in both grid and non-grid cases; it sees a cost reduction from grid connection of almost 60 USD/t. Some sites become active, bi-directional participants in the market: at 63 sites, the revenue from power sales exceeds the cost of power purchased. The main cause of cost reduction is the replacement of energy storage equipment. No site with a grid connection requires battery storage or hydrogen fuel cells, and the hydrogen storage equipment reduces in size by ~30% compared to a no-grid case. Figure 2 (b) shows that sites which are grid connected are more concentrated near population centres on the coastline. This geographical spread could enable both (a) power generation at non-coastal sites that is transmitted through the grid; and (b) access to infrastructure, including a skilled workforce, spare equipment, water for desalination, and ports for export.

4. Impacts of Scale

Without grid connection, the production LCOA at industrial scale has limited dependence on production rate; while downstream infrastructure may benefit from economies of scale, most equipment required for ammonia production is modular. However, including a grid connection introduces scale dependency. At small scales (<0.1 MMTPA), the fixed cost of grid connection is not worthwhile; at large scales (i.e. ~10 MMTPA), a grid connection will not be sufficient to supply a meaningful amount of electricity to the ammonia plant without impacting the local network (meaning the grid connection is small relative to overall plant size). Between these extremes lies a minimum production cost, at which the plant makes maximum use of its investment in the electricity grid.

The precise value of this minimum production cost depends on the distance of the plant from the electricity grid, the extent to which the plant makes use of the grid connection, and the size of grid connection available. For this analysis, it was assumed that the maximum size of the grid connection was ~175 MW, which is equivalent to a large LVAC connection, and equivalent to ~10% of the average demand in Australia's smaller states (meaning larger sizes are very unlikely to be suitable). At small scales, the per unit cost of the Haber-Bosch plant increases. To factor this effect into the analysis, the Haber-Bosch plant size was estimated using a load factor of 80% before optimisation to estimate its per-unit cost (which is required for the linear optimisation problem).

The carbon intensity of the local grid must also be considered; if grid electricity represents too large a fraction of the input electricity, the ammonia may not be considered 'green'. Because of its limited use, there is no common standard at present for the maximum carbon intensity of ammonia to be considered 'green'; one European model, CertifHy, specifies a maximum of 36 g CO_2-e/MJ of fuel for hydrogen; that limit is marked on Figure 3 (b). For carbon accounting, we here assume that green electricity sold onto the electricity grid displaces fossil fuels, and therefore counts as a carbon credit.

Figure 3 shows the impact of scale on production in two locations. At the first location, which uses more grid electricity than it exports, the minimum production cost corresponds to a high carbon intensity; to reduce these emissions, a larger plant scale is required at which the grid provides less of the power, slightly increasing costs. On the other hand, at the second location in Tasmania (the cheapest identified for 2019 data), which sells more electricity than it purchases, the ammonia is carbon negative at all scales. At this location, it is profitable to sell electricity from the renewable energy production; that sale effectively subsidises the cost of green ammonia, and therefore benefits from very small scales – the larger the production rate, the smaller the subsidy from electricity export per ton.

The different behaviour at the two locations is caused both by different weather patterns, and different electricity grids; Western Australia has a stable, carbon intensive grid powered dominantly by gas; Tasmania's grid uses mostly hydropower and wind, enabling more opportunity for cost arbitrage and low-carbon grid connection.

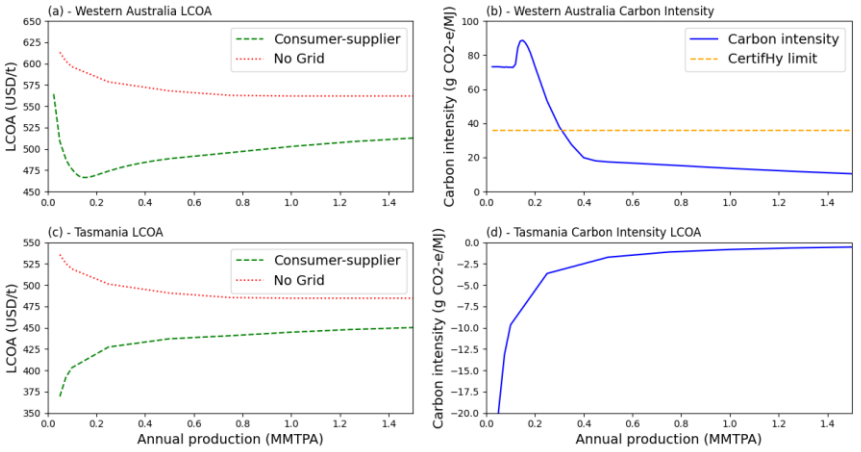


Figure 3 - Impact of scale on LCOA and carbon intensity at two low-cost sites, one in Western Australia (top), and the other in Tasmania (bottom).

5. Operating Considerations

Because the plant is optimised to minimise the LCOA, the plant (grid-connected or not) may not operate as efficiently in different weather conditions than in those under which it was designed. To some extent, grid-connected sites can use back-up power to maintain stable operation, but doing so increases costs, so electricity inputs should be minimised. Non-grid connected (islanded) sites rely on energy and hydrogen storage when they cannot generate power; if storage is too small, there is a risk of system failure (shutdown).

The main operating challenge is the requirement of the Haber-Bosch (HB) plant to operate above a minimum rate. Ambitious estimates put this minimum rate at around 20% of rated capacity. If the designed plant cannot maintain this rate for a given weather profile, regardless of how it is operated, the operating model will fail to converge, which occurs for many grid-connected and islanded cases. To reduce the likelihood of failure, the plant was overdesigned by imposing increasingly tight restrictions on the green ammonia minimum rate (i.e. > 20%) (in design mode only), but allowed to operate at the most flexible minimum rate (i.e. = 20%); this reduces failure frequency, but increases costs.

Two sets of islanded sites are compared to grid-connected sites: The Islanded (I) set refers to sites in the same location as grid-connected sites at which the model was re-run without the grid; the Islanded (II) set refers to different sites where grid connection is not optimal, whether or not it is allowed. At islanded sites, costs are mostly capitalised; at grid-connected sites, electricity costs are operational, and therefore impact cash flow. For fair comparison between sets, we report the "Cash flow delta", which is given by the cash flow in the operating year minus the cash flow that was anticipated in the design year.

Both grid-connected and islanded sites require overdesign to reduce the plant failure risk, but grid-connected sites still outperform islanded sites on two fronts. Firstly, the cost to overdesign is higher at both sets of islanded sites than at grid-connected sites – see Figure 4 (a). Secondly, while imposing stricter overdesign requirements reduces the failure rate at all sites, there are fewer failures at grid connected sites than at either set of islanded sites – see Figure 4 (b). Figure 4 (c) also indicates that sites are more likely to connect to the grid during the design process if the constraints imposed on the HB plant are tighter.

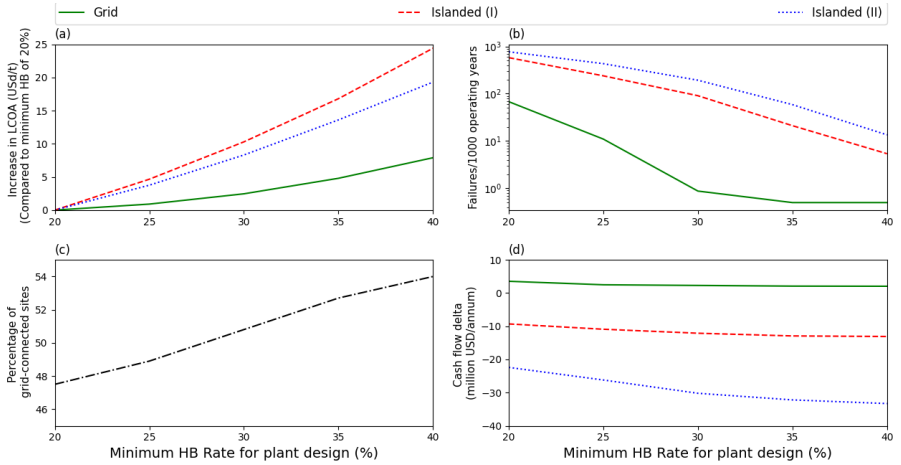


Figure 4 - Plots showing operating performance of the three sets of sites at different minimum HB rates in plant design. No operating failures were recorded for grid sites with minimum rates of 35% or greater, which is recorded as a rate of 0.5/thousand years so it can be read on a log scale.

With no overdesign, the average cash flow delta overall is below 0, which indicates that performance is generally worse during operation than was anticipated during system design, because the plant is not optimised for the new weather or grid electricity profile. However, grid-connected sites on average had a cash flow delta slightly greater than 0, even though the operating timeframe included both years in which the grid was cheaper and years in which it was more expensive than the design year. The cash-flow delta worsens for islanded sites as the plant is more overdesigned, while for grid-connected sites it stays relatively constant: this is an artefact of cases which were previously non-converging being factored into the averages.

6. Conclusions

This research explores the benefits of sector coupling for green ammonia production. It demonstrates that significant reductions in the LCOA are achievable using a grid connection. The most substantial cost reductions, which are in the order of 10%, occur when the plant is located near the electricity grid. In Australia, these sites are mostly coastal, which will locate them close to other supporting industry and to export ports. Further cost reductions are achievable by optimising the plant scale relative to the maximum allowable size of the grid connection. In some locations, ammonia production can be significantly subsidised by profitable participation in the grid.

When a site connects to the grid, it is less likely to fail during operation, and will generate more cash flow than if a grid connection is not used. Regardless of whether a site is grid-connected, it requires some overdesign, which can be achieved by designing with tighter limitations on the minimum rate of the HB plant than are achievable during operation; the cost of overdesign at grid-connected sites is less than at islanded sites.

The integration of optimised green ammonia production and grid electricity is a first step in understanding how sector-coupling in the energy system can reduce the costs of decarbonisation. Further research should consider other industries which may have synergies with electricity grids, and how electricity grids themselves will transform over time. Additionally, the operating model demonstrated that there is a risk of plant failure caused by a shortage of back-up power or hydrogen that can occur whether connected to the grid or not; further research is required to understand how ammonia plants will be operated with imperfect weather forecasting information.

References

BLOOMBERG NEW ENERGY FINANCE 2020. Sector Coupling in Europe: Powering Decarbonization.

FASIHI, M., WEISS, R., SAVOLAINEN, J. & BREYER, C. 2021. Global potential of green ammonia based on hybrid PV-wind power plants. *Applied Energy,* 294**,** 116170.

NAYAK-LUKE, R. & BAÑARES-ALCÁNTARA, R. 2020. Techno-economic viability of islanded green ammonia as a carbon-free energy vector and as a substitute for conventional production. *Energy & Environmental Science,* 13**,** 2957-2966.

SALMON, N., BAÑARES-ALCÁNTARA, R. & NAYAK-LUKE, R. 2021. Optimization of green ammonia distribution systems for intercontinental energy transport. *iScience,* 24**,** 102903.

SRINIVASAN, V., TEMMINGHOFF, M., CHANRNOCK, S. & HARTLEY, P. 2019. *Hydrogen Research, Development and Demonstration: Priorities and Opportunities for Australia, CSIRO* [Online]. Available: https://www.csiro.au/en/Do-business/Futures/Reports/Energy-and-Resources/Hydrogen-Research [Accessed December 2020].

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50318-3

Development of Multi-Purpose Dynamic Physical Model of Fuel Cell System

Shigeki Hasegawa[a*,b], Shun Matsumoto[b], Yoshihiro Ikogi[b],
Tsuyoshi Takahashi[b], Sanghong Kim[c], Miho Kageyama[a], and Motoaki Kawase[a]

[a] *Department of Chemical Engineering, Kyoto University, Kyoto 615-8510, Japan*
[b] *Commercial ZEV Product Development Div. Toyota Motor Corporation, Aichi 471-8571, Japan*
[c] *Department of Applied Physics and Chemical Engineering, Tokyo University of Agriculture and Technology, Tokyo 184-8588, Japan*
s.hasegawa@cheme.kyoto-u.ac.jp

Abstract

1-dimentional (1D) physical modeling methods of the fuel cell (FC) system including the FC stack, air system, H_2 system, and cooling system were investigated. To ensure the simulation of life-long system operation (> 100,000 km) of a vehicle with the allowable calculation time, the proper model resolution was selected and the in-house high-speed numerical solvers were developed. The acceptable accuracy was confirmed by the comparison between the model outputs and the actual FC-system data collected with 2nd-generation MIRAI under a variety of operating condition.

Keywords: Fuel cell system; Model-based development; Physical modeling;

1. Introduction

Hydrogen energy is regarded as one of the most promising alternative energies to fossil fuels from the view of CO_2 emission and energy efficiency. The FC-system manufacturers are required to develop the products for the wide range of applications such as passenger vehicles, commercial vehicles of buses and tracks, railways, marine vessels, aviation, and stationary power generator purposes.

On the other hand, due to the complexity in the hardware and software configuration of the FC systems, neither a system model nor a systems approach for the FC-system simulation, analysis, optimization, and manufacturing has not been proposed, in spite of intensive investigation on the simulation of fuel cell itself (Weber et al., 2014). Thus a significant effort and cost of trail-and-error for the development of each commercial application is required to the FC-system manufacturers, which is one of the largest barriers of entry to the fuel cell industry. The purpose of this research is the development of a state-of-the-art FC-system model, with which the FC-system manufacturers can do the model-based design, evaluate the system, and reduce the difficulty described above.

In this study, the FC system implemented to 2nd-generation MIRAI (Takahashi and Kakeno, 2021) shown in Fig. 1, is taken as an example of the application of the developed model to describe the modeling strategies and detailed implementation methods.

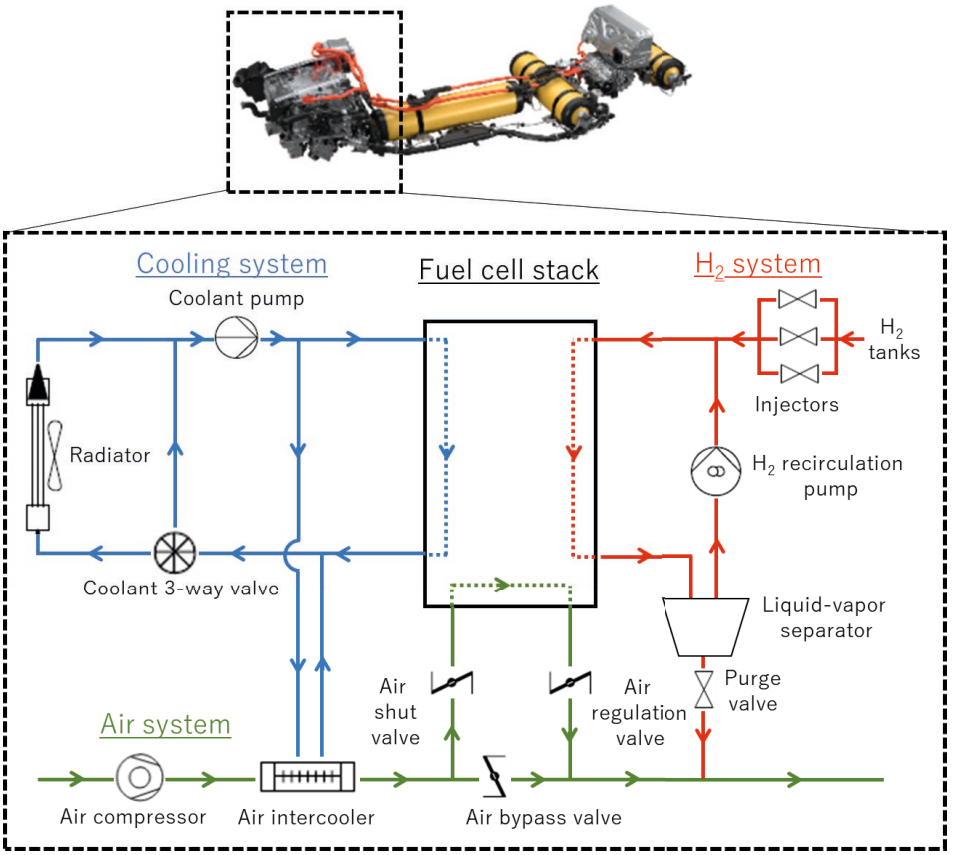


Fig. 1. Flow diagram of the FC system implemented in 2nd generation MIRAI

The FC system in Fig. 1 consists of the FC stack as main engine, air system, H_2 system and cooling system, and the components in each system and their functions are listed in Table 1.

Table 1. The components and their functions in the FC system shown in Fig. 1

System	Component	Function
Air	Air compressor	Air supply
	Air intercooler	Air cooling for the FC-stack materials protection
	Air shut-valve	Seal-up during system shut-down condition
	Air regulation valve	Pressure control, seal-up during system shut-down condition
	Air bypass valve	Control of air flowrate to the FC stack and the exhaust line
H_2	Injectors	Hydrogen supply from high-pressure H_2 tank
	Liquid-vapor separator	Separation of exhausted liquid water
	Purge valve	Exhaust of gaseous impurities (N_2,O_2, and H_2O), and liquid water
	H_2 recirculation pump	Recirculation of unreacted H_2 to inlet-side
Cooling	Coolant pump	Coolant supply
	Radiator	Heat exhaust from the coolant
	Radiator fan	Enhancement of heat exchange rate at the radiator
	Coolant 3-way valve	Control of coolant flowrate to the radiator and the bypass lines

2. Physical modeling methods of FC system

The modeling strategies and implementation methods of the FC stack and H_2 system, whose configuration is shown in Fig. 1, are explained. The system configurations are described as the function-block diagram shown in Fig. 2(b), where the overall system configuration is broken-down to the component level. In each function block, state variables (pressure, flowrate, temperature, and gas composition) and individual component models are encapsulated. In more detail, physical models of mass-transfer and electrochemistry in the FC stack are developed (Hasegawa *et al.*, 2021) and implemented in '(5) aFC' block, and physical models of fluid and thermal dynamics of various system components (Bird et al, 2006) are implemented in the other blocks.

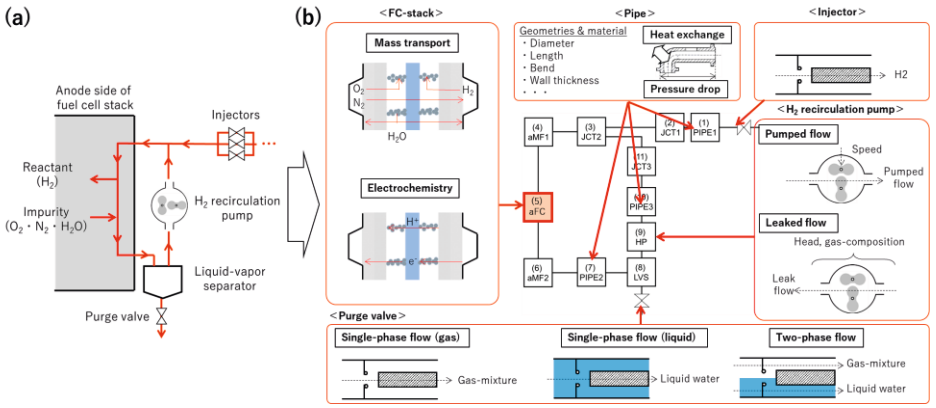


Fig. 2. Schematic drawings of the configurations of the FC stack and the H_2 system in a (a) flow-diagram and (b) function-block diagram

The dynamic relationships between state variables in each function-block are described as the algebraic equations of pressure balance, material balance, and energy balance. An example of pressure balance is shown in Eq. (1),

$$
\begin{array}{l}
(1)\ \text{PIPE1} \\
(2)\ \text{JCT1} \\
(3)\ \text{JCT2} \\
(4)\ \text{aMF1} \\
(5)\ \text{aFC} \\
(6)\ \text{aMF2} \\
(7)\ \text{PIPE2} \\
(8)\ \text{LVS} \\
(9)\ \text{HP} \\
(10)\ \text{PIPE3} \\
(11)\ \text{JCT3}
\end{array}
\begin{pmatrix}
a_{1,1} & a_{1,2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
a_{2,1} & a_{2,2} & a_{2,2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & a_{3,2} & a_{3,3} & a_{3,4} & 0 & 0 & 0 & 0 & 0 & 0 & a_{3,11} \\
0 & 0 & a_{4,3} & a_{4,4} & a_{4,5} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & a_{5,4} & a_{5,5} & a_{5,6} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & a_{6,5} & a_{6,6} & a_{6,7} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & a_{7,6} & a_{7,7} & a_{7,8} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & a_{8,7} & a_{8,8} & a_{8,9} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{9,9} & a_{9,10} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{10,9} & a_{10,10} & a_{10,11} \\
0 & 0 & a_{11,3} & 0 & 0 & 0 & 0 & 0 & 0 & a_{11,10} & a_{11,11}
\end{pmatrix}
\begin{pmatrix}
P_{\text{tot},1} \\ P_{\text{tot},2} \\ P_{\text{tot},3} \\ P_{\text{tot},4} \\ P_{\text{tot},5} \\ P_{\text{tot},6} \\ P_{\text{tot},7} \\ P_{\text{tot},8} \\ P_{\text{tot},9} \\ P_{\text{tot},10} \\ P_{\text{tot},11}
\end{pmatrix}
=
\begin{pmatrix}
b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \\ b_8 \\ b_9 \\ b_{10} \\ b_{11}
\end{pmatrix}
\qquad (1)
$$

where $P_{\mathrm{tot},i}$ is the total pressure in function-block i [Pa], $a_{i,j}$ is the (i, j) element of the coefficient matrix, and b_i is constant term of the function-block i. The coefficient matrix is sparse since only the elements of connected function blocks have non-zero values. The relationship between the function-block diagrams and the equations is considerably simple and clear. Because these equations are derived by implicit methods (Patankar, 1980), the solutions of these equations directly result in the pressure, concentration, and temperature distribution, throughout the entire FC system. To reduce the error caused by the linearization of the physical models, which is necessary to implement the physical models in the Eq. (1) expression, the numerical solver for error-convergence was implemented. Owing to the function-block-diagram modeling method and numerical implementation methods of in the linear algebraic equations, it is remarkably easy to modify, replace, add, and remove the component specifications in the proposed model. This reduces the lead-time to develop the system models for the individual applications. The other system models of air and cooling system in Fig. 1 were developed with the same strategies and integrated as an entire dynamic FC-system simulation model.

The parameters of each system component model are determined with the test data collected in the unit-testbeds before the integration of the components as an entire system. Fig. 3(a) is an example of the unit-testbed configuration, where the pump speed and valve angle are changed in the various gas composition conditions and the response of the flowrate and pressure head of the pump are measured. The dynamics of hydrogen pump is expressed by Eq. (2) (Akaike et al., 1983),

$$\dot{v} = C_1 N - \left(C_2 N C_3 \frac{\Delta P^{C_4}}{\mu^{C_5} \rho^{C_6}} \right) \tag{2}$$

where $\dot{v}$ is volumetric flowrate [m^3/s], N is rotational speed [rad/s], ΔP is pump head [Pa], μ and ρ are fluid viscosity [$\mathrm{Pa \cdot s}$] and density [kg/m^3], and C_1 - C_6 are the tuning parameters defined by pump geometries. The parameters C_1 - C_6 were determined by non-linear least-square method to fit the experimental data in Fig. 3(b). Such parameter determination procedures require only unit-testbed data. In other words, no system-testbed data, and the effort of data collection can be considerably reduced. The parameters for the other system component models were determined by such simple methodologies with the unit test data.

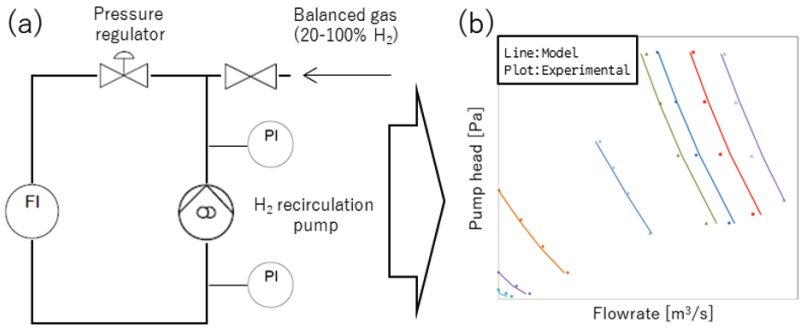


Fig. 3. (a) The unit-testbed configuration and (b) collected test data for parameter determination for H_2 recirculation pump

3. Model validation and verification

For the model validation and verification (V&V), the special test vehicles were manufactured by attaching many sensors in addition to the original ones throughout the entire FC system as shown in Fig. 4. Within the FC system, the H_2 system has a unique hardware configuration and controller specification compared with other powertrains as internal combustion engine, and it was necessary to develop special measurement instruments to collect V&V data efficiently. Moreover, ultra-compact and highly-integrated sensors were essential to collect the V&V data in the realistic dynamic vehicle operating conditions due to the limitation in the packaging space of the test vehicle. These sensors also have to be durable even in high humidity condition where large amount of liquid water exists around the sensors. To meet such requirements, H_2 concentration sensor and liquid-water level sensor were newly developed and utilized in V&V data collection by the test vehicles (Hasegawa *et al*, 2021). As described in the previous section, these data were not used for the parameter determination but only for V&V of the integrated model. When the experimental data and model output do not agree with each other, the possibilities of the missing physics or the deficiencies of parameter determination procedures were investigated repeatedly.

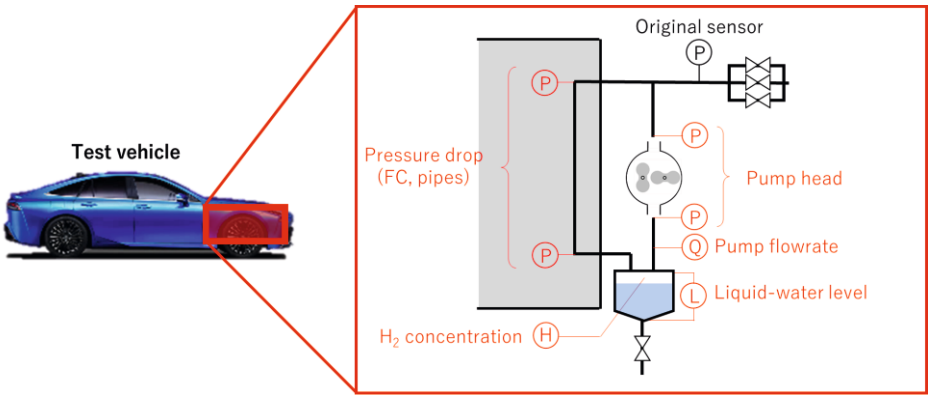


Fig. 4. Schematic drawing of the experimental configuration of data collection for the validation and verification of the developed models

Fig.5 is an example of the results of V&V of the models. A considerable amount of test data was collected by the test vehicle under a wide range of operating condition of low to high loads, operating temperatures, and atmospheric pressures. The same inputs were given to the model and the model output data were compared with the experimental data. It was confirmed that the simulation results and the measured fuel cell voltage were in good agreements within 10 % error as shown in Fig. 5 (e) and the other model outputs in Fig. 5 agreed with the experimental data, although the deviations between measured and calculated liquid-water levels in the liquid-vapor separator were observed during 300-400 s and 500-550 s in Fig. 5 (c). The deviations were not mainly caused by modeling error, but by erroneous measurements from the liquid-water level sensor in Fig. 4, whose accuracy is not guaranteed when FC-current is low.

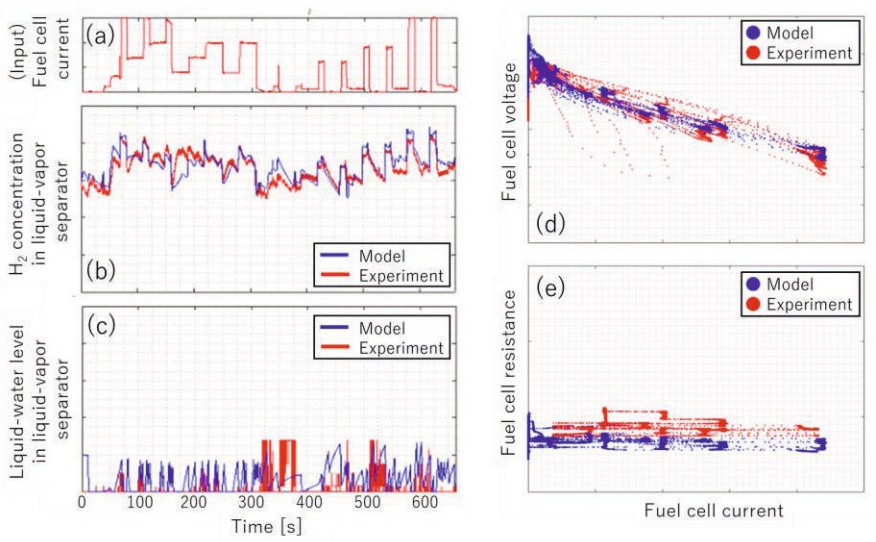


Fig. 5. An example of validation and verification results in (a) FC current as the system input, (b) H_2 concentration and (c) liquid-water level in liquid-vapor separator of H_2-system, and (d) IV and (e) IR plot of the FC stack

Conclusions

Physical modeling methods and numerical solving methods for the dynamic simulation of the entire FC system were investigated. The parameter determination procedures for each system component with unit-testbed data were proposed. The integrated model was validated and verified with a considerable amount of test data collected with the 2nd-generation MIRAI test vehicles. It is expected that the developed model can be a powerful platform of the FC-system simulation, analysis and optimization, which enables the FC-system manufacturers to investigate a wide range of specifications of the FC stack and the system components with less difficulty.

Acknowledgement

This work was supported by the FC-Platform Program: Development of design-for-purpose numerical simulators for attaining long life and high performance project (FY 2020–2022) conducted by the New Energy and Industrial Technology Development Organization (NEDO), Japan.

References

A. Z. Weber et al., 2014, *Journal of The Electrochemical Society*, **161** (12) F1254-F1299

S. Akaike and M. Nemoto, 1983, *Mechanical Engineering Journal of Japan*, Vol.49, No.447.

S. Hasegawa, M. Kimata, Y. Ikogi, M. Kageyama, S. Kim, and M. Kawase, 2021, *ECS Transactions*, 104 (8) 3-26

S. Hasegawa *et al*., 2021, *EVTeC 2021 Proceedings*, No.B1.2.

S. V. Patankar, 1980, Numerical Heat Transfer and Fluid Flow, *McGraw–Hill, New York*

T. Takahashi and Y. Kakeno, 2021, *EVTeC 2021 Proceedings*, No.B1.1.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50319-5

Embracing the era of renewable energy: model-based analysis of the role of operational flexibility in chemical production

Thomas Knight[a], Chao Chen[a], Aidong Yang[a]

[a] *Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, UK*
aidong.yang@eng.ox.ac.uk

Abstract

The chemical industry, like other industrial sectors, is expected to embrace renewable resources as replacement for fossil fuels in the coming decades. Focusing on the special challenges arising from the use of intermittent and variable renewable energy in large-scale chemical production, this paper presents two case studies, one of methanol and another of aluminium, to explore the role of operational flexibility in mitigating the burden of energy storage requirement. Whole-system optimisation and unit-level dynamic simulation were applied in the two cases, respectively, with both results showing potential benefits of shifting process operation from a constant-load mode to one that accommodates a certain degree of flexibility.

Keywords: Renewable energy, chemical production, aluminium, methanol, operational flexibility

1. Introduction

Grand environmental challenges such as climate change, ecosystem deterioration and resource depletion have greatly re-shaped the landscape of energy supply in recent decades, with the production and use of renewable energy emerging clearly as the preferred direction. The transition to renewable energy systems will contribute greatly to the decarbonisation of the chemical industry which conventionally depends on energy and feedstock derived from fossil sources.

To date, much work has been done on incorporating renewable feedstock, in particular biomass, based on the concept of biorefinery (Kokossis and Yang, 2010). On the energy front, considerations have been given to the utilisation of renewable energy such as wind and solar power in the framework of "power-to-X" (Sternberg and Bardow, 2015), where "X" typically represents hydrogen and chemicals and materials that can be derived from hydrogen in combination of other molecules such as CO_2. Currently, most large-scale chemical production processes have been designed to operate with a relatively stable load, supported with stable energy input from either the grid or on-site generation by consuming fossil fuels. In contrast, renewable energy generation is often intermittent and variable, which requires costly energy storage to supply to chemical processes with a rigid demand profile. To overcome this barrier, it is desirable to explore the possibility of operating a chemical process with a certain degree of flexibility in its load, to reduce the energy storage requirement via a closer match between the energy supply and demand profiles. While the perspective of flexible operation has been discussed recently in the context of power to ammonia (Cheema and

Krewer, 2018), methane (Matthischke et al., 2018) and methanol (Hank et al., 2018), detailed model-based assessment of the potential of this strategy is still rather limited.

In this paper, we present two distinctive case studies to demonstrate the potential of using operational flexibility as a tool to reduce the burden of energy storage when powering large-scale chemical production with variable renewables. The first study extends our earlier work (Chen and Yang, 2021) on methanol production based on CO_2 hydrogenation, which uses an optimisation model of the combined system of energy supply, storage and chemical production to reveal the potential of a holistically optimised system enabled by different levels of operational flexibility. This study considers H_2 for both energy and feedstock storage. It focuses on the impact (i.e. "what-if"), not the realisation (i.e. "how"), of process flexibility in a complex system comprising a number of subsystems. In contrast, the second study offers a detailed, process unit-level analysis of aluminium production, considering the replacement of grid electricity with power from a wind energy facility to run the smelter, where the energy storage to be tackled is in the form of batteries. It is based on dynamic simulation of the smelter to predict (1) the extent to which flexible load can be achieved within operational constraints and (2) the corresponding impacts.

2. Case study 1: methanol production

2.1. System overview and modelling approach

As shown in Fig. 1, the system consists of electrolytic H_2 production, CO_2 capture, methanol synthesis and purification. Compressed H_2 storage and fuel cell-based H_2-to-power conversion are included to reconcile the mismatch between variable renewable energy generation and the demands for H_2 and power by methanol production, while supplementary power supply from a dispatchable source (e.g. grid) is also available. In terms of operational flexibility, the electrolyser was assumed to be fully flexible (i.e. with no restrictions with respect to the minimum load). In the methanol production subsystem, the methanol synthesis reactor was assumed to be flexible to a certain degree (specified by a minimum load), while the other components, namely CO_2 capture (by an amine-based absorption-desorption cycle) and methanol purification (by distillation) were assumed to operate at a constant load, thus giving rise to the potential need for CO_2 storage and raw methanol storage. Further details of the system can be found in Chen and Yang (2021).

For a specified minimum load of the methanol synthesis reactor and a targeted level of annual methanol production rate, an optimisation model developed in our previous work (Chen and Yang, 2021) can be used to identify the design and operational decisions leading to the minimisation of the levelized methanol production cost (L_{MeOH}), which comprises the cost for power supply and all the components for conversion and storage of energy and chemicals. A key factor to be explored is the trade-off between the oversizing of the methanol reactor and the reduction of H_2 storage requirements, varying with the assumed degree of flexibility of the synthesis reactor. Our previous study investigated two specific geographical locations, namely Norderny in Germany and Kramer Junction in the US, which are known to have superior wind and solar energy sources, respectively. The optimisation model has now been further applied to more locations to assess the impact of process flexibility under a wider range of

conditions, including particularly an area near Tokyo (latitude 35.75; longitude 140.75) for which the results are presented below for a system producing 400,000 t/y of methanol. The global horizontal radiation and wind speed data were obtained from NASA's worldwide energy resource (POWER) database; monthly averages were taken over Jul 1983 - Jun 2005 and Jan 1984 - Dec 2013 for solar and wind, respectively. Either wind, solar or a combination of the two may be utilised to power the production.

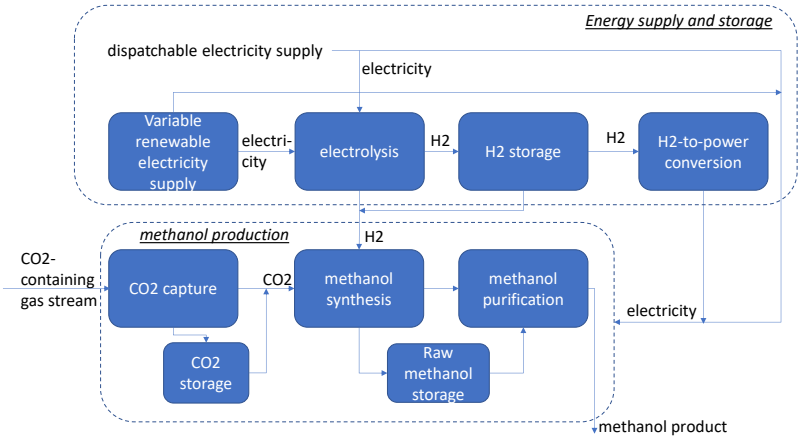


Figure 1. Methanol production based on CO_2 hydrogenation with renewable energy.

2.2. Results

As shown in Fig. 2 (left), a system with the greatest operational flexibility in the methanol reactor (with minimum load = 10%, referred to as "flexible design") could achieve a L_{MeOH} up to 12% lower than that with no flexibility in the load of the reactor (with minimum load = 100%, referred to as "nonflexible design"), which occurs at the higher end of the dispatchable energy price, a circumstance where the use of (variable) renewable energy is highly preferred. The detailed optimisation results (not shown) reveal that under such circumstance, the flexible design for a system fully powered by renewables would need to increase the annualised capital cost of the methanol reactor from ~$28m (of the nonflexible reactor) to ~$40m due to a larger reactor size needed to compensate for operating not always at its full load, but the corresponding annualised cost of hydrogen storage would reduce dramatically from ~$46m to ~$6m, which was predicted to be economically advantageous even after adding the costs of ~$11m for intermediate storage of CO_2, raw methanol and heat (the latter is omitted in Fig. 1 for clarity) needed to accommodate flexibility of the reactor. However, compared to the cost reduction (in terms of L_{MeOH}) previously predicted of Norderny (~20%) and of Kramer Junction (~30%), the benefit of process flexibility predicted of this near-Tokyo site is lower. The discrepancy is due to the difference in the renewable energy profiles between these sites in terms of the combination of diurnal and seasonal variations which affect the relative importance of energy storage in the system for meeting a specified annual methanol output, which in turn affects the impact of process flexibility.

In addition to the reduction in L_{MeOH}, Fig. 2 (right) shows that in a cost-optimized system incorporating process flexibility, the proportion (indicating the penetration level) of renewable energy in total energy supply also increases particularly when the

dispatchable energy price is in its mid-range, which is particularly ideal from the perspective of carbon emission reduction if the dispatchable energy is generated with a high carbon intensity. Note that in this case study, the range of flexibility tested by simulation was an assumption; its operational feasibility has not been evaluated. In principle, such evaluation could be carried out using detailed dynamic simulation, as demonstrated for a different process in the second case study.

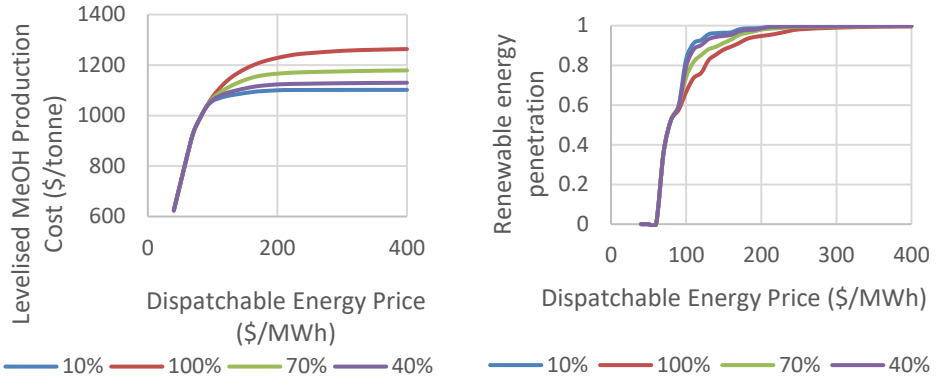


Figure 2: Impact of minimum load of methanol synthesis reactor on levelized MeOH production cost (left) and renewable energy penetration (right), plotted against dispatchable energy price. The four series indicate levels of flexibility, which is measured in terms of the minimum load of the reactor.

3. Case study 2: aluminium production

3.1. System overview and modelling approach

This second study considers aluminium production by the typically adopted Hall Héroult process, in which alumina is dissolved in molten cryolite (as the electrolyte) and electrolysed with a direct current (Fig. 3, right). The overall chemical reaction, taken into account the loss of current to side reactions, is

$$(\eta + 1)\, Al_2O_3 + 3\, C \rightarrow 2(\eta + 1)\, Al + 3\eta\, CO_2 + 3(1 - \eta)\, CO \qquad \text{Eq. (1)}$$

where η is the current efficiency. The (simplified) energy balance equation for the smelter bath is

$$m \cdot c_p \frac{dT_{bath}}{dt} = I \cdot V_{bath} - \Delta H_R - \Delta H_{feed} - Q \qquad \text{Eq. (2)}$$

where m, c_p and T_{bath} are the mass, heat capacity and temperature of the bath, respectively, I is the current, V_{bath} is the bath voltage, ΔH_R, ΔH_{feed}, and Q are the net enthalpy of all the reactions in the bath, the thermal energy required to heat and dissolve the feed alumina and the heat loss from the bath, respectively. The bath voltage has multiple components including the reversible voltage of the electrolysis reaction, resistive voltage of the bath and the bubble layer, and voltage drops at the anode and the cathode. The detailed model of these components and that of the other terms in Eq. (2), together with the detailed mass balance model and the model for estimating current efficiency, can be found in Knight (2021).

Aluminium reduction cells are operated within a very tight region which provides the best conditions for efficiency, alumina dissolution and cell life. This means the reaction

has to take place at approximately 970°C with less than 40°C tolerance. Due to these constraints, most aluminium smelters operate under a constant energy input to help control the bath conditions. To allow a smelter to run with a more flexible load, the use of a shell heat exchanger (SHE) as developed by Energia Portior (Depree *et al.*, 2016) was considered. A SHE can be installed on the outside of a cell by acting as either an insulator or conductor (hence regulating Q in Eq. (2)) to give smelter operators a new tool to aid in control of cell conditions. As part of the temperature regulation, the use of a SHE also offers extra control to facilitate the maintenance of another operational constraint, the minimum thickness of the cell's ledge (consisting of cryolite) which melts and freezes during operation.

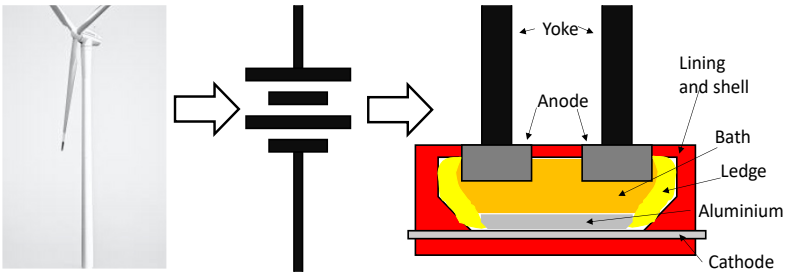


Figure 3. Structure of a system comprising an aluminium smelter (right) powered by variable renewable energy (left), mediated by a battery-based electricity storage (middle).

Dynamic simulation was performed to a system with a battery-based electricity storage to connect a wind farm and a smelter (Fig. 3). The degree of operational flexibility of the smelter was represented by a simple approach that manipulates the deviation between the current profile provided by the wind farm (I_{wind}) and that taken up by the smelter ($I_{smelter}$):

$$I_{smelter}(t) = I_{constant} - b(I_{wind}(t) - I_{constant}) \quad \text{Eq. (3)}$$

where b is an adjustable parameter ranging from 0 to 1; these two extremes represent an operation with no flexibility (where the smelter takes a constant current, $I_{constant}$, hence requiring greatest energy storage) and one completely following the current output from the wind farm (hence requiring no energy storage), respectively. The feasibility of a chosen value for b is predicted by the dynamic simulation of the smelter, against the permitted range of bath temperature and minimum ledge thickness.

3.2. Results

Table 1 summarises representative simulation results for a 24-hour operation feeding on wind power with a profile scaled from German national data (https://www.amprion.net/Netzkennzahlen/ Windenergieeinspeisung/) recorded on 5 April 2020. The bath temperature was restricted to the range of 955°C to 985°C and the minimum ledge thickness set to 5 cm (i.e. 1/3 of the initial thickness) which equates to ~2500 kg of frozen cryolite for the system considered. One can see that a "conventional" smelter (without using a SHE) can already accommodate a certain degree of load flexibility. However, the use of the SHE, thanks to its extra regulation of heat loss, allows the system to implement greater load fluctuation (reflected by a larger b value) and hence achieve greater reduction in electricity storage size. Although a more robust comparison needs to balance the storage cost with the cost of the SHE and the

slight penalty in aluminium production rate and current efficiency (shown in Table 1), the more than 25% of battery size reduction due to the use of a SHE suggests that technologies enhancing the operational flexibility of aluminium smelters may hold considerable potential.

Table 1. Results of applying different load modulation strategies to the smelter.

Current modulation	b = 0	b = 0.14 (without SHE)	b = 0.34 (with SHE)
Maximum temperature (°C)	965.9	982.3	984.9
Minimum ledge mass (kg)	3780	2685	2559
Battery size (kWh)	2599	2001	1489
Al produced (kg)	1399.6	1382.7	1359.7
Energy to produce Al (kWh/kg)	11.48	11.76	11.78
Average current efficiency (%)	92.0	90.1	90.3

4. Conclusions

Transitioning to the era of renewables calls for re-consideration of how large-scale chemical processes should be designed and operated. Through the two complementary case studies presented in this work, it is evident that increasing operational flexibility to allow load fluctuation could bring considerable benefits when feeding on variable renewable energy, in terms of reducing the electricity or chemical storage burden arising from the mismatch in the "rhythm" of energy supply and process operation. In this context, PSE tools such as detailed dynamic simulation and whole-system optimisation can play an important role, respectively, in assessing flexibility-accommodating operational feasibility of existing and adapted process units and revealing best system designs corresponding to location-specific energy supply patterns.

References

I. Cheema, U. Krewer, 2018. Operating envelope of Haber–Bosch process design for power-to-ammonia. RSC Adv 8:34926–36.

C. Chen, A. Yang, 2021. Power-to-methanol: The role of process flexibility in the integration of variable renewable energy into chemical production. Energy Conversion and Management 228: 113673

N. Depree, R., Düssel, P. Patel, T. Reek, 2016. The 'Virtual Battery' — Operating an Aluminium Smelter with Flexible Energy Input. Light Metals, 2016: 571-576.

C. Hank, S. Gelpke, A. Schnabl et al., 2018. Economics & carbon dioxide avoidance cost of methanol production based on renewable hydrogen and recycled carbon dioxide – power-to-methanol. Sustain Energy Fuels 2:1244–61.

T. Knight, 2021. The role of operational flexibility in aluminium production driven by variable renewable power. Research project report, Department of Engineering Science, University of Oxford.

A.C. Kokossis, A. Yang, 2010. On the use of systems technologies and a systematic approach for the synthesis and the design of future biorefineries, Computers & Chemical Engineering, 34 (9), 1397-1405.

S. Matthischke, S. Roensch, R. Güttel, 2018. Start-up time and load range for the methanation of carbon dioxide in a fixed-bed recycle reactor. Ind Eng Chem Res 57(18):6391–400.

A. Sternberg, A. Bardow, 2015. Power-to-What? — Environmental assessment of energy storage systems. Energy and Environmental Science, 8 (2): 389–400.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50320-1

Hollow Fiber-based Rapid Temperature Swing Adsorption (RTSA) Process for Carbon Capture from Coal-fired Power Plants

Kuan-Chen Chen[a], Jui-Yuan Lee[b], Cheng-Liang Chen[a*]
[a]*Department of Chemical Engineering, National Taiwan University, Taipei 10617 Taiwan*
[b]*Department of Chemical Engineering and Biotechnology, National Taipei University of Technology, Taipei 10608, Taiwan*
[*]*CCL@ntu.edu.tw*

Abstract

Post-combustion carbon capture is one of the feasible methods to reduce emission of carbon dioxide (CO_2) from coal-fired power plants. This work proposes a hollow fiber based rapid temperature swing adsorption (RTSA) method for capturing CO_2 from a typical 550 MW coal-fired power plant. The proposed RTSA approach can shorten the operating time and using low-grade energy for regeneration of adsorption elements.

This work studies the impact of using low-grade steam extracted from a low-pressure turbine as the heating source of the dual column vacuum RTSA (DC-vRTSA). The DC-vRTSA at 120 °C-55 °C will reduce the efficiency of the coal-fired power plant by 8.2 % to 1.9 %. The lowest CO_2 capture cost, 19.20 US dollars per tons of captured CO_2, is located at 60 °C desorption temperature.

Keywords: CO_2 capture; Coal-fired power plant; Hollow fiber; Rapid temperature swing adsorption (RTSA).

1. Introduction

1.1. Global Warming and Carbon Capture

Carbon dioxide emissions are one of the major contributors to greenhouse effects and global warming. IEA (2021) indicates that there are 31.5 Gt of CO_2 emissions worldwide in 2020 where fossil-fired power generation plays the dominant role in manmade CO_2 emissions. The carbon capture technique becomes a key to reduce the greenhouse gas effect in the near future.

In the carbon capture adsorption process, there are several available methods that can be used to desorb CO_2, such as Pressure Swing Adsorption (PSA), Temperature Swing Adsorption (TSA), and Vacuum Swing Adsorption (VSA). Haghpanah et al. (2013) presented a systematic analyses of several VSA cycles with Zeochem zeolite 13X as the adsorbent to capture CO_2 from dry flue gas. Joss et al. (2017) studied the design of TSA cycles and analyzed how individual steps within TSA cycles would affect the purity and recovery of the CO_2. Liu et al. (2019) investigate the relationship between CO_2 recovery, productivity rate, purity, specific energy consumption, and second-law efficiency based on experimental data. The effect of CO_2 concentration, desorption duration, adsorption temperature and desorption temperature has been considered in a lab-scale 4-step TSA system. However, the pressurizing cost of PSA or long heating time for regeneration of TSA is a limitation for large-scale carbon capture (Rezaei and Webley, 2010). For the

carbon capture technology mentioned above, there are many literature that compared the cost for different capture technologies. Wang et al. (2017) have reviewed the economic performance of the post-combustion CO_2 capture technologies from a coal-fired power plant, including chemical absorption and membrane-based separation.

1.2. Hollow Fiber-Based Rapid Temperature Swing Adsorption

Lively et al. (2009) proposed a novel porous hollow fiber to overcome the temperature swing limitation. The hollow fiber has a large contact area to speed up the heat transfer rate much faster than the conventional packed bed tower. Therefore hollow fiber based adsorption column is expected to realize the Rapid Temperature Swing Adsorption process. This process utilizes hollow fiber morphology to pass cooling water through the pores to maximize the adsorption capacity, and steam to pass through the pores to effectively desorb CO_2. Rezaei et al.(2014) developed a two-dimensional mathematical model of a rapid temperature swing adsorption (RTSA) process for the first time to predict polymer supported amine hollow fiber sorbent performance during post-combustion CO_2 capture from flue gas. This work focuses on manufacturing a single fiber model to simulate a four step RTSA process. Also, the sensitivity analysis to parametrric values such as gas and water velocities and initial temperatures are evaluated. Hosseini et al. (2017) developed a two-dimensional mathematical model to analyze the effects of operating variables on RTSA performance (recovery, purity, productivity, the amount of separated pure carbon dioxide in 24 h, and specific energy consumption).

However, the existing work is limited to small-scale single column simulation for the RTSA process. The cost analysis of the RTSA process has not been thoroughly discussed. This paper proposes a large-scale hollow fiber for the dual-column RTSA process to capture CO_2 from a 550 MW coal-fired power plant. In order to enhance the performance of the RTSA process, the vacuum system is used in the desorption step.

2. Process Description

The process configuration and mathematical model details are described in previous work (Chen et al., 2020). In this paper, the research focus will be economic analysis. The DC-vRTSA (Dual Column vacuum Rapid Temperature Swing Adsorption) is evaluated to a typical coal-fired power plant. The stream data of the illustrated coal-fired power plant come from Urueli (2010). Figure 1(a) shows approximate flowsheet diagram of steam from the coal-fired power plant. The high pressure water first sends to steam generator to generate high pressure steam and then electricity by pass thought the high-pressure (HP) turbine. It is then sent to steam generator again for reheating, and goes thought intermedium-pressure (IP) turbine and low-pressure (LP) turbine for generation of more electricity. The outlet of low pressure turbine is sent to condenser using cooling water to condense. Finally, it passes through condensate pump, de-aerator and feed water pump to make high pressure water than recycle to steam generator. Figure 1(b) also shows the process detail stream data of the typical power plant with the DC-vRTSA, which includes steam temperature, pressure, enthalpy, flow rate, turbine energy output, condense heat, cooling water flow rate and the power plant efficiency.

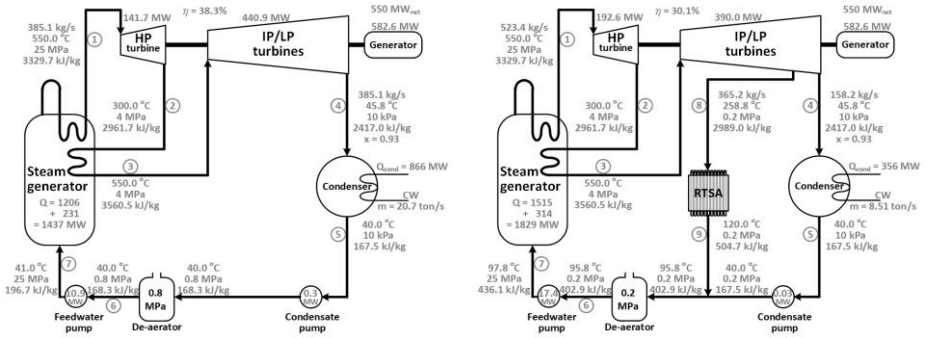


Figure 1. The illustrative coal-fired power plant (a) without carbon capture and (b) with the DC-vRTSA at 120 °C desorption temperature.

The steam extracted from the low-pressure turbine is used to provide the heat required by the DC-vRTSA process. In the first step of this simulation, stream enthalpy, temperature, pressure and DC-vRTSA desorption temperature data are given. Then saturated temperature and pressure are searched from steam table. Using saturated pressure to find the superheated temperature that should be used to provide heat for the DC-vRTSA process extracted from the low pressure turbine. Then one can find the superheated steam enthalpy. The appropriate amount of superheated steam will be extracted to provide sensible heat and latent heat for desorbing CO_2 from the DC-vRTSA process. For example, in the 120 °C desorption temperature, the extracted superheated steam temperature is 258.8 °C. This will provide the sensible heat of the superheated steam from 258.8 °C to 120 °C, then the latent heat from saturated 120 °C steam to 120 °C liquid water. Therefore the inlet is 258.8 °C superheated steam, and the outlet is 120 °C saturated liquid water. The latter will be mixed with the condensate water and returned to the boiler. Next, one makes a guess of the total steam flow rate to estimate the heat duty and the CO_2 emission (88 kg of CO_2 per GJ) from the steam generator. The DC-vRTSA is set to capture 90 % of CO_2 emission from the steam generator. Following one can calculate steam flow rate to the DC-vRTSA process and all stream data for the power plant. One calculates a new total steam flow rate which can provide 582.6 MW (including 550 MW output and total auxiliaries 32.6 MW, Zoelle et al., 2015) power and see if the new total steam flow rate equals to the guessed value afterward. If not, one uses the new one as an updated guess value and calculates again until the new value is equal to the previous one, which means the simulation is completed. Then the simulation results will be used in economic analysis.

3. Economic Analysis

For economic analysis, the detailed model of cost estimation can be found elsewhere. The desorption temperature from 120 °C to 55 °C DC-vRTSA has been used to analyze the total cost (50 °C cannot achieve 90 % purity and 90 % capture ratio).

The total capital cost (TCC) includes the RTSA material and frame, vacuum pump (for low desorption pressure). The annual capital cost (ACC) is the total capital cost divided by the designed power plant lifetime (25 years). The annual operating cost (AOC) includes the penalty, vacuum pump electricity, cooling water and RTSA material replacement. The penalty is estimated by the electricity lost in the power plant due to the

extraction of low pressure steam. For example, for the 120 ℃ desorption temperature DC-vRTSA process (see Figure 1(b)), the total steam flow rate increases from 385.1 kg/s to 523.4 kg/s. The penalty for this process is thus calculated by the difference of electricity generation between these two steam flow rates and then times the unit electricity price (0.065 US$/kWh, Ramasubramanian et al., 2012). Similarly, the cooling water cost is calculated by the amount of saved cooling water demands times unit cooling water price (0.354 US$/GJ, Turton et al., 2008).

The total annual cost (TAC) is the sum of annual capital cost and annual operating cost. Furthermore, the capture cost is estimated by TAC divided by the captured CO_2. The cost of DC-vRTSA in the 550 MW coal-fired power plant is listed in Table 1. From the table, it is found that TAC of the process is increased as desorption temperature increases, where the penalty is the biggest contributor of TAC and the higher desorption temperature leads to a higher penalty. We can invent the penalty of 120 ℃ desorption temperature process (90.67 US$/year) is about four times of 55 ℃ desorption temperature process. But the lower desorption temperature process has a lower gas volume flow rate (see Table 1). It needs more DC-vRTSA unit to capture 90 % of carbon emission. The DC-vRTSA unit will affect the DC-vRTSA price and the RTSA replacement price.

Table 1. The cost summary of DC-vRTSA in a 550 MW coal-fired power plant

Desorption temperature(℃)	120	110	100	90	80	70	60	55
Capital cost								(million US$)
DC-vRTSA unit								
Material	9.13	9.63	9.61	11.14	9.76	12.25	16.21	21.89
Frame	97.61	101.09	100.93	111.09	101.98	118.19	141.75	172.34
Total	106.74	110.72	110.54	122.23	111.74	130.44	157.96	194.23
Vacuum pump	65.96	64.22	78.48	76.87	112.83	110.09	108.38	108.33
TCC	172.70	174.94	189.02	199.10	224.57	240.53	266.34	302.56
ACC (/25 years)	6.91	7.00	7.56	7.96	8.98	9.62	10.65	12.10
Operating cost								(million US$/year)
Penalty	90.67	77.53	62.37	52.79	37.45	27.58	21.28	20.88
Electricity (vacuum pump)	16.08	15.65	19.13	18.74	27.50	26.83	26.42	26.40
Cooling water	3.22	2.73	2.18	1.82	1.23	0.83	0.58	0.57
Rplacement	1.46	1.54	1.54	1.78	1.56	1.96	2.59	3.50
AOC	111.43	97.45	85.22	75.13	67.74	57.20	50.87	51.35
TAC	118.34	104.45	92.78	83.09	76.72	66.82	61.52	63.45
Captured CO_2 (Mton/year)	3.74	3.64	3.53	3.45	3.34	3.25	3.21	3.20
CO_2 capture cost (US$/ton)	31.68	28.71	26.32	24.07	23.00	20.53	19.20	19.81
(US$/kWh)	0.0289	0.0255	0.0226	0.0203	0.0187	0.0158	0.0150	0.0154

Table 1 shows the lowest desorption temperature process (55 ℃) has the highest DC-vRTSA capital cost and replacement cost. The vacuum pump price mainly depends on the desorption pressure, lower desorption pressure leads to higher vacuum pump price and electricity demand. Also, the desorption temperature will affect the cooling water cost, too. The best total annual cost and CO_2 capture costs of these processes is 61.52 million US dollars and 19.20 US dollars per tons of captured CO_2. The capture cost that needs to be shared for each kWh also be calculated. For 120 ℃ and 55 ℃ desorption

temperature process, each kWh of electricity needs to bear 0.0289 and 0.0154 US$, respectively.

3.1. Sensitivity of Price Annualize Factor

For the economic analysis above, the price annualize factors are fixed. In this section, we will investigate how the material price of DC-vRTSA, power plant lifetime and electricity price affect the capture cost.

Figure 2(a) shows the effect of material price on the capture cost. For those process with high material area, applying high material price brings the addition in the capture cost; on the other hand, it reduces the difference of capture cost between high and low desorption temperature processes. Figure 2(b) shows the effect of power plant lifetime on capture cost. The price of electricity will mainly affect the penalty and the vacuum pump electricity in the annual operating cost. In the above economic analysis, the electricity price is 0.07 US$/kWh. Figure 2(c) illustrates how the price of electricity affects the capture cost.

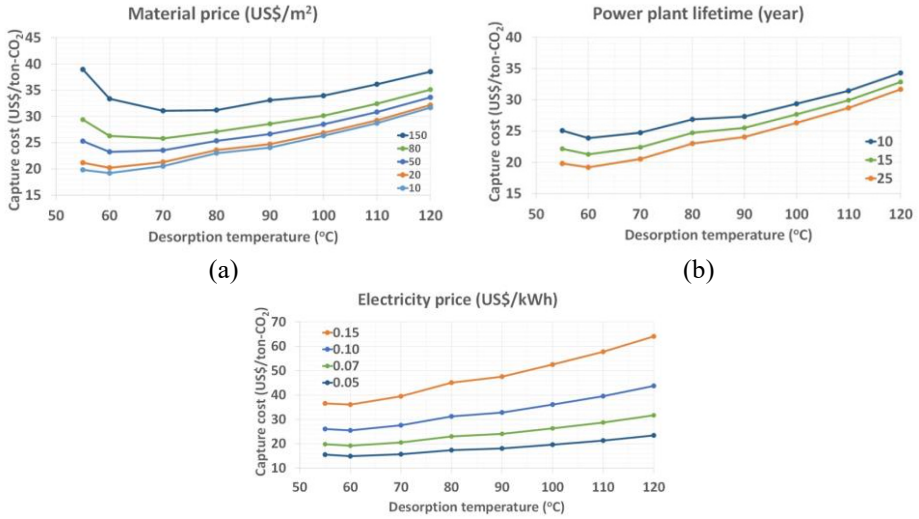


Figure 2. Sensitivity analysis of (a) material price, (b) power plant life time and (c) electricity price.

4. Conclusions

In this study, the Rapid Temperature Swing Adsorption (RTSA) process for capturing CO_2 from the coal-fired power plants has been simulated. With considering the possibility of using steam extracted from the low-pressure turbine as the heat source for the DC-vRTSA process, the impact on the efficiency and stream data of a typical coal-fired power plant were studied. The DC-vRTSA at 120 °C-55 °C reduces the efficiency of the coal-fired power plants by 8.2 % to 1.9 %. The economic analysis of the DC-vRTSA (120 °C-55 °C desorption temperature) used in the coal-fired power plant was performed. The best total annual cost and CO_2 capture costs of these processes located at 60 °C desorption temperature, which are 61.52 million US dollars and 19.20 US dollars per tons of captured CO_2. The sensitivity analysis shows that material price and power plant lifetime have significant effect on lower desorption temperature area.

References

IEA (2021), Global Energy Review: CO2 Emissions in 2020, IEA, Paris https://www.iea.org/articles/global-energy-review-co2-emissions-in-2020

R. Haghpanah, R. Nilam, A. Rajendran, S. Farooq, and I. Karimi, Cycle Synthesis and Optimization of a VSA Process for Postcombustion CO_2 Capture. Vol. 59. 2013.

L. Joss, M. Gazzani, and M. Mazzotti, Rational design of temperature swing adsorption cycles for post-combustion CO2 capture. Chemical Engineering Science, 2017. 158, 381-394.

B. Liu, Y. Lian, S. Li, S. Deng, L. Zhao, B. Chen, and D. Wang, Experimental investigation on separation and energy-efficiency performance of temperature swing adsorption system for CO2 capture. Separation and Purification Technology, 2019. 227, 115670.

F. Rezaei and P. Webley, Structured adsorbents in gas separation processes. Vol. 70. 2010. 243-256.

Y. Wang, L. Zhao, A. Otto, M. Robinius, and D. Stolten, A Review of Post-combustion CO_2 Capture Technologies from Coal-fired Power Plants. Energy Procedia, 2017. 114, 650-665.

M.R.M. Abu-Zahra, J.P.M. Niederer, P.H.M. Feron, and G.F. Versteeg, CO_2 capture from power plants: Part II. A parametric study of the economical performance based on mono-ethanolamine. International Journal of Greenhouse Gas Control, 2007. 1(2), 135-142.

A. Rao and E. Rubin, A Technical, Economic, and Environmental Assessment of Amine-Based CO_2 Capture Technology for Power Plant Greenhouse Gas Control. Environmental science & technology, 2002. 36, 4467-75.

P. Maas, N. Nauels, L. Zhao, P. Markewitz, V. Scherer, M. Michael, D. Stolten, and J.F. Hake, Energetic and economic evaluation of membrane-based carbon capture routes for power plant processes. International Journal of Greenhouse Gas Control, 2016. 44, 124-139.

R.P. Lively, R.R. Chance, B.T. Kelley, H.W. Deckman, J.H. Drese, C.W. Jones, and W.J. Koros, Hollow Fiber Adsorbents for CO2 Removal from Flue Gas. Industrial & Engineering Chemistry Research, 2009. 48(15), 7314-7324.

S. Fatemeh Hosseini, M. Reza Talaie, S. Aghamiri, M. Hasan Khademi, M. Gholami, and M. Nasr Esfahany, Mathematical modeling of rapid temperature swing adsorption; the role of influencing parameters. Separation and Purification Technology, 2017. 183, 181-193.

K.-C. Chen, J.-Y. Lee, and C.-L. Chen, Hollow fiber-based rapid temperature swing adsorption process for carbon capture from coal-fired power plants. Separation and Purification Technology, 2020. 247, 116958.

I. Urieli, Case Study - The General James M. Gavin Steam Power Plant, Chapter 8, Applied Engineering Thermodynamics,. http://www.ent.ohiou.edu/~thermo/Applied/Chapt.7_11/SteamPlant/GavinCaseStudy.html, accessed., 2010.

A. Zoelle, D. Keairns, L.L. Pinkerton, M.J. Turner, M. Woods, N. Kuehn, V. Shah, and V. Chou, Cost and Performance Baseline for Fossil Energy Plants Volume 1a: Bituminous Coal (PC) and Natural Gas to Electricity Revision 3. United States: N. p., 2015.

K. Ramasubramanian, H. Verweij, and W.S. Winston Ho, Membrane processes for carbon capture from coal-fired power plant flue gas: A modeling and cost study. Journal of Membrane Science, 2012. 421-422, 299-310.

R. Turton, R.C. Bailie, W.B. Whiting, and J.A. Shaeiwitz, Analysis, Synthesis and Design of Chemical Processes. 3rd Ed. Pearson Education, 2008.

W.L. Luyben, Comparison of extractive distillation and pressure-swing distillation for acetone/chloroform separation. Computers & Chemical Engineering, 2013. 50, 1-7.

S. Swernath, K. Searcy, F. Rezaei, Y. Labreche, R.P. Lively, M.J. Realff, and Y. Kawajiri, Chapter 10 - Optimization and Technoeconomic Analysis of Rapid Temperature Swing Adsorption Process for Carbon Capture from Coal-Fired Power Plant, in Computer Aided Chemical Engineering, Y. Fengqi, Editor. 2015, Elsevier. p. 253-278.

T.C. Merkel, H.Q. Lin, X.T. Wei, and R. Baker, Power plant post-combustion carbon dioxide capture: An opportunity for membranes. Journal of Membrane Science, 2010. 359(1-2), 126-139.

L. Zhao, E. Riensche, L. Blum, and D. Stolten, Multi-stage gas separation membrane processes used in post-combustion capture: Energetic and economic analyses. Journal of Membrane Science, 2010. 359(1-2), 160-172.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50321-3

Determining Accurate Biofuel System Outcomes: Spatially Explicit Methods for Combined Landscape-Feedstock and Supply Chain Design

Eric G. O'Neill[a,b*], Christos T. Maravelias[a,b,c]

[a]*Department of Chemical and Biological Engineering, Princeton University, Princeton, 08540, USA*
[b]*DOE Great Lakes Bioenergy Research Center, USA*
[c]*Andlinger Center for Energy and the Environment, Princeton University, Princeton, NJ 08540, USA*
eoneill@princeton.edu

Abstract

The adoption of sustainably produced second generation biofuels will rely heavily on an optimized and integrated biofuel supply chain (SC) system from field to product, and a large amount of land will need to be converted to dedicated bioenergy crops to support sufficient economies of scale. Efficient models are needed to determine the optimal upstream 'landscape design' decisions that balance trade-offs with more commonly studied SC decisions. Landscape design optimization, deciding where in the landscape to plant bioenergy crops and how to manage that land (e.g. fertilization), has been shown to improve the environmental impact of farm-scale biomass production (including soil carbon (C) sequestration), but has been studied largely separately from biofuel SC network design (SCND). In this paper we present a model for landscape design optimization and a model/data integration strategy that enables the use of both high spatial resolution crop simulations and simultaneous optimization of the downstream biofuel SC. Using crop simulations that include realistic yield and environmental data, we present an illustrative case study in Michigan, USA and highlight the benefits of the model formulation and insights from simultaneously optimizing the SC and the landscape.

Keywords: Sustainable Supply Chains, MILP, Biofuels

1. Background

In previous research related to biofuel supply chain (SC) optimization, researchers assume fixed locations and availability for biomass feedstock and do not explicitly consider design decisions related to the crop and landscape system as decision variables (Ghaderi et al., 2016; O'Neill and Maravelias, 2021). Contrary to this assumption and because dedicated bioenergy crops have yet to be planted in large quantities, decisions related to landscape design and crop management can have downstream effects on the optimal configuration and operation of the biofuel SC. There is an opportunity to extend mathematical optimization approaches for biofuel supply chain network design (SCND) to include the simultaneous optimization of the upstream landscape decisions to find sustainable and economically favorable solutions for both SC and landscape design and operation.

There are two major challenges to integrating a landscape optimization model with the biofuel SCND. First, the yield and soil carbon (C) sequestration potential has been shown

to be highly field specific which necessitates landscape optimization models with a high spatial resolution (Field et al., 2018). Second, accurately predicting the field-specific yield and soil C potential for crops like switchgrass is difficult and relies on biogeochemical crop model simulations for high-resolution data (Basso and Ritchie, 2015). A strategy for efficient model/data integration between crop simulations and optimization models is needed to preserve the computational tractability of the optimization model without sacrificing landscape design accuracy.

Accordingly, in this paper we propose a computationally efficient model formulation designed with the data in mind, and a model/data integration strategy. We then incorporate the landscape design model into a biofuel SCND model and present a case study to demonstrate the benefit from making simultaneous landscape and SC decisions at a high spatial resolution.

2. Landscape Design Model

The landscape design model makes decisions regarding the establishment, fertilization, and harvesting of biomass. Harvest decisions correspond to a multi-period ($t \in \mathbf{T}$) model to remain general to crops having multiple harvests. The primary outputs are the yield at each field and the amount of soil C that is sequestered. An overview of the integrated model and data approach is shown in Figure 1.

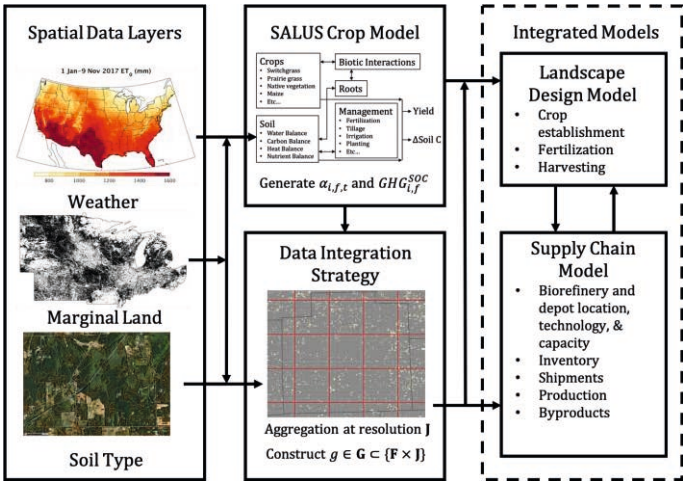


Figure 1. Overview of the integrated model and data approach. Parameter definitions are described below, and the data integration strategy is discussed in Section 3.

First, we introduce the set of fields $f \in \mathbf{F}$. The fraction of field f established with crop $i \in \mathbf{I}^F$ is controlled by decision variable $E_{i,f}$ in Eq.(1).

$$0 \leq E_{i,f} \leq 1 \tag{1}$$

To capture the effect of nitrogen (N) fertilization we also introduce continuous variable $R_{i,f} \in [0,1]$, which represents the fraction of fertilization to apply between 0 and the maximum, ω (kgN/ha). The variable $R_{i,f}$ is not used in the actual formulation, instead we construct auxiliary variable $D_{i,f} = E_{i,f}R_{i,f}$ which is used to calculate the additional yield

from fertilization and its effect on soil C in Eq.(3) and Eq.(4). Because $E_{i,f}$ is bounded by one, $D_{i,f}$ is bounded as in Eq.(2).

$$0 \leq D_{i,f} \leq E_{i,f} \tag{2}$$

The biomass yield $Y_{i,f,t}$ (Mg) of each field f is given by Eq.(4) where $\alpha_{i,f,t}$ is the potential yield (Mg) at time period $t \in \mathbf{T}$ with zero fertilization if the field had been fully planted with crop i. The second term represents the change from fertilization where $\varsigma^1_{i,f,t}$ is the additional yield (Mg) achieved from fertilization (calculated as the potential yield at full fertilization minus the potential yield at zero fertilization). The auxiliary variable is used because a fraction of the additional fertilization yield can be equivalently attained by planting or fertilizing a fractional amount. The soil C sequestration,$GHG^{SOC}_{i,f}$, is calculated similarly (Eq.(4)) with $\Gamma^{\mathrm{SOC}}_{i,f}$ and $\varsigma^2_{i,f}$ being the annualized soil C sequestered at 0 kgN/ha and the change from full fertilization respectively.

$$Y_{i,f,t} = E_{i,f}\alpha_{i,f,t} + D_{i,f}\varsigma^1_{i,f,t} \tag{3}$$

$$GHG^{SOC}_{i,f} = E_{i,f}\Gamma^{\mathrm{SOC}}_{i,f} + D_{i,f}\varsigma^2_{i,f} \tag{4}$$

The actual area of field f that is planted $A_{i,f}$ (ha) and the actual amount of fertilizer applied $F_{i,f}$ (kgN) are given by Eq.(5) and Eq.(6) respectively. Where σ_f is the full area.

$$A_{i,f} = E_{i,f}\sigma_f \tag{5}$$

$$F_{i,f} = D_{i,f}\omega\sigma_f \tag{6}$$

In a realistically sized SC, there are a very large number of potential fields. To connect the output of the landscape optimization model to a SC model while maintaining computational tractability, we consider transportation at a coarser spatial resolution represented by a set of 'harvesting sites' $j \in \mathbf{J}$. Multidimensional set $g \in \mathbf{G} \subset (\mathbf{F} \times \mathbf{J})$ describes the membership of fields to harvesting sites. The amount of harvested biomass $H_{i,j,t}$ transported to downstream SC nodes is constrained in Eq.(7).

$$H_{i,j,t} \leq \sum_{f \in \mathbf{G}_j} Y_{i,f,t} \tag{7}$$

By modelling transportation at the harvesting site level, we preserve landscape design detail but avoid adding flow variables for every field.

The costs associated with landscape design are given in Eq.(8) where λ_i are the per-Mg costs from harvesting, ρ is the cost of fertilization per kgN, and ϕ_i is the combined annualized cost of establishment and per hectare cost of land management.

$$C^{LAND} = \sum_{i,j,t} \lambda_i H_{i,j,t} + \sum_{i,f} \rho F_{i,f} + \sum_{i,f} A_{i,f}\phi_i \tag{8}$$

Similarly, the emissions from landscape activities are given by Eq.(9) where Γ_i^{MG} are the per-Mg emissions, Γ^N are the emissions from fertilization, and Γ_i^{HA} are the annualized per-hectare establishment and management emissions.

$$GHG_s^{LAND} = \sum_{i,j,t} \Gamma_i^{MG} H_{i,j,t,s} + \sum_{i,f} \Gamma^N F_{i,f} + \sum_{i,f} A_{i,f} \Gamma_i^{HA} + \sum_{i,f} GHG_{i,f}^{SOC} \quad (9)$$

For brevity, the full mixed-integer linear SC model is not presented here, but a similar model without extensions to measure SC specific GHG emissions or landscape design is presented by (Ng et al., 2018). The modified SC model is a multi-period MILP formulation that minimizes the total annualized cost. Key decisions include the location, technology, and capacity of pre-processing depots and biorefineries, inventory levels, and shipment and production schedules. GHG emissions are considered in the case study via the ϵ-constraint method and include emissions from shipment, production, landscape, and soil C sources, but do not include any credit for replacing fossil fuels.

3. Model and Data Integration

Because biomass yield and soil C sequestration potential are field specific and depend on local weather and soil quality, modelling landscape design at high resolution is critical. The model presented in section 2 uses large amounts of spatial data for parameters $\alpha_{i,f,t}$, $\Gamma_{i,f}^{SOC}$, $\varsigma_{i,f,t}^1$, and $\varsigma_{i,f}^2$ and requires a convenient way to integrate the model and the data.

Biogeochemical crop models such as SALUS (Basso and Ritchie, 2015) are tools that simulate crop growth and capture the spatial variability of crop yields and soil C sequestration. The underlying data that SALUS uses to capture spatial variability is a weather layer (gridMET (Abatzoglou, 2013)), a soil layer (SSURGO 30m resolution), and a geographic raster defining the available fields for planting biomass. Identifying potential fields is outside the scope of this paper, interested readers can refer to (Lark et al., 2020). Holding all other inputs constant, simulations with the same soil type and weather will result in the same yield and soil C sequestration. With this observation, we define the weather grid's 4x4 km pixels as harvesting sites and the soil types as fields. For a given weather grid, if there are fields with identical soil types, the fields are aggregated into a single element of **F** without losing landscape design accuracy.

The model data integration strategy allows for the aggregation of 'identical' fields which reduces the model size without aggregating unlike fields which could reduce the impact of the landscape design decisions. The model data integration strategy can also be extended to a user defined harvesting site resolution. A coarser resolution may be of interest because it can reduce the model size and enable the modelling of larger study areas. A coarser harvesting site grid can be produced by overlaying an arbitrary grid, labelling the fields enclosed in each cell with a 'grid label', then aggregating (summing yield, soil C, and field area) fields with identical grid labels and soil types.

4. Results

4.1. Case Study Description

We consider 24 potential biorefineries and 500 potential pre-processing depots in the lower peninsula of the state of Michigan, USA. SALUS simulations were performed for switchgrass grown on recently abandoned, expanded, and intermittent cropland identified by (Lark et al., 2020). There are originally 275,502 fields in the study area.

Table 1. Model size and error in yield from applying the model/data integration procedure for various harvesting site resolutions in Michigan, USA.

Harvesting Site Resolution	$\|\mathbf{F}\|$	$\|\mathbf{J}\|$	Mean error (std.) in yield (Mg/ha)
4x4 km	69,835	5,720	reference
8x8 km	41,120	1,656	0.57 (0.59)
12x12 km	29,874	794	0.58 (0.59)
16x16 km	22,801	453	0.60 (0.62)

Results from applying the data integration strategy for several harvesting site resolutions are shown in Table 1. The error in yield for each soil type is calculated as the absolute value of the yield at the 4x4 km resolution minus the aggregated yield at the coarser resolution for that same soil type. The mean and standard deviation are taken for all soils in the study area.

4.2. Landscape Design Benefit

Figure 2 shows the benefit of designing the landscape simultaneously with the supply chain. In panel (a) we use the ϵ-constraint method to find the minimum supply chain cost (Eq.(10)) while constraining the GHG emissions. Note that the GHG emissions are negative due to a sequestration of soil C which offsets the other sources of CO_2 (CO_2 equivalents from transportation, production, and land management) and does not include any credits from bioethanol replacing fossil fuels. Panels (b) and (c) correspond to the configuration of the minimum cost SC for a high and low amount of GHG emissions respectively.

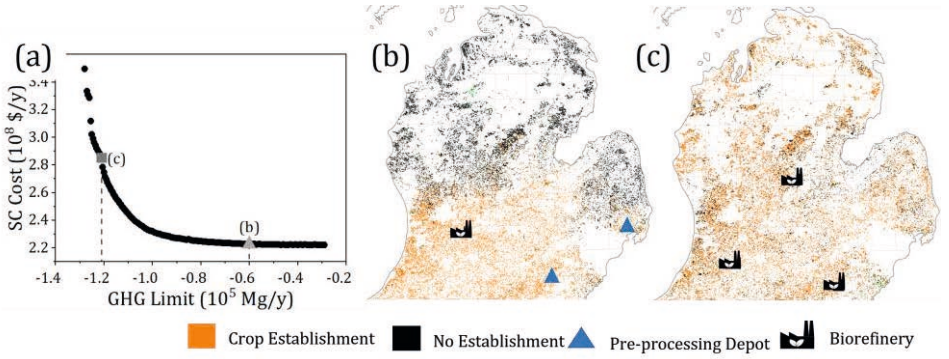


Figure 2. (a) The pareto frontier for the ϵ-constraint method constraining GHG emissions. (b) The SC configuration at a high emission level. (c) The SC configuration at a low emission level.

As the upper bound on GHG emissions decreases (becomes more restrictive), the landscape design becomes more important. Panel (c) shows that the optimal configuration avoids planting biomass at fields which are 'poor' at sequestering soil C and instead transports biomass further for the added benefit of sequestering carbon at more distant fields, opposed to panel (b) where biomass is planted close to the refinery to lower transportation costs. Furthermore, at lower GHG solutions, additional biorefineries are constructed to reduce the emissions from transportation, and depots are avoided which have a GHG penalty from using local grid energy to pelletize biomass. Interestingly, by designing the SC and landscape simultaneously, significant reductions in GHG emissions are possible for only a marginal increase in costs as shown by the flat region in the bottom right of Figure 2(a).

5. Conclusions

The high-resolution landscape design model described in section 2 allows the incorporation of the data integration strategy and results in a computationally efficient way to consider landscape design and management simultaneously with SC optimization. The large amount of spatial data needed for modelling dedicated bioenergy crops is aggregated in a way that maintains the separation of un-like fields (allowing the model greater control over the outcomes from landscape design) while ensuring a tractable model. We showed that by introducing the upstream landscape design model, integrated solutions could be found that leverage crop establishment locations to find attractive GHG solutions for only a relatively small increase in SC costs. The flexible model and data integration strategy can be used to model biofuel SCs with landscape design considerations on a much larger scale. Decision makers could use the integrated model to analyze the environmental and economic trade-offs between land quality, land distribution, SC configuration, and SC operation and the influence of key parameters on the integrated system.

References

J.T. Abatzoglou, 2013. Development of gridded surface meteorological data for ecological applications and modelling. Int. J. Climatol. 33, 121–131.

B. Basso, J.T. Ritchie, 2015. Simulating Crop Growth and Biogeochemical Fluxes in Response to Land Management Using the SALUS Model, in: Hamilton, S., JE, D., Robertson, P. (Eds.), The Ecology of Agricultural Landscapes: Long Term Research on the Path to Sustainabiliy. pp. 252–274.

J.L. Field, S.G. Evans, E. Marx, M. Easter, P.R. Adler, T. Dinh, B. Willson, K. Paustian, 2018. High-resolution techno–ecological modelling of a bioenergy landscape to identify climate mitigation opportunities in cellulosic ethanol production. Nat. Energy 3, 211–219.

H. Ghaderi, M.S. Pishvaee, A. Moini, 2016. Biomass supply chain network design: An optimization-oriented review and analysis. Ind. Crops Prod.

T.J. Lark, S.A. Spawn, M. Bougie, H.K. Gibbs, 2020. Cropland expansion in the United States produces marginal yields at high costs to wildlife. Nat. Commun. 11, 1–11.

R.T.L. Ng, D. Kurniawan, H. Wang, B. Mariska, W. Wu, C.T. Maravelias, 2018. Integrated framework for designing spatially explicit biofuel supply chains. Appl. Energy 216, 116–131.

E.G. O'Neill, C.T. Maravelias, 2021. Towards integrated landscape design and biofuel supply chain optimization. Curr. Opin. Chem. Eng.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50322-5

Assessing the Environmental Potential of Hydrogen from Waste Polyethylene

Cecilia Salah[a], Selene Cobo[a], Gonzalo Guillén-Gosálbez[a*]

[a] *Institute for Chemical and Bioengineering, Department of Chemistry and Applied Biosciences, ETH Zürich, 8093 Zürich, Switzerland*
gonzalo.guillen.gosalbez@chem.ethz.ch

Abstract

In 2019, nearly 370 million tonnes of waste plastic were generated, an amount that has been steadily increasing over the years. Here we assess hydrogen production from waste polyethylene in the context of a circular economy of plastics. Based on the gasification of polyethylene waste (wPG), we performed a Life Cycle Assessment (LCA) study following the ReCiPe method. Our results show that the wPG process coupled with carbon capture and storage (CCS) performs very well environmentally relative to other H_2 production routes, outperforming steam methane reforming (SMR) with and without CCS and biomass gasification (BG) in the three endpoint impact categories.

Keywords: Hydrogen; Waste polyethylene; Circular economy; Life cycle assessment

1. Introduction

Every year, the global demand for single-use polymers increases and, with it, the generation of plastic waste. According to Geyer et al. (2017), as of 2015, 79 % of all plastic ever made had been disposed of in landfills or the environment. Although the percentage of polymer waste destined for recycling has been increasing over the years, millions of tonnes of residues are annually mismanaged globally. This continuous accumulation underlines the need for a circular economy that valorizes polymer residues.

The circular economy of plastics is based on the recycling, repurposing, refurbishing, and revalorization of the generated waste. Notably, valuable feedstocks to the chemical industry could be produced from waste polymers through chemical recycling, as pointed out by Pacheco-Lopez et al. (2021). For instance, waste polyethylene and polypropylene can be processed to recover their respective monomers through pyrolysis or to produce synthesis gas (syngas), a key feedstock for chemical production, through gasification (Saebeaa et al. 2020). While the first route would be ideally preferred due to the higher value of monomers, the second is significantly easier to implement due to the more mature gasification technologies.

Here we explore the benefits of recycling the hydrogen chemically stored in polymers via gasification of waste polyethylene (wPE). Hydrogen has attracted increasing attention as an energy carrier and low-carbon feedstock for various fields. Steam methane reforming (SMR), currently the standard and cheapest route for hydrogen production, relies on natural gas (Parkinson et al. 2019), which leads to significant carbon emissions. Alternatives with lower emissions include SMR with carbon capture and storage (CCS) and electrolysis powered by renewable energies. Based on process modeling and LCA, here we investigate whether H_2 from recycled plastics is environmentally appealing, which at present remains unclear.

2. Methods description

2.1. Process description

We consider a hydrogen production process from waste polyethylene, alone or coupled with CCS, based on data collected from Saebeaa et al. (2020), Luyben (2018), and Susmozas (2015). The resulting block flow diagram is shown in Figure 1. The waste-polymer gasification (wPG) process consists of two main parts: syngas generation through steam gasification of wPE, followed by H_2 production and purification through water-gas shift (WGS) and pressure swing adsorption (PSA). CCS (wPG+CCS) is done on a CO_2-rich stream (95.6 mol%), which is compressed to 150 bar prior to injection.

The gasification takes place at 800 ºC and 1.013 bar with steam as the gasifying agent to generate syngas with 67.25 mole% H_2, 25.24 mole% CO, 7.33 mole% CO_2, and 0.18 mole% CH_4, as in Saebeaa et al. (2020). The stream is then compressed to 32.5 bar with intercooling before entering the WGS section to meet the conditions in Luyben (2018). The syngas undergoes a high-temperature water-gas shift (HT-WGS) at 400 ºC with 88 % conversion and a low-temperature water-gas shift (LT-WGS) at 250 ºC with 95 % conversion, following the reaction in Eq.(1).

$$CO + H_2O \leftrightarrow CO_2 + H_2 \tag{1}$$

The H_2-enriched stream is cooled down to 40 ºC and flashed before being sent to the pressure swing adsorption (PSA) unit, obtaining H_2 at 99.9 mole% purity based on Susmozas (2015). The tail gas is decompressed and undergoes combustion at 1300 ºC and 1.5 bar, covering the energy needs of the gasification. Combustion is performed with air, and the post-combustion stream is vented after cooling and flashing at 40 ºC.

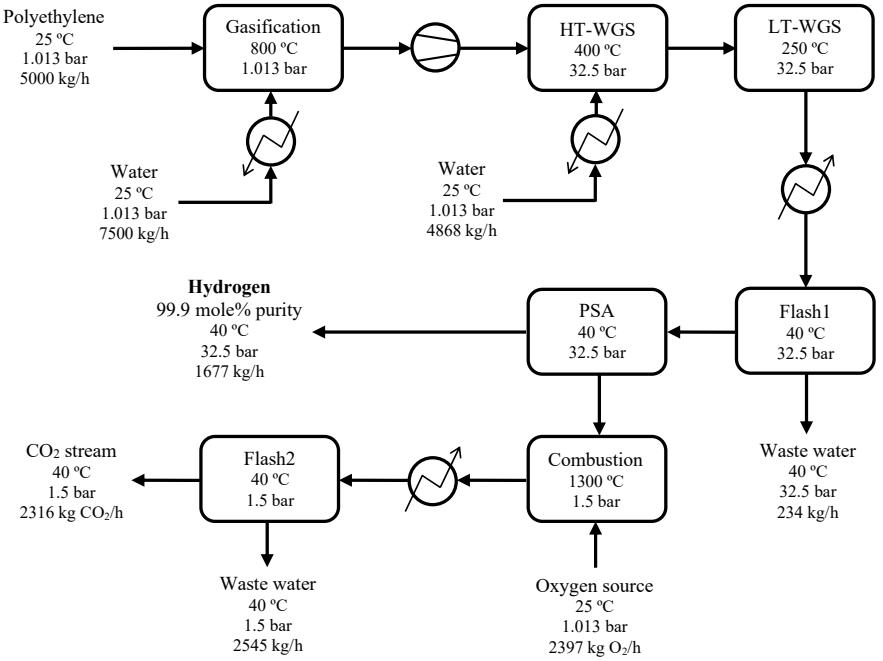


Figure 1. Block flow diagram of the process for hydrogen production from waste polyethylene.

For the process with CCS (wPG+CCS), oxy-combustion with a near stoichiometric O_2 ratio is considered at the same conditions as in wPG, obtaining a stream 95 mole% in CO_2 after cooling and flashing, which is compressed to 150 bar in four steps with inter-cooling and flashing, requiring 0.09 kW/kg CO_2 (Pipitone and Bolland of 2009). We considered that the flashed streams from this multi-stage compression are decompressed to 1.013 bar and undergo a final flash, recycling the gaseous stream to the combustion reactor. Therefore, no direct emissions are produced in the wPG+CCS process. The waste water streams from all flash units are sent to water treatment. Only cooling utilities are required.

2.2. Life cycle assessment (LCA) and scenarios definition

The environmental assessment was performed according to the ISO 14040 (2006). For each technology, the endpoint impacts on human health (HH), ecosystems quality (EQ), and resource depletion (RD) were calculated for 1 kg of hydrogen with the ReCiPe2016 method (Huijbregts et al. 2017) using data of global activities from Ecoinvent 3.7 (Wernet et al. 2016) in Simapro 9.2. The assessment follows a cradle-to-gate approach that considers the impacts from raw materials, electricity, process utilities, products, and direct emissions at point of substitution (APOS), disregarding the end-of-life of the plant infrastructure and use phase of the produced hydrogen. The ReCiPe2016 methodology quantifies the impacts on HH in disability-adjusted life years (DALY), which are the number of years during which individuals are not in total health; the effects on EQ are measured as the fraction of species that may be lost over time due to changes in environmental systems (species.y); RD in USD 2013 represents the extra cost required for the exploitation of resources in the future.

We quantified the life cycle inventories (LCI) for wPG and wPG+CCS from the mass and energy balances of the process in section 2.1. Moreover, we expanded the system boundaries to account for the treatment of waste polyethylene. Hence, we assume that our process avoids the landfilling and incineration of wPE, considering the proportion destined to each alternative worldwide in 2015, showcased by Geyer et al. (2017): 55 % of wPE to landfills and 25.5 % to incineration, using processes from Ecoinvent v3.7. The impacts of the remaining 19.5 % of waste, which is recycled, were omitted.

The wPG and wPG+CCS processes were compared to other hydrogen production routes, of which the inventories were taken from literature: SMR with and without CCS, following Dufour et al. (2012); biomass steam gasification (BG) with and without CCS, based on Susmozas et al. (2016); proton exchange membrane (PEM) electrolysis powered by various energy sources, following Lee et al. (2010), i.e., bioenergy with CCS (BECCS), hydropower, nuclear power, solar power from photovoltaic cells, wind power and the electricity mix from the 2018 power grid.

3. Results and Discussion

Figure 2 displays the endpoint environmental impacts of 1 kg of H_2 for the 12 scenarios studied here. The total values per impact category are also available in Table 1.

wPG and wPG+CCS perform differently for various reasons. wPG+CCS requires 14 % more electricity and 78 % more cooling water per kg of H_2 than wPG due to the CO_2 compression unit. Additionally, wPG+CCS includes the impacts embodied in the oxygen feedstock for oxy-combustion, avoided in the standard wPG route that uses excess air as an oxygen source and vents the post-combustion stream. Overall, wPG+CCS is worse in RD and better in HH and EQ.

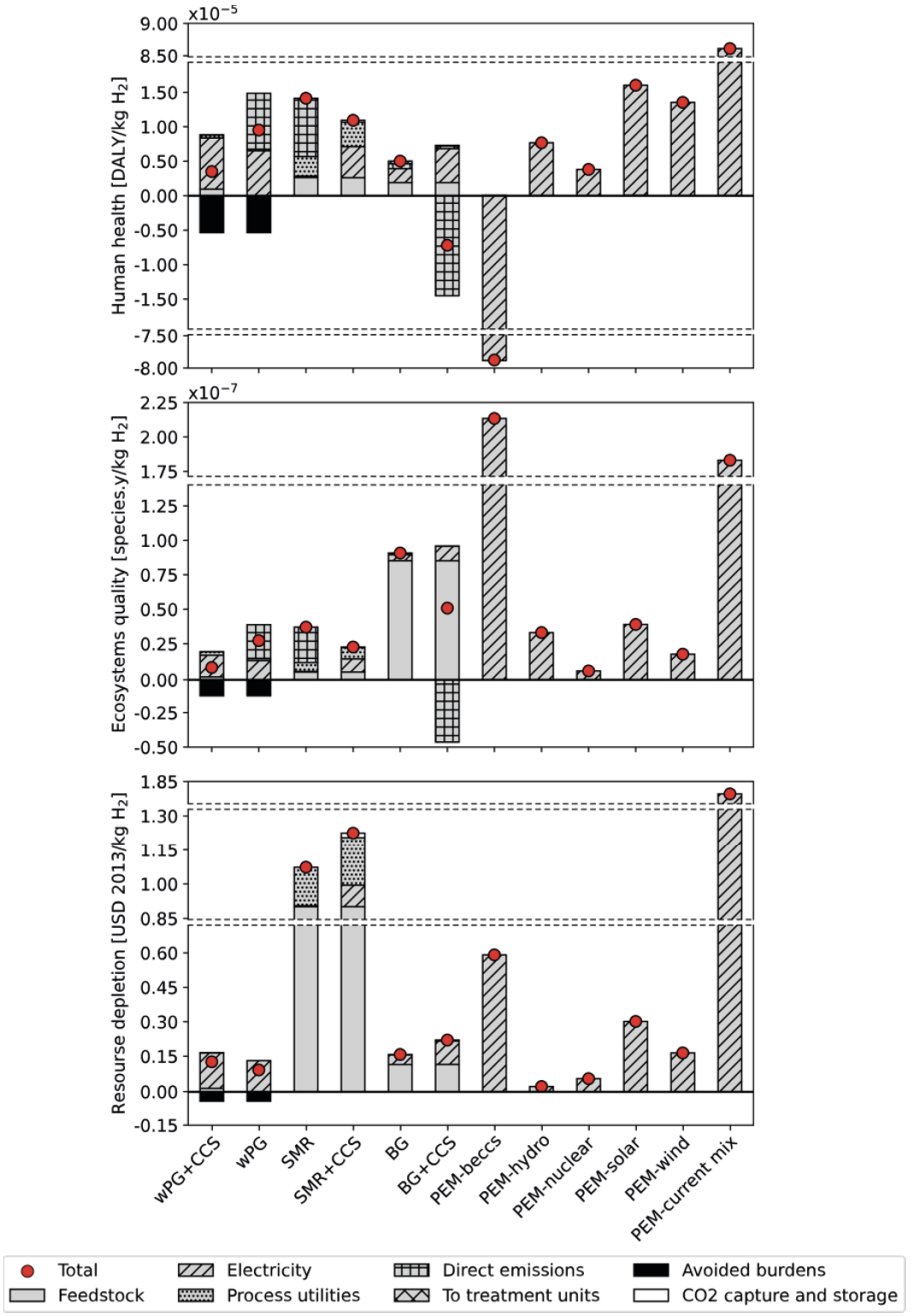


Figure 2. Endpoint environmental impacts of the hydrogen production routes assessed in this study, broken down by process components. The electricity term corresponds to the impacts of the energy sources used in the process (different for each PEM scenario). The following acronyms are employed: wPG: waste polyethylene gasification; CCS: carbon capture and storage; SMR: steam methane reforming; BG: biomass gasification; PEM: proton exchange membrane electrolysis; beccs: bioenergy with CCS; hydro: hydropower; nuclear: nuclear power plant; solar: photovoltaic energy; wind: wind power; current mix: electricity from the power grid of 2018.

Table 1. Total endpoint environmental impacts of the different technologies per impact category.

	Human health [DALY/kg H_2]	Ecosystems quality [species.y/kg H_2]	Resource depletion [USD 2013/kg H_2]
wPG+CCS	3.62×10^{-6}	9.16×10^{-9}	0.13
wPG	9.84×10^{-6}	2.92×10^{-8}	0.10
SMR	1.46×10^{-5}	3.93×10^{-8}	1.09
SMR+CCS	1.13×10^{-5}	2.44×10^{-8}	1.24
BG	5.19×10^{-6}	9.43×10^{-8}	0.17
BG+CCS	-7.43×10^{-6}	5.33×10^{-8}	0.23
PEM-beccs	-7.88×10^{-5}	2.13×10^{-7}	0.62
PEM-hydro	7.95×10^{-6}	3.51×10^{-8}	0.02
PEM-nuclear	3.95×10^{-6}	6.63×10^{-9}	0.06
PEM-solar	1.66×10^{-5}	4.12×10^{-8}	0.32
PEM-wind	1.40×10^{-5}	1.91×10^{-8}	0.18
PEM-current mix	8.62×10^{-5}	1.82×10^{-7}	1.79

In terms of impacts on HH, technologies involving biomass coupled to CCS (PEM-beccs and BG+CCS) are the most favorable scenarios, with negative impacts. wPG+CCS, PEM-nuclear, BG, PEM-hydro, wPG, SMR+CCS, PEM-wind, SMR, PEM-solar, and PEM-current mix follow. In terms of EQ, PEM-nuclear has the least impact, followed by wPG+CCS, PEM-wind, SMR+CCS, wPG, PEM-hydro, SMR, PEM-solar, BG+CCS, BG, PEM-current mix, and PEM-beccs. As for RD, PEM-hydro and PEM-nuclear are the least impactful technologies. wPG, wPG+CCS, BG, PEM-wind, BG+CCS, PEM-solar, PEM-beccs, SMR, SMR+CCS, and PEM-current mix follow.

wPG H_2 with and without CCS outperforms SMR H_2 in all three endpoint categories. Moreover, except for technologies involving biomass coupled with CCS (PEM-beccs and BG+CCS), wPG+CCS has the lowest impact on HH among the studied routes.

In terms of EQ, wPG+CCS is the second least harmful route, only behind PEM-nuclear. The technologies involving biomass and CCS, which performed very well in HH, are the most damaging to the ecosystems. This is due to the significant land requirements of biomass plantations (i.e., poplar).

As for RD, the PEM-current mix presents the highest impact because the 2018 energy mix is heavily reliant on fossil resources, as reported by the IEA 2019 World Energy Outlook. SMR with and without CCS follow as they require natural gas as feedstock, representing 83 % of the total RD impact for SMR and 73 % from SMR+CCS.

Interestingly, PEM-solar and PEM-wind are not the preferred option in any category. This is due to their life-cycle impacts, linked to the manufacture of photovoltaic panels and wind turbines, which are higher than those associated with other energy sources. For photovoltaic panels, these impacts mostly come from material and the energy provision for manufacturing, while for wind turbines, it is the material provision and construction. These results are consistent with the ones presented by Turconi et al. (2013).

4. Conclusions

Our work assessed the potential environmental benefits of producing hydrogen from waste polyethylene (wPE). Our results show that wPG+CCS outperforms SMR (business as usual) and SMR+CCS in the studied impact categories. Moreover, wPG+CCS performs better in human health than biomass (BG) and electrolytic H_2 (PEM) from renewables, excluding BECCS. It also shows lower impacts on ecosystems quality than

most processes (except for PEM-nuclear). The processes using wPE as a feedstock also display lower impacts on resource availability than SMR and BG with and without CCS.

Overall, our results suggest that hydrogen production based on plastic waste via wPG+CCS is environmentally appealing. This technology would help realize the circular economy concept in chemicals production by recycling polymer residues that would otherwise end up in landfills or incineration facilities.

Acknowledgments

This publication was created as part of NCCR Catalysis, a National Centre of Competence in Research funded by the Swiss National Science Foundation.

References

A. Pacheco-López, A. Somoza-Tornos, A. Espuña, M. Graells, 2021, Systematic generation and targeting of chemical recycling pathways: A mixed plastic waste upcycling case study, Computer Aided Chemical Engineering, 50, 1125-1130.

A. Susmozas, 2015, Analysis of energy systems based on biomass gasification (Doctoral thesis, Universidad Rey Juan Carlos), retrieved from BURJC-Digital database.

A. Susmozas, D. Iribarren, P. Zapp, J. Linßen, Javier Dufour, 2016, Life-cycle performance of hydrogen production via indirect biomass gasification with CO_2 capture, International Journal of Hydrogen Energy, 41, 42, 19484-19491.

B. Parkinson, P. Balcombe, J.F. Speirs, A.D. Hawkes, K. Hellgardt, 2019, Levelized cost of CO_2 mitigation from hydrogen production routes, Energy and Environmental Science, 12, 1, 19-40

D. Saebeaa, P. Ruengritb, A. Arpornwichanopc, Y. Patcharavorachot, 2020, Gasification of plastic waste for synthesis gas production, Energy Reports, 6, 202-207

G. Pipitone, O. Bollav, 2009, Power generation with CO_2 capture: Technology for CO_2 purification, International Journal of Greenhouse Gas Control, 3, 5, 528-534.

G. Wernet, C. Bauer, B. Steubing, J. Reinhard, E. Moreno-Ruiz, B. Weidema, 2016, The ecoinvent database version 3 (part I): overview and methodology, International Journal of Life Cycle Assessment, 21, 1218–1230.

IEA, World Energy Outlook 2019, 2019, International Energy Agency.

ISO, 2006, ISO 14040:2006(E): Environmental Management – Life Cycle Assessment – Principles and Framework, International Standards Organization.

J. Dufour, D.P. Serrano, J.L. Gález, A. González, E. Soria, J.L. Fierro, 2012, Life cycle assessment of alternatives for hydrogen production from renewable and fossil sources, International Journal of Hydrogen Energy, 37, 2, 1173-1183.

J. Lee, S. An, K. Cha, T. Hur, 2010, Life cycle environmental and economic analyses of a hydrogen station with wind energy, International Journal of Hydrogen Energy, 35, 6, 2213-2225.

M. Huijbregts, Z. Steinmann, P. Elshout, G. Stam, F. Verones, M. Vieira, A. Hollander, M. Zijp, R. Van Zelm, 2017, Recipe2016: a harmonized life cycle impact assessment method at midpoint and endpoint level, International Journal of Life Cycle Assessment, 22, 138-147.

Plastics Europe, 2021, Plastics – the facts 2020, Plastics Europe.

R. Geyer, J. R. Jambeck, K. L. Law, 2017, Production, use, and fate of all plastics ever made, Science Advances, 3, 7.

R. Turconi, A. Boldrin, T. Astrup, 2013, Life cycle assessment (LCA) of electricity generation technologies: Overview, comparability and limitations, Renewable and Sustainable Energy Reviews, 28, 555-565.

W. L. Luyben, 2018, Plantwide control of a coupled/reformer ammonia process, Chemical Engineering Research and Design, 134, 518-527.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50323-7

A Systematic Comparison of Renewable Liquid Fuels for Power Generation: Towards a 100% Renewable Energy System

Antonio Sánchez[a]*, Elena C. Blanco[a], Mariano Martín[a], Pastora Vega[b]

[a]*Department of Chemical Engineering. University of Salamanca. Plz. Caidos 1-5. 37008. Salamanca (Spain)*
[b]*Department of Computer Science and Automatic. Univeristy of Salamanca. Plz. Caidos 1-5. 37008. Salamanca (Spain)*
antoniosg@usal.es

Abstract

Energy store will be essential in the future power system due to the inherent fluctuations of the renewable resources. The use of liquid fuels, such as methanol or ammonia, allows for seasonal storage and, as energy carriers, can be used for different energy applications. In this work, a systematic evaluation of different alternatives to produce power from methanol and ammonia has been performed. Particularly, for each of them, the thermochemical (combustion) and electrochemical (fuel cells) routes are evaluated. The operating conditions of each of the alternatives are optimized yielding energy efficiencies between about 15-40 %. From an economic point of view, the cost of electricity for the proposed power facilities is around 0.8 €/kWh for the fuel cell based production and 0.25 €/kWh for the combustion one, with lower prices for ammonia based alternatives. The fuel cells are more suitable for small scale applications, hence, with lower economies of scale. Therefore, this techno-economic assessment demonstrates the feasibility of liquid fuels to provide a robust power grid based on renewable resources meeting sustainable development.

Keywords: Ammonia, Energy Storage, Methanol, Power-to-X, Renewable fuels

1. Introduction

Renewable energy is a vital resource for the energy transition and sustainable development, in order to achieve the UN target of ensuring access to affordable, reliable, sustainable, and modern energy for all. Therefore, an increase in the share of renewable energy sources (RES) is expected in the coming years. IRENA (2018) forecasts that, by 2050, renewable generation will account for about 85 % of the global power generation, and around 60 % of the total final energy consumption. However, this increase in renewable penetration threatens the stability of the power system due to the inherently stochastic nature of the two main RES, wind and solar. At this point, energy storage will be crucial to ensure demand satisfaction at every time regardless of weather conditions in an optimal way. In this area, two alternatives emerge as the most promising in the future scenario, lithium-ion batteries and hydrogen and its derivatives (Schmidt et al., 2019). This latter option is particularly suitable for long-term energy storage (more than 700h). Apart from the power system, the share of renewables in other activities must be increased to achieve a high sustainable final energy

consumption rate. In this field, energy carriers arise as one of the main tools for different applications where direct electrification is difficult.
Regarding energy storage/carriers, liquid fuels, such as methanol or ammonia, emerge as one of the most attractive options. Liquid fuels have a high volumetric energy density, a scalable and flexible behavior for different time scales and power capacities, and easy conditions of storage and transportation. Several works in the literature have analysed Power-to-liquids (PtL) technology taking into account the variability of the solar/wind availability and its influence on the chemical production. Once the fuel has been synthesised, the second stage will be the generation of electricity when it is needed, for instance, when renewable generation is low. At this point, two different alternatives are available for this purpose. The first one is based on the combustion of the liquid fuels to produce power using, mainly, a gas turbine. On the other side, liquid fuels can be converted into power using a fuel cell. Direct methanol/ammonia fuel cells are being developed, but another alternative is a first step based on the reforming of the fuels and, subsequently, using a hydrogen fuel cell which is a more mature technology. Different experimental works have evaluated the transformation of liquid fuels to power at laboratory scale from different perspectives. However, a process-scale analysis is required, including all the stages involved in the transformation, determining the energy efficiency of the process, and computing the cost of electricity for the different alternatives. This step is essential for planning an orderly introduction of these liquid fuels in real applications such as grid storage, mobility, etc.
In this work, a holistic comparison of the two main hydrogen liquid carriers, methanol and ammonia, is performed. For each of them, two different transformation routes have been studied: thermochemical (combustion) and electrochemical (fuel cell). The preparation of the raw material, the power conversion, and the different gas treatment operations have been included in this assessment. With this work, a fair comparison of both liquid fuels is conducted from a technical and economic perspective.

2. Process Description

Two different alternatives are possible to transform methanol or ammonia into power: thermochemical or electrochemical (Figure 1). The combustion of methanol alone is difficult due to the cold start (related to the high latent heat of vaporization of methanol) and the associated emissions. Therefore, the methanol/hydrogen blends have been proposed as one of the most promising fuels in order to overcome the issues associated with methanol combustion. Hence, the first step of the process is the preparation of the mixture, in this case, a proportion of 85 % methanol and 15 % of hydrogen is used (Zhen & Wang, 2015). The hydrogen required is generated using methanol steam reforming in a catalytic reactor using $Cu/ZnO/Al_2O_3$ as catalyst. In the combustion of the blend, undesirable NO_x is generated. Its concentration is computed using an experimental-based correlation. The maximum temperature in the combustion chamber is limited to 1873 K due to material limitations. To produce power using the gases from the combustion chamber, a combined cycle is employed with a first step based on a gas turbine and a second part including a Rankine cycle.
Ammonia combustion follows a similar pattern than methanol. Due to issues related to ammonia combustion, a mixture of 70 % of ammonia and 30 % of hydrogen is used as fuel (Sánchez et al., 2021). Hydrogen is produced through ammonia decomposition in a membrane catalytic reactor. As in the methanol alternative, a combined cycle is set up to generate power. To overcome the maximum gas turbine operating temperature, an inert component (argon) is introduced into it. Due to the combustion characteristics,

different gas clean-up operations are required. Firstly, NO_x removal can be introduced if necessary to meet the environmental restrictions. Secondly, hydrogen is leaving the combustion chamber and must be recovered using a selective membrane. After the gas clean-up section, a mixture of N_2 and Ar is obtained and, to reuse these components, a distillation column is used to separate them.

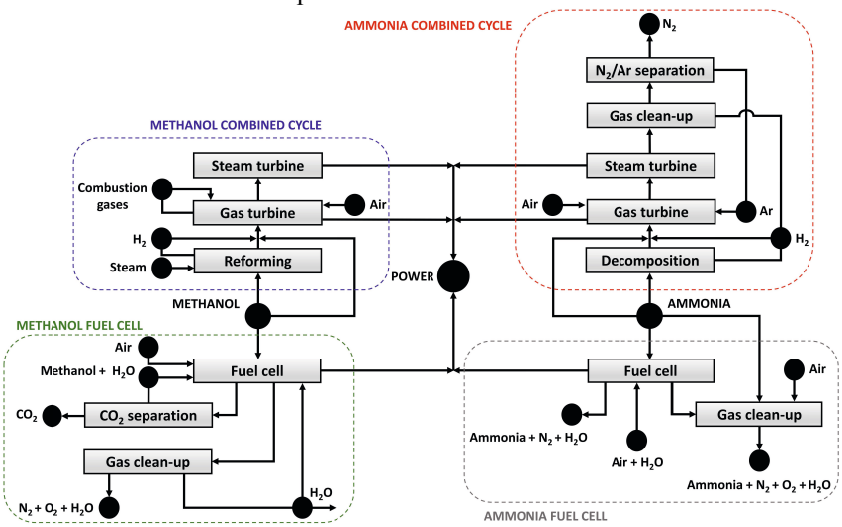


Figure 1: Process superstructure diagram of power production using methanol or ammonia

The electrochemical alternative is also evaluated. Methanol can be transformed into power in a fuel cell. An aqueous solution of methanol (1 mol/L) is fed to the unit along with air as oxidizer (Goor et al., 2019). From the anode of the cell, carbon dioxide, water and traces of methanol are obtained and, from the cathode, water, carbon dioxide, oxygen and nitrogen leave the cell. The operating conditions of the fuel cell (mainly voltage and intensity) determine the amount of power that can be produced from methanol. The anode stream is separated to recycle methanol and water into the fuel cell. Finally, the CO_2 from the cathode stream is also recovered by a zeolite system for synthesis purposes.

The last analysed system in this work is the electrochemical conversion of ammonia into power (Siddiqui & Dincer, 2019). In this alternative, ammonia is fed to the fuel cell together with wet (50 %) air. The main products of the reaction are nitrogen and water. From the anode, a mixture of nitrogen, water and traces of ammonia is produced. In the cathode, nitrogen, water, oxygen from the air but also NO_x is obtained that must be removed using a selective catalytic removal (SCR).

All the units and operations of each section in Figure 1 have been modelled using an equation based approach. Different modelling techniques have been used including mass and energy balances or experimental correlations. The entire superstructure is decomposed as a set of non-linear programming problems, one for each of the alternatives and fuels (up to 2000 equations per alternative). The formulation is implemented in GAMS and CONOPT 3.0 has been used as preferred solver in a multistart optimization approach. As objective function, a simplified operating cost of the facility is used such that (Eq.(1)):

$$Z = \sum_{i \epsilon IN} f_i C_i - \sum_{j \epsilon OUT} f_j C_j \quad (1)$$

where f_i is the flow of the inlet or outlet resources of the facility and C_i is the cost associated. A price of 0.43 €/kg and 0.55 €/kg is set for methanol and ammonia respectively (Matzen et al., 2015). The optimization procedure determines the optimal operating conditions of the different alternatives proposed in order to improve the energy performance of the system minimizing the cost of electricity.

3. Results

First, some technical and operating results of the power facilities using liquid fuels are presented. The main operating variables of each of the alternatives are optimized minimizing the cost of electricity. For this optimized scenario, the efficiency and specific energy are shown in Table 1 for two different fixed capacities. The fuel cells are more suitable for small scale applications; therefore, the production capacity is set considerably lower than in the thermochemical way.

Table 1: System efficiency results for the methanol/ammonia to power transformation

Capacity (MW)	Technology alternative	Efficiency (%)	Specific energy (kWh/kg)
100.0	Methanol combined cycle	38.08	2.122
	Ammonia combined cycle	33.95	1.768
1.0	Methanol fuel cell	22.99	1.032
	Ammonia fuel cell	15.00	0.710

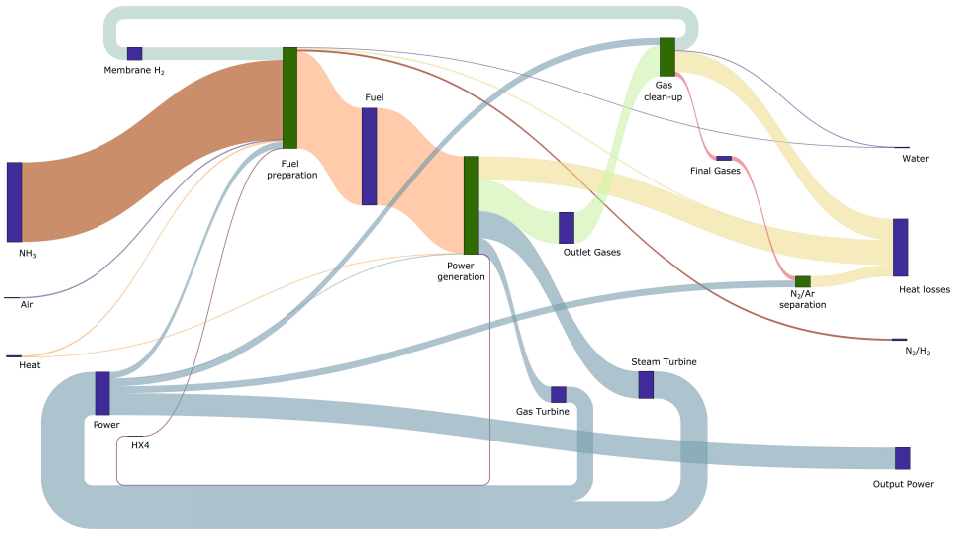


Figure 2: Sankey diagram for the energy flows in thermochemical ammonia transformation

In general, the thermochemical alternatives present better efficiencies than the electrochemical options with differences of almost double. The better efficiency is achieved in the methanol combined cycle reaching about 38 %. The thermochemical processes are based on the combination of a gas turbine and a Rankine cycle, achieving, a good performance in efficiency terms. Nevertheless, the main limitation to improve the efficiency values of these alternatives is the maximum temperature allowed in the gas turbine due to material restrictions. If this constraint is removed, a significant increase in the efficiency values could be expected. Therefore, a detailed research in the gas turbine construction should be made to widen the operating conditions of these units. Methanol combined cycle shows better efficiency results than ammonia based alternatives. The ammonia transformation is a more complex process that includes gas clean-up operations and N_2/Ar separation. These operations, particularly the last one, are energy demanding reducing the energy performance of the whole system.

The electrochemical systems show a drastically reduction in the efficiency values, about 15-20 %. These technologies are still under development at laboratory scale and, therefore, further research is required to improve the energy efficiency of these devices. These values are significantly lower than those obtained in the hydrogen alternative because the liquid fuels studied have a lower electrochemical reactivity. The lower values of efficiency in the methanol/ammonia fuel cells are due to the low operating voltage of these devices. At this point, the temperature is limiting the cell voltage to avoid the damage of the electrochemical catalyst. A better design of the cell membrane could improve the performance of these systems reducing the fuel crossover or improving the catalyst features. Methanol fuel cell shows better efficiency due to, according to experimental results, can operate with higher voltage and intensity and, therefore, with higher power density.

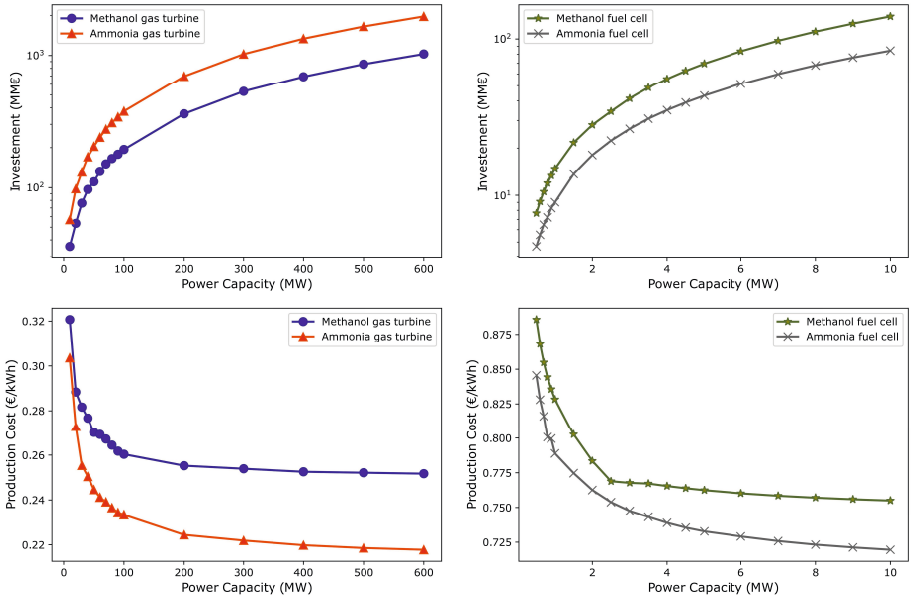


Figure 3: Capital and production costs of the different methanol/ammonia power generation system

The performance of the entire system, including all the operations involved, is analysed to determine the flows of energy in each section. For instance, in Figure 2, a Sankey diagram is included to analyse the energy operation of the thermochemical conversion of ammonia. A significant amount of energy is required in the different operations in the facility reducing the total performance of the system. For example, around 50 % of the produced power is devoted to internal operations. Similar results are obtained for the rest of the alternatives, however, for the sake of brevity, only one is included in this chapter. This holistic analysis is necessary to a fair evaluation of the system because including only the fuel cell or the combustion system is not enough to determine the real energy operation.

An economic evaluation of this power generation system based on liquid fuels is also performed in this work (as shown in Figure 3). The thermochemical alternatives exhibit a production cost about 0.2-0.3 €/kWh with an associated investment of about 3-5 MM€/MW depending on the capacity. For the electrochemical alternatives, the production costs significantly increase to values around 0.75-0.85 €/kWh requiring an investment of around 10 MM€/MW due to the lower production capacities than the previous alternatives. The lower energy efficiency of these systems and the lower power capacities (with reduced economies of scale) determine the worst economic performance of this electrochemical route.

4. Conclusions

In this work, a systematic evaluation of the transformation of methanol/ammonia into power is performed. Two different main routes have been assessed: thermochemical (combustion) and electrochemical (fuel cell). The entire process is analysed including the preparation of the raw materials, the power production and the final purification operations. The technical performance of these systems is demonstrated yielding efficiency values of about 35 % for the thermochemical and 18 % for the electrochemical alternatives. The operating cost is about 0.25 €/kWh for the combustion based processes and 0.8 €/kWh for the fuel cells due to the lower energy performance and the reduced power capacities. This analysis leads to a successful integration of these technologies in real applications with the objective of a fully decarbonized energy system.

References

M. Goor, S. Menkin, E. Peled, 2019, High power direct methanol fuel cell for mobility and portable applications, International Journal of Hydrogen Energy, 22, 3138-3143.

IRENA, 2018, Global Energy Tranformation: A roadmap to 2050, International Renewable Energy Agency, Abu Dhabi.

M. Matzen, M., Alhajji, Y. Demirel, 2015, Technoeconomics and Sustainability of Renewable Methanol and Ammonia Productions Using Wind Power-based Hydrogen. Advanced Chemical Engineering, 5,4, 1000128.

A. Sánchez, E. Castellano, M. Martín, P.Vega, 2021, Evaluating ammonia as green fuel for power generation: A thermo-chemical perspective, Applied Energy, 293, 116956.

O. Schmidt, S. Melchior, A. Hawkes, I. Staffell, 2019, Projecting the Future Levelized Cost of Electricity Storage Technologies, Joule, 3 (1), 81-100.

O. Siddiqui, I. Dincer, 2019, Development and performance evaluation of a direct ammonia fuel cell stack, Chemical Engineering Science, 200, 285-293.

X. Zhen, Y. Wang, 2015, An overview of methanol as internal combustion engine fuel, Renewable and Sustainable Energy Reviews, 52, 477-493.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50324-9

Guiding innovations and Value-chain improvements using Life-cycle design for Sustainable Circular Economy

Vyom Thakker, Bhavik R. Bakshi*

William G. Lowrie Department of Chemical and Biomolecular Engineering, The Ohio State University, Columbus, Ohio 43210, USA

bakshi.2@osu.edu

Abstract

While the current linear state of the economy has led to large scale natural-resource exploitation and pollution, circular economy can also lead to unexpected harm to environmental sustainability. Thus, there is a need to design product value-chains to achieve a Sustainable and Circular Economy (SCE). Previous work has focused on developing a systems engineering framework using life-cycle assessment with 'superstructure' network optimization to find optimal value-chain pathways while considering product life-cycles. However, the role of innovations in the form of novel technologies, societal action and new policy action has become increasingly crucial to establish SCE. In this work, we propose a sensitivity optimization framework to find the most attractive innovation directions within the value-chain using parameter perturbations as additional decision variables over pathway choice. The objectives include maximizing circularity and minimizing carbon dioxide emissions. We quantify the trade-off between these objectives and determine win-win innovative solutions using pareto and perturbation fronts. The method is demonstrated for an illustrative example, and its applicability to real value-chain networks has been probed.

Keywords: Life-cycle design, Sustainable and circular economy, Multi-objective optimization, Innovation modeling

1. Introduction

The current 'linear' state of the economy is contributing to many man-made disasters like climate change, plastic oceanic gyres, resource scarcities, harmful algal blooms in lakes, etc. While 'circular economy' is expected to bring about reduction in waste and pollution, it may not always be aligned with sustainability requirements such as curtailing climate change and respecting nature's carrying capacity. Progress toward a Sustainable Circular Economy (SCE) is crucial to mitigate large-scale exploitation of natural resources and pile-up of man-made materials like plastics in the environment. For achieving a SCE there is a need to holistically design entire value-chains of products and services while considering the environmental, economic, and social implications of potential alternatives. The field of Process Systems Engineering (PSE) has the potential to contribute towards establishing SCE for material life-cycles provided it expands its system boundary to account for the life cycle, economy, and ecosystems (Bakshi, 2019). This work is aimed at expanding PSE models and methods toward Sustainable Engineering to find optimal value-chain reforms and discover most attractive innovation directions.

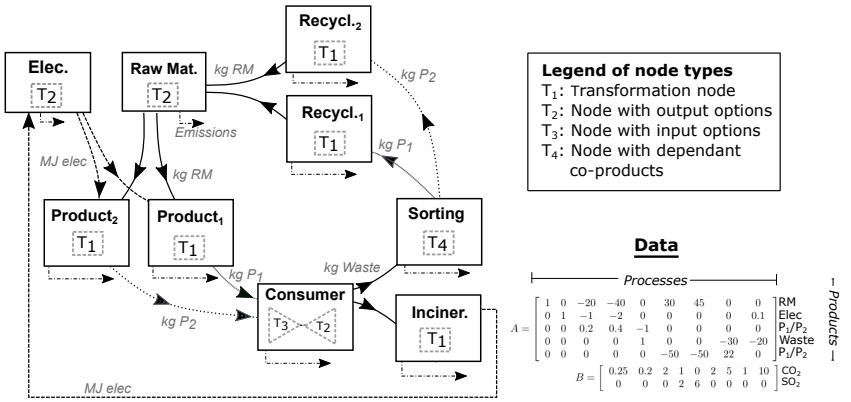


Figure 1: Illustrative example of a typical circular value-chain network

We construct the connection between SCE and PSE using the framework of life-cycle assessment (LCA), which focuses on calculating net environmental impact of 'singular' product value-chains, while considering entire life-cycles of the product - right from natural resource extraction to disposal. The computational structure of LCA involves solving a linear programming (LP) problem on life-cycle inventory data of value-chains (Heijungs and Suh, 2002). In our previous work (Thakker and Bakshi, 2021a), we have developed the SCE design framework which uses optimization to evaluate multiple alternative value-chains by creating a 'superstructure' network of alternatives. In this work, we expand the scope of design to identify the optimal perturbations in technology efficiencies, supply chains, policies and behavior that can be brought about by innovations. These perturbations do not consider systemic disruptions to value-chains that can be brought about by innovations, which need to be modeled as separate processes in the superstructure network. Sections 2 and 3 of this paper are devoted to finding SCE optimal value-chain pathways from the illustrative example network in figure 1 using previous work. Section 4 then describes the novel sensitvity optimization framework developed for innovation guidance, which is followed by insights on potential applications of the framework in conclusions.

2. Life Cycle Assessment (LCA) framework

LCA is used to find the impact of catering to a particular demand of a product, from the processes in its entire life cycle. The input, output and emissions data for each process is found from national averages, and are included within columns of the technology and intervention matrices (A and B). These matrices are available from national agencies and commercial organizations. LCA method consists of the equations, $As = f$, $g = Bs$ which ensures flow conservation of all products (rows), while meeting the final demand 'f', specified for the LCA study. 's' is the scaling vector, which represents the scale of operation of each process in A to meet the demand. 'g' is a vector of total resource use and emissions for meeting the demand, and is found by scaling the interventions B with the same 's'. One short-coming of LCA is that it requires separate analysis for each alternative value-chain pathway, which is rectified using the SCE design method described below.

3. SCE Design Framework

In our previous work (Thakker and Bakshi, 2021a), we have developed a multi-objective superstructure optimization method to find optimal value-chain pathways for SCE objectives, and develop pareto fronts to quantify trade-offs between these objectives.

3.1. Node-alphabet representation of superstructure networks

The framework is generally applicable to any circular system owing to the node-alphabet representation. Nodes in a typical life-cycle network are classified into 4 types (T_{1-4}) according to the substitutability of inputs and outputs, and whether the streams undergo transformations. Any SCE network may be represented as a combination of these node types, as shown for the illustrative example in figure 1.

3.2. Illustrative example

This paper describes the foundational work and methodological developments using an illustrative example, shown in figure 1. This example involves finding the optimal-value chain to meet a consumer demand from one of the two-products (P_1 & P_2), which are sent either to segregation (and recycling) or to incineration. The goal is to find optimal pathways to meet SCE requirements. This illustration is chosen since the network is representative of typical product life-cycles, e.g. plastic containers, semi-conductors, laptops, etc., thereby highlighting the wide applicability of this work to relevant SCE problems.

3.3. Constraints

The decision variables of the SCE design problem are the scaling factors 's' denoting the pathway selection, i.e. s_j is 0 if value-chain process j is inactive. Since the technology matrix A represents a 'superstructure' network of alternatives, A is a rectangular matrix (not full rank) and pathways design for an arbitrary objective $Z(s)$ is possible. However, optimizing value-chains requires flow conservation with the life-cycle which is established by specifying the LCA equations as constraints on the decision variables.

$$\begin{aligned} &\min_s z := Z(s) \\ &\text{Subject to:} \quad As = f \\ &\qquad\qquad\ \ g = Bs \end{aligned} \tag{1}$$

In addition, the network needs to be scaled to meet consumer demand, which is added as a constraint on the life-cycle flows (Hs). Furthermore, governing equations such as material, energy and component balances are specified as balance constraints ($\mathcal{F}_B$). Node efficiencies are also specified as constraints ($\mathcal{F}_n$) for each node based on its type (T_{1-4}).

$$Hs \geq u \tag{2}$$

$$\mathcal{F}_B(s) \leq 0 \tag{3}$$

$$\mathcal{F}_n(s) \leq 0 \tag{4}$$

Finally, non-negative scaling ($s \geq 0$) and the net-zero final demand of intermediate flows ($f_i = 0$) are ensured using variable bounds. All these constraints yield a feasible design space of pathway choices and technology options in a non-linear problem (NLP), which is optimized for various SCE objectives and to characterize the trade-offs between them.

3.4. Objectives and Pareto-front generation

SCE objectives must comprise of Environmental, Economic and Circularity aspects. The emission and effluent flows are captured within the g vector and can be used as environmental objectives. In the illustrative example, there are two emission flows, i.e. carbon dioxide (CO_2) and sulfur dioxide (SO_2), which need to be minimized. However, real life-cycle inventory data contains hundreds of emission flows, which can be aggregated into midpoint indicators such as global warming, acidification and eutrophication potential.

The economic objective can be formulated as life cycle cost (LCC), which would consider the cost of directly and indirectly used natural resources. For simplicity, this objective has been excluded from the illustrative example. Within the circularity domain, we formulate a novel metric θ using life-cycle flows of the network to quantify the circularity of the network. It is calculated as the ratio of the value of circular flows within the system to the value of manufactured products ($M.s$). Circular flows comprise of recycling, refurbishment, down-cycling and up-cycling in the technological system ($C \in A$) and valuable effluents such as compost g_c to the environment. The general expression for θ is as follows,

$$\theta = \frac{\Sigma\gamma_i C.s + \Sigma\gamma_k g_c}{\Sigma\gamma_i M.s} \quad i \in \text{products}, k \in \text{emissions}, C \in A_{\text{circular}},\ M \in A_{\text{manufacturing}} \tag{5}$$

γ denotes the value function, typically in monetary, exergetic or physical units, and it determines the nature of the circularity metric, θ. In this example, we consider monetary circularity with both recycled raw material and electricity generating monetary value. Since we consider multiple objectives of SCE, there are bound to be trade-offs and win-win solutions. These are quantified using pareto optimal solutions, found using the ϵ-constraint method. The pareto-optimal solutions form a front which represent the best possible solutions without bias to any one of the three objective domains. The points lying above the front are sub-optimal and the ones lying below are infeasible. Thus, pareto front generation provides quantification and visualization of trade-offs.

3.5. SCE designs for the illustrative example

The SCE design solutions for the illustrative example are shown on the extremities of the black solid line pareto front in figure 2. The P_1-C-S-R_1 pathway corresponds to minimum CO_2 emissions, whereas the P_2-C-S-R_2 pathway has maximum circularity. ϵ-constraint is used to find points on the pareto front (solid line) which correspond to impartial compromise solutions between the objectives. The objective space below the front is sub-optimal, whereas the space above corresponds to win-win solutions which can only be achieved by innovations and pathways outside the superstructure network.

4. Sensitivity optimization for innovation discovery

It is crucial to identify the most attractive directions for innovations in the value-chain to improve SCE objectives. This is achieved by modifying the SCE design framework to

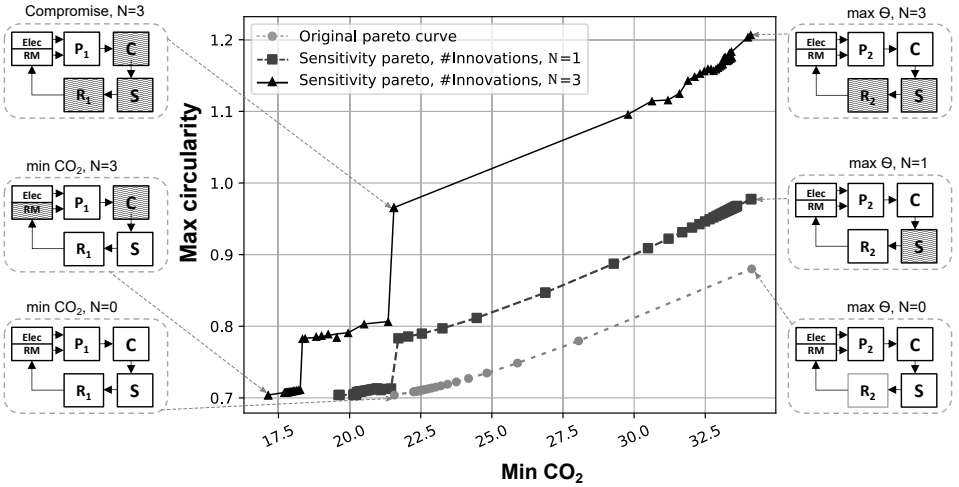


Figure 2: Pareto and Perturbation fronts for SCE design of network in figure 1.

include additional decision variables corresponding to the sensitivity of the parameters in the technology matrix(A). Each element in A (a_{ij}) represents a particular property of a life-cycle activity (j). We introduce binary variables y_{1ij} and y_{2ij} to assume the value '1' if a_{ij} can be increased or decreased (respectively), by a factor of $\Gamma \in [0, 1]$. These perturbations to a_{ij} are assumed to be brought about by innovations, better technologies, or improved policy, and the number of such permitted perturbations (N) are set by the user of the framework. Flow conservation and material-energy balances need to hold despite these perturbations, which is done by changing the original design formulation to the following.

$$
\begin{aligned}
&\min z := \text{kgCO}_2,\ \text{kgSO}_2 \\
&\max_{s,y_1,y_2}\ z := \text{Circularity } (\theta) \\
\text{Subject to:}\quad & A \odot (1 + \Gamma y_1 - \Gamma y_2) s_j = f \\
& y = y_1 + y_2 \\
& \sum_i \sum_j y_{ij} \leq N,\ \text{\# of permitted perturbations} \\
& G(s) \geq 0,\ \text{other constraints on } s \\
& y_1,\ y_2, y \in \mathbb{Z}^{i \times j};\ \Gamma \in [0, 1];\ s \geq 0
\end{aligned}
\tag{6}
$$

Here, '$\odot$' represents element-wise multiplication, and (y_1, y_2) are binary variables to identify the optimal innovative perturbations in the positive and negative directions. The resulting optimization is a MINLP that finds best value-chain pathways and most attractive perturbations within a user-defined range (Γ=0.1). The pareto fronts for SCE objectives using this formulation are referred to as 'perturbation fronts'. These fronts provide win-win solutions over the original pareto front, and greater win-win is obtained when more innovation perturbations (N) are allowed. For the illustrative example, two perturbation fronts are developed; with $N = 1$ (dashed line), $N = 3$ (dotted line), as shown in figure 2.

While the value-chain pathways on the extremes are identical to the original pareto front, the perturbed value-chain activities (shaded boxes) vary. For instance, minimum CO_2 emissions demand innovations to focus on manufacturing and consumer use, whereas highest circularity requires them to perturb sorting and recycling of P_2. A 'compromise' solution on the perturbation front is found (top left pathway in figure 2), which improves both objectives from the original SCE design. This solution says that innovations must be focused towards improving efficiencies and yields of sorting and recycling, while also increasing consumer re-use. Through this illustrative example, we prove the utility of the method to guide value-chain reforms and innovations for any SCE network of relevance.

5. Conclusions

In this work, we have expanded the previously developed SCE design framework (Thakker and Bakshi, 2021a) to include sensitivity optimization for finding most attractive innovation directions, along with pathway design of value-chains. The new modeling framework is demonstrated upon an illustrative example, to find the optimal perturbations in technology and societal parameters that can lead to win-win solutions from circularity and CO_2 emissions viewpoint. Pareto front generation allows quantification of trade-offs and selection of pareto-optimal solutions, which can inform new research directions based on a reasonable 'compromise' between SCE objectives. Future work will pertain to application of this methodology to a real-life value-chains of products, such as plastic-containers, windmills, etc. While the general applicability of the framework is established in section 3.1., the tractability of sensitivity optimization for large value-chain network is currently being explored using a case study on plastic grocery bags. In addition, it may be needed to introduce a physico-chemical transformation network using a multi-scale approach (Thakker and Bakshi, 2021b) to provide a realistic constraints on allowable perturbations.

6. Acknowledgments

The authors acknowledge the support from "The Global KAITEKI Center" at Arizona State University (ASU), and the National Science Foundation (NSF) grant EFMA-2029397.

References

Korhonen, J., Honkasalo, A. and Seppälä, J., 2018. Circular economy: the concept and its limitations. Ecological economics, 143, pp.37-46.

Bakshi, B.R., 2019. Toward sustainable chemical engineering: the role of process systems engineering. Annual review of chemical and biomolecular engineering, 10, pp.265-288.

Thakker, V. and Bakshi, B.R., 2021. Toward sustainable circular economies: A computational framework for assessment and design. *J of Cleaner Production*, 295, p.126353.

Heijungs, R. and Suh, S., 2002. The computational structure of life cycle assessment (Vol. 11). *Springer Science & Business Media.*

Thakker, V. and Bakshi, B.R., 2021. Multi-scale Sustainable Engineering: Integrated Design of Reaction Networks, Life Cycles, and Economic Sectors. *Computers & Chemical Engineering*, p.107578.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50325-0

Simultaneous Optimal Operation and Design of a Thermal Energy Storage Tank for District Heating Systems with Varying Energy Source

Caroline S. M. Nakama[a], Agnes C. Tysland[a], Brage R. Knudsen[b], Johannes Jäschke[a*]

[a]*Department of Chemical Engineering, Norwegian University of Science and Technology (NTNU), Sem Sælandsvei 4, Trondheim 7491, Norway*
[b]*SINTEF Energy Research, Kolbjørn Hejes vei 1B, Trondheim, 7491, Norway*

johannes.jaschke@ntnu.no

Abstract

District heating systems based on industrial waste heat play an important role in using energy efficiently. Combined with a thermal energy storage technology, such as pressured-water tanks, they have the potential of significantly reducing greenhouse gas emissions as well. However, installing thermal energy storage requires capital and, therefore, it is important to find an optimal design that balances the benefits of energy storage with the costs of installing such system. In this work we formulate a dynamic optimization model for designing a thermal energy storage tank based on operational conditions and apply it to a case study using historical data from a district heating system that recovers heat from an industrial plant in Norway. We found that a relatively large tank (greater than 5000 m3) would be necessary to store all excess energy provided by the plant that cannot be immediately used for the period and input data considered. However, the results can be used to investigate uncertainties and their effects on the optimal tank volume and return of investment.

Keywords: Energy systems; thermal energy storage; optimal operation; optimal design.

1. Introduction

Environmental, energetic and climate issues of today require a shift from society's fossil fuel dependency to renewable energy sources. The pace of this change must accelerate, and significant measures are taken to increase the development and use of renewable-energy-based technologies (Mirandola and Lorenzini, 2016), and environmental policies implemented by governments. For such shifted scenario, decarbonized energy system, district heating (DH) systems and thermal energy storage (TES) can play a critical role and contribute significantly to Europe's 2050 emission goals (Connolly el al., 2014). An important DH system type is those utilizing industrial waste heat; however, due to the commonly high variation of the waste heat availability, its combination with TES is of interest to further reduce the use of peak-heating sources. Pressured-water tanks are the most suitable TES technology for DH systems, yet they can be very costly and space availability may be limited (Knudsen et al., 2021).

In this work we focus on the optimal *operation and design* of a TES tank for utilization in a DH system based on waste–heat recovery. Integrating operation into the sizing

problem is important, as operational conditions have a significant impact on how efficiently the waste heat is utilized, which in turn can influence the size of the TES tank. We present an approach that formulates a single nonlinear dynamic optimization problem that accounts for optimal operation and sizing simultaneously, as opposed to combined optimization/simulation-based methods previously proposed, e.g., Knudsen et al., 2021; Li et al., 2021. We demonstrate this method on a historical data set from a DH plant in Norway that recovers heat from a ferrosilicon plant.

2. Case Study

We consider a case study for designing a TES tank for the heating plant of the DH system of Mo i Rana in Norway. The DH plant is located inside Mo Industry Park and receives waste heat from a ferrosilicon plant. The objective of the TES is to increase the waste-heat utilization and thereby reduce necessary peak-heating.

The DH system has 6 boilers heating up the water that is sent back to the city. Two of them use waste-heat from the industrial park and four of them are peak-heating boilers. They run primarily on electricity or CO-gas as energy source, the latter being a by-product from a manganese plant in the industry park and thus with varying availability. Since today waste-heat availability does not exactly match demand, excess heat is dumped, and deficit heat is supplied by the peak-heat boilers. Figure 1 shows a simplified diagram of the process with a TES tank; the waste-heat boilers (WHB) and peak-heat boilers (PHB) are lumped together and represented as one unit. Nodes A and B represent split or merging of the main water flow, depending on whether the TES tank is charging or discharging, since there is no variation of volume in the TES tank. A description of the variables is presented in the Modelling section.

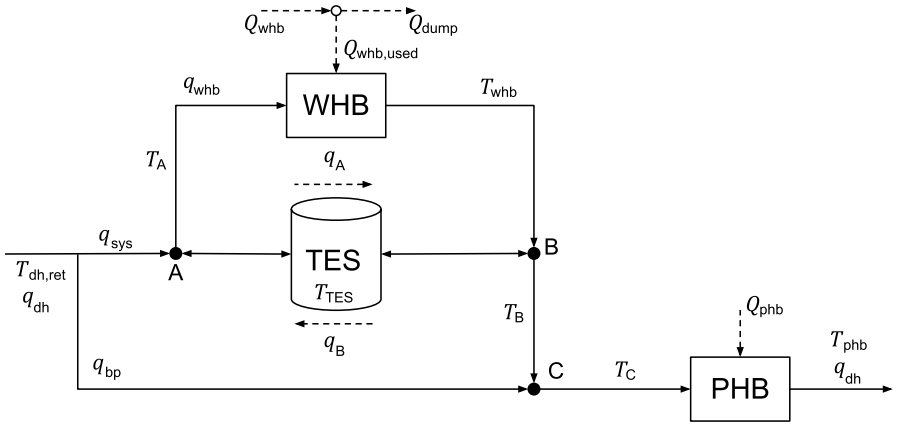


Figure 1. Flow diagram of the DH system of Mo i Rana.

2.1. Historical Data

For this case study, we selected March of 2019 as a representative month in which waste-heat availability oscillates from shortage to excess when compared against the heat demand from the city, as seen in Figure 2. This behaviour, usually seen during the transition months between summer and winter, has a potential for short-term savings, as opposed to long periods of shortage (winter) or excess (summer) of heat availability that would require long-term storage. From the DH system, we also have given the return

and supply temperatures and mass flow rate of water for every hour available as input data; the temperatures are shown in the top Figure 3.

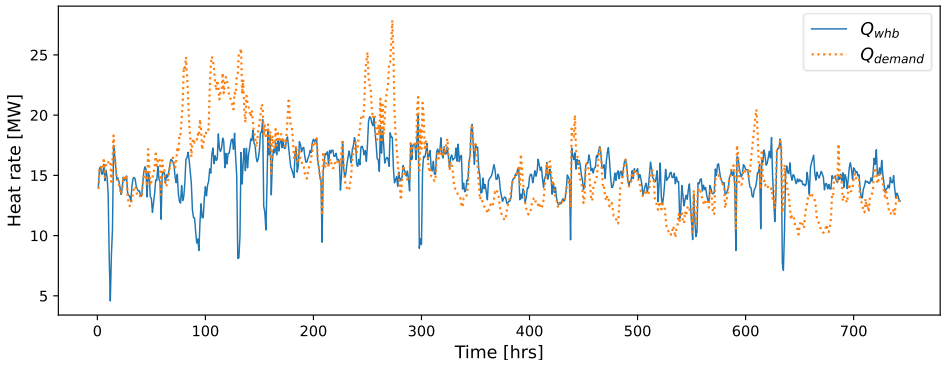


Figure 2. Historical waste-heat and heat demand data from Mo i Rana district heating system for March 2019.

3. Methodology

We formulate an optimization model to obtain the optimal volume of a TES tank for the Mo i Rana DH system taking operational conditions into account. For that, we need mass and energy balances of the process, as well as operational and cost functions that can be minimized to express our main goal. The mass and energy balances act as constraints in the model and are as follows

$$q_{\text{dh}}(t) - q_{\text{sys}}(t) - q_{\text{bp}}(t) = 0 \tag{1a}$$

$$q_{\text{sys}}(t) - q_{\text{whb}}(t) - q_{\text{A}}(t) + q_{\text{B}}(t) = 0 \tag{1b}$$

$$q_{\text{sys}}(t)C_pT_{\text{dh,ret}}(t) + q_B(t)C_pT_{\text{TES}}(t) - q_{\text{whb}}(t)C_pT_A(t) - q_A(t)C_pT_A(t) = 0 \tag{1c}$$

$$q_{\text{whb}}(t)C_pT_{\text{whb}}(t) + q_A(t)C_pT_{\text{TES}}(t) - q_{\text{sys}}(t)C_pT_B(t) - q_B(t)C_pT_B(t) = 0 \tag{1d}$$

$$q_{\text{bp}}(t)C_pT_{\text{dh,ret}}(t) + q_{\text{sys}}(t)C_pT_B(t) - q_{\text{dh}}(t)C_pT_C(t) = 0 \tag{1e}$$

$$Q_{\text{phb}}(t) - q_{\text{dh}}(t)C_p\left(T_{\text{phb}}(t) - T_C(t)\right) = 0 \tag{1f}$$

$$Q_{\text{whb,used}}(t) - q_{\text{whb}}(t)C_p\big(T_{\text{whb}}(t) - T_A(t)\big) = 0 \tag{1g}$$

$$Q_{\text{whb}}(t) - Q_{\text{whb,used}}(t) - Q_{\text{dump}}(t) = 0 \tag{1h}$$

$$\frac{d}{dt}\Big(\rho V_{\text{TES}}C_pT_{\text{TES}}(t)\Big) = q_A(t)C_p\big(T_{\text{TES}}(t) - T_A(t)\big) - q_B(t)C_p\big(T_{\text{TES}}(t) - T_B(t)\big) \tag{1i}$$

where $q_{\cdot}$ are flow rates in kg/s, $T_{\cdot}$ corresponds to the temperature at the outlet of the subscript reference in °C, C_p is the specific heat capacity of the water in kJ/(kgK), $Q_{\cdot}$ are heat rates in W, ρ is the density of the water in kg/m^3, and V_{TES} is the volume of the TES tank in m^3. It is important to point out that q_A and q_B correspond to the same flow but in opposite direction. For example, when the TES tank is charging, $q_B > 0$ and q_A must be zero, and vice versa. If we enforced this condition in the optimization model, we would get a mathematical program with complementarity constraints, which is a class of

nonconvex optimization models that can be particularly challenging to solve. To avoid that, we rely on information we have available; we enforce that, if the waste-heat available is higher than the city demand, then the tank can only be charged, i.e., $q_A = 0$ while q_B is a free positive variable. The opposite is also added as constraint to the model.

For the operational term in the objective function, we choose to minimize dumped waste-heat that could be later used during periods of low waste-heat availability. Peak-heat use, which we also wish to minimize, is considered in operational costs. The economic term in the objective function to be minimized is the payback period since it is one of the most relevant economic aspects in designing a tank. It relates both investment and operational costs, allowing for one term to account for them simultaneously and avoiding tuning separate weights.

The dynamic optimization model is then given by

$$\min_{q,V_{\mathrm{TES}}} \quad N + C\int_0^T Q_{\mathrm{dump}}(t)dt + 10^{-7}\int_0^T q_{\mathrm{whb}}dt + 10^{-5}\int_0^T q_{\mathrm{bp}}dt \tag{2a}$$

$$\text{s.t.} \quad N = \frac{ln(S/(S - I\{V\}r))}{ln(1+r)} \tag{2b}$$

$$S = n\,C\int_0^T \left(Q_{\mathrm{phb,noTES}}(t) - Q_{\mathrm{phb}}(t)\right)dt \tag{2c}$$

$$I(V) = 4.7V^{0.6218} \tag{2d}$$

$$q_A(t) = 0 \quad \text{if } Q_{\mathrm{whb}}(t) < Q_{\mathrm{demand}}(t) \tag{2e}$$

$$q_B(t) = 0 \quad \text{if } Q_{\mathrm{whb}}(t) \geq Q_{\mathrm{demand}}(t) \tag{2f}$$

$$x_{\mathrm{lb}} \leq x \leq x_{\mathrm{ub}} \tag{2g}$$

The model in Eq. (1)

where N is the payback period in years, T is the total length of the considered period in hours, $I(V)$ is an expression describing initial investment cost in 10^3 euros as a function of the volume of the tank in m^3 (Li et al., 2021), r is the annual interest rate, S is financial savings in 10^3 euros/year, n is the number of representative periods in a year, C is the cost of heat composed by the price of the energy source (in this case, C_{CO} or C_{elect}) and associated tax emissions (C_{CO_2} and C_{NO_x}), x is a vector containing all variables in the model, and x_{lb} and x_{ub} are the corresponding lower and upper bounds, respectively. The extra two terms in the objective function are regularization terms, which help the solver converge to a local solution, since the flow distribution within the DH system is not necessarily unique for some Q profiles and V_{TES}. Note that, here, C is also used as a weighting parameter for the waste-heat dump term.

Eq. (2) was discretized using implicit Euler with time step of one hour and implemented in Julia using JuMP as the mathematical modelling language (Dunning et al., 2017) and IPOPT as the nonlinear programming solver (Wächter and Biegler, 2006). Table 1 shows the values of parameters and variable bounds used for the calculation.

Table 1. Parameters and variable bounds for Eq. (2) (The cost of CO-gas is confidential).

Parameter	Value	Bounds	Value
Electricity cost, C_{elect}	€ 0.087/kWh	T lower bound	40 °C
CO_2 emission tax, C_{CO_2}	€ 58.82/t CO_2	T upper bound	120 °C
NO_x emission tax, $C_{\text{NO\_x}}$	€ 2,340.9/t NO_x	$Q_{\text{whb,used}}$ upper bound	22 MW
Annual interest rate, r	5 %	Q and q lower bound	0
Initial tank temp., $T_{\text{TES}}(0)$	95 °C	q_{whb} upper bound	333 kg/s

4. Results

The results for operational conditions considering electricity and CO-gas as peak heating were the same. In both cases, the optimal volume was 6323 m^3 and Figure 3 shows some of the optimal operational conditions. The bottom plot shows peak heating and waste heat used, as well as the peak heating use without a TES tank. The total peak heating originally used during the period considered was 876.4 MWh. With the implementation of a TES tank of the optimal volume, this consumption is reduced in 48 % in total for the period. The top plot shows the TES tank, the supply temperatures to the DH system, and the corresponding return temperature. Initially during this the month, up to around 300 h, heat demand from the city is mostly greater than waste-heat supply, so the energy initially stored in the tank is consumed. Then, the TES tank temperature increases as excess waste-heat is available and reaches the maximum temperature at the end of the period.

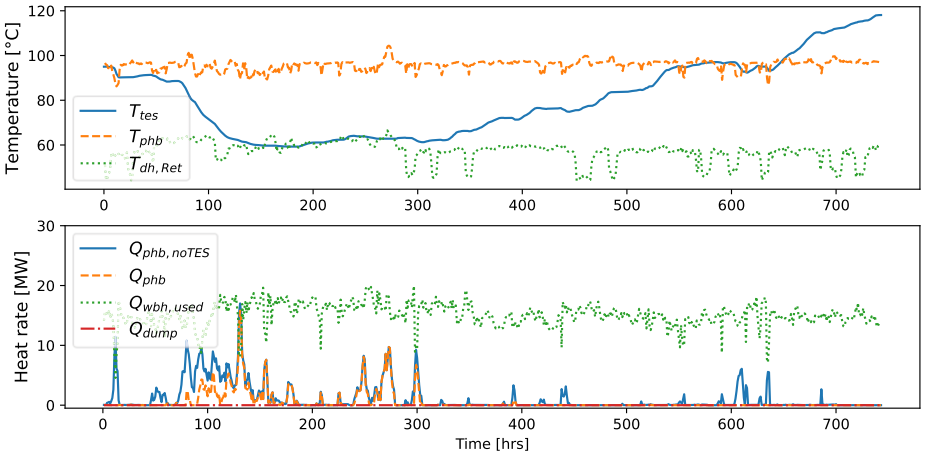


Figure 3. Optimal operation conditions for the optimal TES tank.

Regarding the economic aspect, if we consider that there are 3 months such as the representative period per year, and that the remaining months are not able to induce significant savings, the payback time for electricity as peak-heating source would be 13.7 years. Since in Norway, electricity is mainly from hydropower, the corresponding emission tax is lower. Consequently, for the case with CO-gas as peak-heating source the payback period is reduced to 12.2 years. Although these values imply large investment costs, uncertainties in the cost parameters, such as varying electricity price and emission taxes, the latter expected to increase in the next years (Klima- og Miljødepartementet, 2021), can reduce the payback time. Indeed, if the CO_2 tax is

increased to the value expected by the Norwegian government in 2030, the payback time is reduced in about half for CO-gas as peak-heating source.

Since the investment cost of the TES tank is directly related to its volume, the payback period is also dependent on it. The bottom plot of Figure 3 shows that no waste heat is discarded, i.e., $Q_{\text{dump}} = 0$, and the large volume obtained for this TES tank is due to minimizing heat dump. The weighting parameter C can be seen as a cost for dumping heat and, in this case study, we used the actual cost of peak-heating. Decreasing its value could potentially allow for some excess waste-heat to be discarded, which, in turn, could reduce the tank volume. However, that would also increase peak heating and a balance should be found.

5. Conclusions and Future Work

The results show that using a single dynamic optimization model based on operation conditions can indeed be applied to design a TES tank and systematically investigate the influence of parameters subjected to uncertainties. The calculated TES tank volume for the case study is relatively large, which is a result from the selected input data (one month), and the available price parameters. For future work, we seek to apply this optimization model to a longer horizon that can comprehend an entire season and find a systematic approach to balance storing enough heat to obtain significant savings while keeping the tank as small as possible to reduce investment costs.

Acknowledgements

This work was supported by the Research Council of Norway (RCN) through FRIPRO Project SensPATH and HighEFF - Centre for an Energy Efficient and Competitive Industry for the Future under the FME-scheme (Centre for Environment-friendly Energy Research, 257632).

References

D. Connolly, H. Lund, B.V. Mathiesen, S. Werner, B. Möller, U. Persson, T. Boermans , D. Trier, P.A. Østergaard and S. Nielsen, 2014, Heat Roadmap Europe: Combining District Heating with Heat Savings to Decarbonise the EU Energy System. Energy policy, 65, 475-489

I. Dunning, J. Huchette and M. Lubin, 2017, JuMP: A modeling language for mathematical optimization, SIAM review, 59, 2, 295-320.

Klima- og Miljødepartementet. Heilskapeleg plan for å nå klimamålet. Regjeringen. 8 Jan. 2021: https://www.regjeringen.no/no/aktuelt/heilskapeleg-plan-for-a-na-klimamalet/id2827600/. Accessed 1 July 2021.

B. Knudsen, D. Rohde, and H. Kauko, 2021, Thermal Energy Storage Sizing for Industrial Waste-Heat Utilization in District Heating: A Model Predictive Control Approach, Energy, 234, 121200.

H. Li, J. Hou, T. Hong, Y. Ding and N. Nord, 2021, Energy, Economic, and Environmental Analysis of Integration of Thermal Energy Storage into District Heating Systems Using Waste Heat from Data Centres, Energy, 219, 119582.

A. Mirandola and E. Lorenzini, 2016, Energy, Environment and Climate: From the Past to the Future. International Journal of Heat and Technology, 34, 2, 159-164

A. Wächter and L.T. Biegler, 2006. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming, Mathematical programming, 106, 1, 25-57.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50326-2

A flexible energy storage dispatch strategy for day-ahead market trading

Jude O. Ejeh[a], Diarmid Roberts[a], Solomon F. Brown[a*]

[a]*Department of Chemical & Biological Engineering, The University of Sheffield, Mappin Street, Sheffield, S1 3JD, United Kingdom*

s.f.brown@sheffield.ac.uk

Abstract

In this work, we present a two-stage optimisation-based approach to obtain key metrics for use in a rules-based energy storage dispatch strategy. In electrical power systems, electrical energy storage (EES) devices have been shown to improve power reliability, quality and reduce electricity bills in behind-the-meter applications. However, owing to problems of a prolonged pay-back period, the scheduling of these EES devices play an important role in for asset owners. Existing optimisation-based approaches heavily rely on a rigid implementation of the obtained solutions, perfect foresight, and may not perform well even when uncertainties are considered. A flexible alternative, quite common in practice, involves the use of rules to guide battery actions. In our approach, we propose a two-stage approach to determine the value of key metrics which can be used in rules-based strategy. The first stage solves a 2-step optimisation model to determine the optimal charging and discharging electricity price from previous historical data, and the second stage simulates, in real-time, the battery actions based on the price rules initially created. This proposed method was applied to a microgrid with local load and PV power generation with access to the UK Day-Ahead energy market, with results showing an improvement in electricity cost savings across board when compared with the more popular time-based rules dispatch strategy.

Keywords: Energy storage dispatch, Energy arbitrage, Optimisation

1. Introduction

In electrical power systems, electrical energy storage (EES) devices have been shown to improve power reliability, flexibility, and quality, and reduce electricity bills in front-of-meter and/or behind-the-meter applications, especially with the increased penetration of intermittent renewable energy (RE) generators (Ma et al., 2018). Owing to problems of a prolonged pay-back period in large scale deployment of these devices, quite a number of research efforts have been focused on revenue stacking – where a collection of tasks are performed by the EES device in order to generate more revenue (Roberts and Brown, 2020). In such cases, the scheduling of the EES device plays an important role both in the total revenue generated as well as the lifetime of the device.

A great deal of research has thus focused on optimising the operation of such devices under differing conditions of energy demand, energy generation sources, electricity

prices and/or accessible revenue streams (García Vera et al., 2019). Hannan et al. (2020) presented a review on existing optimisation methods/algorithms, amongst others, for EES sizing and scheduling in microgrid (MG) applications. Zia et al. (2018) also presented a critical review on methods and solutions for energy management systems - generation dispatch, frequency regulation, etc. Despite the accuracy these optimisation-based methodologies reviewed achieve given the underlying assumptions, results obtained are almost always a single set of time-dependent decision variables, which must be followed precisely to obtain the same objective function value considered in the model. In reality, these solutions require rigid implementation, and do not always perform well when conditions of energy demand, price, generation, or any other adopted data fall far from assumptions or predictions, even with uncertainty considerations. An alternative and more flexible approach towards energy dispatch scheduling involves using a rules-based approach. The rules-based approach uses a simple algorithm, iterating over each time step to determine the EES devices operating mode and action in real-time. Key metrics or system parameters, for example, the threshold prices or times during a day, which characterise the optimal operation of the EES device and thus specify periods for certain battery action (charging/discharging), are examples of rules-based strategies. Kanwar et al. (2015) compared optimisation and rules-based strategies for a MG on a time-of-use (ToU) tariff, with the former strategy obtaining just a 2% increase in savings with additional computational complexities. These rules-based methods also present an easy-to-implement strategy for EES device owners and are applicable over a wider range of system variability. They do however, have the drawback of lacking a guarantee on cost optimality. An ideal strategy will therefore comprise a combination of the cost optimality merits of the optimisation-based strategy and the ease-of-implementation and flexibility of the rules-based approach. Zhang et al. (2017) proposed rules-based strategies with some metrics determined via a linear programming (LP) problem to increase the accuracy of the rules generated. These LP problems were, however, solved at each period of operation.

To this end, we present a 2-step optimisation-based approach to obtain key metrics for use in a rules-based energy storage dispatch strategy. The key metrics to be identified in this work include the minimum/maximum prices, and thus the corresponding times, to charge/discharge an EES device in real-time. Kanwar et al. (2015) adopted a ToU tariff with known electricity price values per time which may not be applicable to wholesale market trading. Our proposed strategy further seeks to evaluate the threshold prices for rules-based battery operation. The first step solves an optimal energy dispatch optimisation model to obtain a set of distinct optimal solutions, and the second is a feature extraction stage which finds the optimal price range where charge and discharge actions are executed so as to minimise the total electricity cost of the MG. The approach is applied to a MG with an EES asset having access to the UK day-ahead energy market. Given a load demand which it must satisfy, this approach proffers the optimal price range (and times) for charging and discharging the EES device to minimise the electricity cost.

In the rest of the paper, the proposed strategy and associated mathematical formulation of the mixed integer linear programming (MILP) models are described in

section 2. The strategy is then applied to a case study is section 3., with findings discussed and some conclusions drawn in section 4.

2. Methodology

Figure 1 gives a flow diagram of the proposed strategy. The first stage involves two optimisation models. The first model (an MILP scheduling model - MILP_SH) minimises the total electricity cost of the MG. The second model (an MILP selection model - MILP_SL) determines similar battery actions from a set of distinct optimal solutions generated from MILP_SH, by maximising the number of similar battery actions.

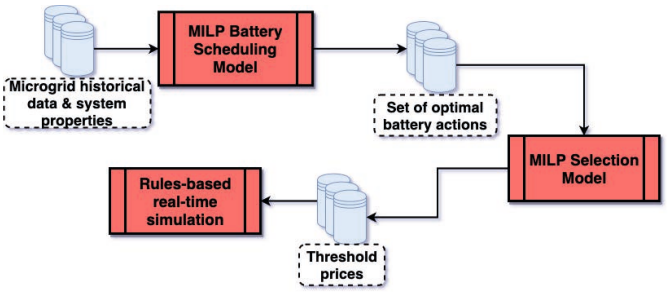


Figure 1: Proposed rules-based strategy

2.1. Problem description & assumptions

The problems solved by these models are described as follows. For the MILP scheduling model (MILP_SH), *given* a microgrid (MG) with known local electricity demands over time (D_t^B), local photovoltaic (PV) power generation (G_t), and an EES device (battery) with a known initial state of charge (SOC, SOC^0), maximum power output (P^{max}), safety capacity ranges (SOC^{min}, SOC^{max}) and charge (η^C) and discharge (η^D) efficiencies; with direct access to the UK Day-Ahead (DA) markets with known historical buy (ρ_t^I) and sell (ρ_t^E) prices;
Determine the optimal schedule of the EES device - charging and discharging times, power (P_t^{s+}, P_t^{s-}) and SOC (SOC_t) - *so as to* minimise the total cost of electricity for the MG.
For the MILP selection model (MILP_SL), *given* a set of differing optimal solutions from the MILP_SH model, and their corresponding battery charging and discharging actions, *determine* the set of common battery charging and discharging actions per time amongst a subset of optimal solutions, *so as to* maximise the total number of similar battery actions.
The corresponding buy and sell electricity prices associated with the solution of MILP_SL model provides the threshold price ranges for charging and discharging the ESS which are used in a price-based rules strategy. It is assumed that the MG acts as a price taker, with its electricity demand not having an impact on the DA market prices.

2.2. Mathematical Formulation

The objective of the proposed MILP_SH model given by eq. (1) is to minimise the total electricity cost subject to eqs. (2) - (9).

$$min \quad \sum_{t} \Delta\left(\rho_t^I \cdot P_t^I - \rho_t^E \cdot P_t^E\right) \tag{1}$$

subject to:

$$SOC^{min} \leq SOC_t \leq SOC^{max} \qquad \forall\, t \tag{2}$$

$$SOC_t = SOC^0 \mid_{t=0} + SOC_{t-1} \mid_{t>0} + \Delta \cdot (\eta^C P_t^{s+} - \frac{P_t^{s-}}{\eta^D}) \qquad \forall\, t \tag{3}$$

$$P_t^I - P_t^E = \frac{D_t^B}{\Delta} + P_t^{s+} - P_t^{s-} - \frac{G_t}{\Delta} \qquad \forall\, t \tag{4}$$

$$P_t^I \leq M \cdot (1 - B_t^N) \qquad \forall\, t \tag{5}$$

$$P_t^E \leq M \cdot B_t^N \qquad \forall\, t \tag{6}$$

$$B_t^C + B_t^D \leq 1 \qquad \forall\, t \tag{7}$$

$$P_t^{s+} \leq P^{max} \cdot B_t^C \qquad \forall\, t \tag{8}$$

$$P_t^{s-} \leq P^{max} \cdot B_t^D \qquad \forall\, t \tag{9}$$

Eq. (2) ensures that the SOC of the battery is within predefined safety limits. The SOC at time t is evaluated using eq. (3) as the net charging/discharging action with respect to its initial SOC, where Δ represents the time step. The total power imported/exported is determined by eq. (4) as the local energy demand less the amount of solar and battery energy generated. At any given time, energy may only be imported or exported (eqs. (5) and (6); $B_t^N \in \{0, 1\}$) to/from the MG. Finally, the battery cannot simultaneously charge and discharge (eq. (7); $B_t^C, B_t^D \in \{0, 1\}$), and its output power must not exceed its rating (eqs.. (8) - (9)).

Distinct optimal (& near-optimal) solutions for MILP_SH are obtained by including the integer cut given by eq. (10) where τ is an index denoting the saved optimal solution in question, U_τ^C and U_τ^D represent the set of time periods in which B_t^C and B_t^D equal 1 respectively; L_τ^C and L_τ^D represent the set of time periods in which B_t^C and B_t^D equal 0 respectively; and σ is a positive integer which represents the degree of variation between generated optimal solutions. MILP_SH model with eq. (10) included is solved repeatedly, updating the sets $U_\tau^C, U_\tau^D, L_\tau^C, L_\tau^D$ in order to generate a set of distinct optimal solutions, $\mathcal{T}$.

$$\sum_{t \in U_\tau^C} B_t^C + \sum_{t \in U_\tau^D} B_t^D - \sum_{t \in L_\tau^C} B_t^C - \sum_{t \in L_\tau^D} B_t^D \leq \mid U_\tau^C \mid + \mid U_\tau^D \mid - \sigma \qquad \forall\, \tau \in \mathcal{T} \tag{10}$$

The objective of the proposed MILP_SL model given by eq. (11) is to maximise the total number of similar battery actions (charging, β_t^C, and discharging, β_t^D) amongst distinct optimal solutions subject to eqs. (12) - (14).

$$max \quad \sum_{t} \beta_t^C + \beta_t^D \tag{11}$$

subject to:

$$\sum_{\tau} \overline{B}_{t\tau}^C + \mid \mathcal{T} \mid (1 - \beta_t^C) \geq \psi \qquad \forall\, t \tag{12}$$

$$\sum_{\tau} \overline{B}^{D}_{t\tau} + \mid \mathcal{T} \mid (1 - \beta^{D}_{t}) \geq \psi \qquad \forall\, t \tag{13}$$

$$\beta^{C}_{t} + \beta^{D}_{t} \leq 1 \qquad \forall\, t \tag{14}$$

$\overline{B}^{C}_{t\tau}$ and $\overline{B}^{D}_{t\tau}$ denote the saved battery charging and discharging actions for each generated optimal solution τ at time t respectively. Eqs. (12) - (14) ensure that a battery action is evaluated as similar only if it occurs in some predefined minimum number of distinct optimal solutions ψ. Eq. (14) is a feasibility constraint that ensures that whatever similar battery actions are selected by the model, the battery still does not charge and discharge at the same time period. The threshold price range for charging and discharging is then obtain by matching the buy and sell electricity prices with the optimal values of β^{C}_{t} and β^{D}_{t} for each time period respectively.

3. Case study

For a case study, a MG with local load, a 3kW PV cell and a 25kWh, 16.7kW Li-ion battery with a round-trip efficiency of 81% was considered. The MG has direct access to the UK DA energy market to purchase/sell electricity. Local electricity demand was generated from ELEXON's 10-year average for 2018 and 2019. Each of the proposed optimisation models were solved using Pyomo 5.6.8 with Gurobi 9.0 to a 0% relative gap using an Intel Xeon E-2146G with 32GB and 4 threads running Windows 10.

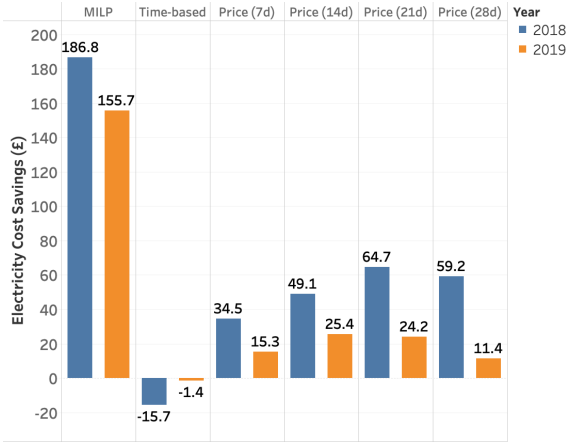


Figure 2: Annual electricity cost savings

Six scenarios were explored to demonstrate the impact of the proposed strategy. First, the MILP_SH model was solved as is using the historical data to provide a basis for the maximum electricity savings possible for the MG assuming perfect foresight (MILP). A rules-based approach based on set times of charging and discharging the battery in any day (Time-based) was simulated to reflect a popular approach used by asset owners. Finally, using the proposed 2-stage approach four additional scenarios were solved differing in how many historical days were used by models MILP_SH and MILP_SL to obtain the threshold prices for the price-based rules approach. "Price (7d)" thus consisted of a scenario where for each 7-day

period simulated in a year using the price-based rules strategy, the threshold prices were generated using the previous 7-days load, PV and energy prices data. In each of the rules-based strategies, the algorithm was implemented over a 7-day period. Figure 2 shows the electricity savings for each of the scenarios solved for the years 2018 and 2019 respectively. In both years, results show that the time-based approach (charging at night and discharging at peak hours in the afternoon) made a loss when compared to scenarios without any battery installed. All cases using the proposed rules-based strategy obtained electricity savings in varying degrees. The maximum electricity cost savings were however obtained from generating threshold prices using 14 or 21 days of historical data.

4. Conclusions

In this work, a flexibile energy storage dispatch strategy was proposed for DA market trading by a MG. The strategy consisted of a two-stage approach. In the first stage, threshold buy and sell prices required for real-time simulation of battery actions were obtained by generating multiple optimal solutions for an MILP scheduling model which minimises the total electricity cost of the MG. The MG was assumed to have a local load, PV generation, and an installed battery with known capacity with access to the UK DA market. Next, an MILP selection model, is used to extract the threshold prices by maximising the total number of similar battery actions from the set of optimal solutions. The first stage thus uses historical data to generate the threshold prices which are used in a real-time simulation of battery actions in the second stage. This strategy was applied to a MG with a 3kWp PV system and 25kWh, 16.7kW Li-ion battery for years 2018 and 2019. Results showed that up to 34% of the maximum possible electricity cost saving was captured, exceeding the time-based simulations popularly adopted which made a loss in both years. As future work, additional metrics to improve the performance of the proposed rules-based strategy will be included.

References

García Vera, Y. E., Dufo-López, R., Bernal-Agustín, J. L., 2019. Energy Management in Microgrids with Renewable Energy Sources: A Literature Review. Applied Sciences 9, 3854.

Hannan, M., Faisal, M., Jern Ker, P., Begum, R., Dong, Z., Zhang, C., 2020. Review of optimal methods and algorithms for sizing energy storage systems to achieve decarbonization in microgrid applications. Renewable and Sustainable Energy Reviews 131, 110022.

Kanwar, A., Hidalgo Rodriguez, D. I., Von Appen, J., Braun, M., 2015. A comparative study of optimization and rule-based control for microgrid operation. Power and Energy Student Summit, 1–6.

Roberts, D., Brown, S. F., 2020. Identifying calendar-correlated day-ahead price profile clusters for enhanced energy storage scheduling. Energy Reports 6, 35–42.

Zhang, Y., Lundblad, A., Campana, P. E., Benavente, F., Yan, J., 2017. Battery sizing and rule-based operation of grid-connected photovoltaic-battery system: A case study in Sweden. Energy Conversion and Management 133, 249–263.

Zia, M. F., Elbouchikhi, E., Benbouzid, M., 2018. Microgrids energy management systems: A critical review on methods, solutions, and prospects. Applied Energy 222, 1033–1055.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50327-4

Monetizing Flexibility in Day-Ahead and Continuous Intraday Electricity Markets

Niklas Nolzen[ab], Alissa Ganter[ab], Nils Baumgärtner[d], Ludger Leenders[a], André Bardow[abc*]

[a]*Energy & Process Systems Engineering, Department of Mechanical and Process Engineering, ETH Zurich, Tannenstrasse 3, 8092 Zurich, Switzerland*
[b]*Institute of Technical Thermodynamics, RWTH Aachen University, Schinkelstraße 8, 52062 Aachen, Germany*
[c]*Institute of Energy and Climate Research - Energy Systems Engineering (IEK-10), Forschungszentrum Jülich GmbH, Wilhelm-Johnen-Straße, 52425 Jülich, Germany*
[d]*Energy Trading and Dispatching, Currenta GmbH & Co. OHG, Kaiser-Wilhelm-Allee 80, 51373 Leverkusen, Germany*
abardow@ethz.ch

Abstract

The rising share of renewable energies increases supply uncertainty in the energy system. To make short-term adjustments more cost-efficient, the continuous intraday market was introduced. The continuous intraday market allows flexible capacity to exploit the price volatilities by asset-backed trading. In asset-backed trading, flexible capacity is continuously traded depending on the real-time electricity price and the marginal cost for electricity production. However, the flexibility for the continuous intraday market needs already be considered during the commitment on the day-ahead market. Hence, this paper proposes an optimal joint bidding strategy for day-ahead and continuous intraday market participation. For this purpose, we employ option-price theory and stochastic optimization. A case study for a flexible multi-energy system shows savings of 11 % by participating in both markets compared to only the day-ahead market. Thus, the bidding strategy provides efficient decision support in short-term electricity markets.

Keywords: electricity markets, stochastic optimization, optimal bidding strategy, spot markets, continuous trading

1. Introduction

The expansion of renewable energies increases supply uncertainty in the electricity grid. To make short-term adjustments in the grid more economical, several European countries introduced the continuous intraday market that settles imbalances with continuous real-time trading (Koch and Hirth, 2019). The continuous intraday market is characterized by strong electricity price volatilities driven by updates in the renewables forecast (Kremer et al., 2020). These price volatilities can be monetized by asset-backed trading (Löhndorf and Wozabal, 2021). Asset-backed trading uses flexible capacity. Its electricity output is continuously traded depending on the current electricity price and the marginal cost for electricity production. When electricity prices rise above marginal cost, electricity from flexible capacity is sold. When prices fall below marginal cost, electricity is purchased on the market. Hence, asset-backed trading is particularly valuable when the electricity price fluctuates around the marginal costs. Eventually, the sum of all trades determines the actual electricity output of the flexible capacity.

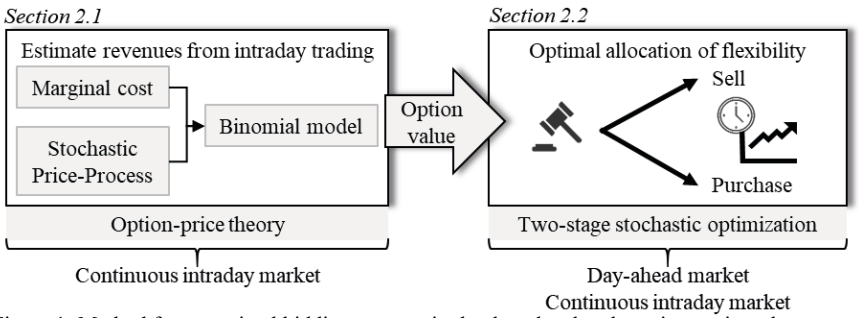


Figure 1: Method for an optimal bidding strategy in the day-ahead and continuous intraday market. First, the option value derives the revenue from trading flexible capacity in the continuous intraday market. Second, a two-stage stochastic optimization models the sequential decision-making process to optimally allocate flexibility in both markets.

Participation in the day-ahead and continuous intraday market is a sequential decision-making process. The day-ahead market clears first. Afterward, trading in the continuous intraday market is possible until shortly before delivery. An optimal bidding strategy considers both markets simultaneously to determine an optimal amount of flexible capacity for asset-backed trading. However, before day-ahead market clearing, only limited information is available on the continuous intraday market.

Due to its recent introduction, only a few studies investigate the continuous intraday market as a trading opportunity. Garnier and Madlener (2015) propose a bidding strategy for renewable energies participating in the continuous intraday market. Based on a multi-period lattice, intraday market participation is simulated using dynamic programming. Corinaldesi et al. (2020) analyze the flexibility of end-user technologies in the day-ahead and continuous intraday market. Intraday market trading is considered with an hourly updated rolling horizon with new price forecasts. Additional intraday market participation saves around 8 %. However, the rolling horizon limits the interaction between both markets as only the first optimization considers the day-ahead market participation.

This paper proposes a method that considers trading in the continuous intraday market in a multi-market optimization. The method determines an optimal bidding strategy for a flexible energy system in the day-ahead and continuous intraday market.

2. Method for optimal bidding strategies in day-ahead and intraday markets

The method allocates the flexible capacity of a market participant to the day-ahead and continuous intraday market to determine an optimal bidding strategy (Figure 1). In sequential decision making, the value of the intraday market opportunity needs to be considered while deciding on the day-ahead market participation. In Section 2.1, the value of trading in the continuous intraday market is derived based on forecast data using option-price theory (Björk, 2009). In Section 2.2, a two-stage stochastic optimization optimizes flexible capacity allocation to the day-ahead and continuous intraday market.

2.1. Deriving the option value for the continuous intraday market

Here, we propose to estimate the revenues from trading in the continuous intraday market with the option value. At the time of day-ahead market clearing, the joint market participation requires knowledge on the revenues in the continuous intraday market. The option-price theory allows estimating the revenues based on parameters available one day before delivery, i.e., at the time of commitment to the day-ahead market.

The option value is derived with the multiperiod binomial model based on Cox et al. (1979). In the multiperiod binomial model (Figure 2), the trader adjusts the

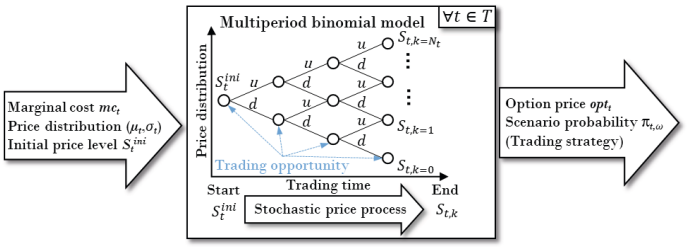


Figure 2: Calculation of revenues from trading in the continuous intraday market using the multiperiod binomial model from option-price theory. Therein, the risk-neutral trading strategy can be derived at each trading opportunity to replicate the option value for each traded hour $t \in T$.

purchased/sold share of flexible capacity at each trading opportunity to realize the estimated option value within a trading session. This adjustment is based on the stochastic price process and the marginal costs for electricity production, assuming a risk-neutral trader. Hence, we refer to this strategy as the risk-neutral asset-backed trading strategy. Using this strategy, option price theory assumes that the option value is realized independent of the price scenario. Therefore, the price volatility in the continuous intraday market is monetized without any financial risk. Therein, we assume that the stochastic price process sufficiently captures the price volatility.

For the sake of simplicity, we consider only trading hourly electricity contracts in the continuous intraday market, thus neglecting other trading opportunities such as half-hourly and quarter-hourly products. The trading session for each traded hour $t \in T$ starts after clearing the day-ahead market and ends shortly before delivery. In each trading session, we discretize continuous trading with N_t trading opportunities. At each trading opportunity, the adjustment of the trading position follows the risk-neutral asset-backed trading strategy. Furthermore, the marginal costs mc_t are known for the flexible capacity. Each trading session starts with the forecasted initial price-level S_t^{ini}. Then, the electricity price follows a stochastic price process modeled as arithmetic Brownian motion (Alexander et al., 2012). In contrast to the often used geometric Brownian motion, the arithmetic Brownian motion models the absolute price change (in €/MWh). Hence, the stochastic price process can also lead to negative electricity prices. Assuming the arithmetic Brownian motion throughout a trading session, the price moves up u_t or down d_t at each trading opportunity. The up-movement u_t and down-movement d_t are determined as follows:

$$u_t = \mu_t \cdot \frac{1}{N_t} + \sigma_t \cdot \sqrt{\frac{1}{N_t}} \quad \text{and} \quad d_t = \mu_t \cdot \frac{1}{N_t} - \sigma_t \cdot \sqrt{\frac{1}{N_t}} \; \forall t \in T, \tag{1}$$

by multiplying the price drift μ_t and the price volatility σ_t with the trading frequency $\frac{1}{N_t}$. Hence, the price drift μ_t and the price volatility σ_t are allocated over the trading session. At the end of a trading session, the last electricity price $S_{t,k}$ deviates from the initial price-level S_t^{ini}. In Eq. (1), we assume that the absolute price deviation from the initial price-level S_t^{ini} is normally distributed with $\mathcal{N}(\mu_t, \sigma_t^2)$. The arithmetic Brownian motion determines the last price as the summation of up-movements and down-movements. In the stochastic price process, $N_t + 1$ last prices $S_{t,k}$ are possible for each traded hour:

$$S_{t,k} = S_t^{ini} + k \cdot u_t + (N_t - k) \cdot d_t \; \forall t \in T, k \in \{0,1,\dots,N_t\}. \tag{2}$$

The option value $opt_t^{sell/pu}$ for positive $(sell)$ and negative (pu) flexible capacity for each traded hour t is derived with Eq. (3) for the multiperiod binomial model:

$$opt_t^{sell/pu} = \sum_{k=0}^{N_t} \underbrace{\binom{N_t}{k}}_{(1)} \cdot \underbrace{\left(\frac{-d_t}{u_t - d_t}\right)^k}_{(2)} \cdot \underbrace{\left(\frac{u_t}{u_t - d_t}\right)^{N_t - k}}_{(3)} \cdot \underbrace{\Phi^{sell/pu}(S_{t,k})}_{(4)} \; \forall t \in T. \tag{3}$$

Eq. (3) is adapted to the continuous intraday market, assuming a zero interest rate due to the short-term nature of the market. In Eq. (3), the binomial coefficient (1) derives the absolute frequency that the last price k is reached. The terms (2) and (3) are the martingale measures, i.e., the risk-neutral probabilities, for an up-movement and down-movement, respectively. The term (4) evaluates the option value $\Phi^{sell/pu}$ at the end of a trading session for the sell option and the purchase option for each traded hour t and last price k:

$$\Phi^{sell}(S_{t,k}) = \begin{cases} S_{t,k} - mc_t, if\ S_{t,k} > mc_t \\ 0, \qquad\quad if\ S_{t,k} \leq mc_t \end{cases} \text{ and } \Phi^{pu}(S_{t,k}) = \begin{cases} 0, \qquad\quad if\ S_{t,k} \geq mc_t \\ mc_t - S_{t,k}, if\ S_{t,k} < mc_t \end{cases}.$$

In summary, the option value sums up and weights the revenues from all last prices. Therein, the scenario probabilities $\pi_{t,\omega}$ for utilizing positive and negative flexibility are obtained from the stochastic price process and the marginal cost. The derived option value $opt_t^{sell/pu}$ accounts for the revenues from asset-backed trading. Both parameters are the input parameters for the two-stage stochastic optimization presented in the following.

2.2. Two-stage stochastic optimization of day-ahead and intraday market participation

The two-stage stochastic optimization derives an optimal bidding strategy in the day-ahead and continuous intraday market. Hence, at the 1st stage, decisions are made to buy or sell electricity in the day-ahead market. Moreover, flexible positive capacity (sell option) and negative capacity (purchase option) are blocked for trading in the continuous intraday market. At the 2nd stage, trading in the continuous intraday market ends, and the energy system operation is adapted according to intraday market trading. The 2nd stage considers two scenarios $\omega \in \Omega = \{S < mc\ ,\ S > mc\}$ that depend on the marginal costs for electricity production mc and the electricity price S at the end of a trading session. In the scenarios, we assume that trading flexibility in the continuous intraday market utilizes the blocked positive and negative flexibility.

The objective function of the two-stage stochastic optimization minimizes the expected costs consisting of operational costs $C_{t,\omega}^{op}$ subtracting the revenues from the day-ahead market R_t^{DA} and the continuous intraday market $R_{t,\omega}^{ID}$:

$$min \sum_{t \in T} \sum_{\omega \in \Omega} \pi_{t,\omega} \cdot \left(C_{t,\omega}^{op} - R_t^{DA} - R_{t,\omega}^{ID}\right) \tag{4}$$

$$s.t.\ energy\ balances\ \forall\ \omega \in \Omega, t \in T, e \in energy\ forms \tag{5}$$

$$market\ constraints\ \forall \omega \in \Omega, t \in T, market \in \{DA, ID\} \tag{6}$$

$$technical\ constraints\ \forall \omega \in \Omega, t \in T, u \in units \tag{7}$$

Therein, the costs and revenues are weighted with the scenario probability $\pi_{t,\omega}$. Furthermore, the energy system is operated to fulfill the product balances and to comply with the constraints set by market participation. In both scenarios $\omega \in \Omega$, trading flexible capacity $ID^{sell/pu}$ in the intraday market is reimbursed with the respective option value for positive and negative flexibility. Hence, the revenues $R_{t,\omega}^{ID}$ can be expressed with

$$R_{t,\omega}^{ID} = ID_{t,\omega}^{sell}(opt_t^{sell} + s_\omega^{sell} \cdot mc_t) + ID_{t,\omega}^{pu}\left(opt_t^{pu} - s_\omega^{pu} \cdot mc_t\right) \forall t \in T, \omega \in \Omega, \tag{8}$$

whereas s_ω^{sell} (s_ω^{pu}) is a binary parameter that is 1 if electricity is sold (purchased) in the continuous intraday market or 0 if no electricity is sold (purchased) in the continuous intraday market. By following the risk-neutral asset-backed trading strategy, the option value for positive and negative flexibility $opt_t^{sell/pu}$ is realized independently from the last price in the intraday market. Hence, the arising (saved) endogenous generation costs are compensated in Eq. (8) for the sell (purchase) option in scenario $S > mc$ ($S < mc$). Finally, the stochastic optimization models trading in the continuous intraday market as a series of sell and purchase options, whereas the sell option (purchase option) models the positive (negative) flexibility of the energy system.

3. Case study: market participation of a multi-energy system

The method is applied to a multi-energy system based on Baumgärtner et al. (2019). The multi-energy system consists of 4 combined-heat-and-power engines, 2 electrode boilers, 3 adsorption chillers, 3 compression chillers, and 4 gas boilers. The multi-energy system covers time-varying demands for electricity, heating, and cooling. The multi-energy system participates in the day-ahead and continuous intraday market in Germany.

The case study is conducted for Wednesday, October 9th, 2019. The day is chosen because the average day-ahead market price is in the same range as the marginal costs. Hence, asset-backed trading in the continuous intraday market is particularly interesting.

The day-ahead market is assumed to be deterministic with time-varying electricity prices taken from Bundesnetzagentur | SMARD.de (2021). For the calculation of the option value, the stochastic price process and the marginal costs are modeled as follows: the day-ahead market price approximates the initial price-level S_t^{ini}. The price drift μ_t and the price volatilities σ_t are derived using historical data from the years 2019 and 2020 from EPEX SPOT based on the ID3 price (EPEX SPOT, 2021). For each traded hour of 2019 and 2020, the price deviation between the continuous intraday market and the day-ahead market is calculated as the difference between ID3 price and day-ahead market price. All price deviations are clustered based on the hourly wind generation forecast, solar generation forecast, and residual load forecast available at ENTSO-E (2021). In each cluster, a normal distribution is fitted to obtain the price drift μ and the price volatility σ. Finally, each traded hour is matched with the respective cluster and u_t and d_t are derived based on Eq. (1). The marginal costs mc_t are derived from operational optimizations with varying electricity demands but without market participation for each traded hour. A linear regression is applied to determine the operational costs as a function of the electricity demands. The slopes are the marginal costs mc_t.

Two cases are compared for market participation: participation only in the day-ahead market (DA) and participation in the day-ahead and continuous intraday market (DA, ID). Both cases are solved to optimality with the solver Gurobi 9.1.1 in less than 5 minutes. Our method lowers expected operational expenditures by 11 % by intraday market participation (DA, ID) compared to only participating in the day-ahead market (DA).

Figure 3 shows the electricity exchange with the day-ahead and continuous intraday market. In case (DA), electricity is sold if the electricity price is higher than the marginal costs and purchased vice versa. In case (DA, ID), the commitment in the day-ahead market differs from the case (DA). For some traded hours, no electricity is purchased or sold in the day-ahead market. Thereby, the multi-energy system uses its positive and negative flexibility in the continuous intraday market. For the remaining traded hours, selling electricity on the day-ahead market maximizes negative flexibility. Overall, the flexibility is used for asset-backed trading in the continuous intraday market, while the

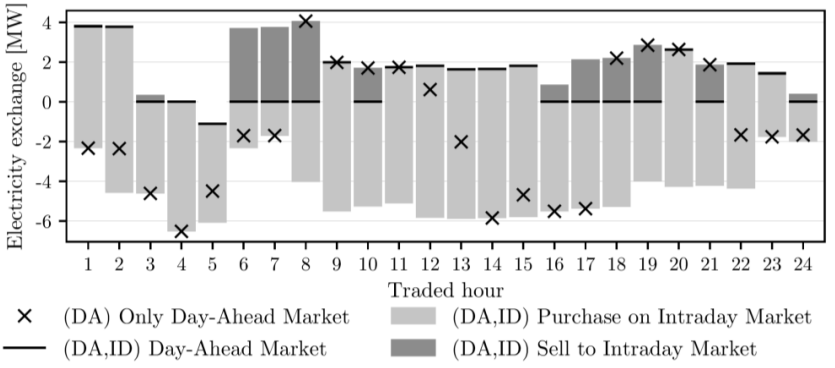


Figure 3: Electricity exchange with the day-ahead and continuous intraday market. In case (DA), electricity is only delivered to the day-ahead market. In case (DA, ID), two scenarios arise as the exchange of electricity in the continuous intraday market depends on the last price in the continuous intraday market S compared to the marginal costs mc.

last price in the continuous intraday market determines the overall physical delivery of electricity. Hence, utilizing the flexibility of the multi-energy system lowers operational expenditures. Finally, adapting the operation of the multi-energy system realizes the flexibility provision while still satisfying the internal heating, cooling, and electricity demand.
Hence, participation in the continuous intraday market allows marketing the flexibility of the multi-energy system. Overall, the method shows the ability to optimally deploy the flexibility of an energy system in the markets.

4. Conclusions

This paper presents a method based on option-price theory and two-stage stochastic optimization to derive an optimal bidding strategy for the day-ahead and continuous intraday market. The option-price theory derives the value of trading in the continuous intraday market based on forecast data. Afterward, the estimated option value serves as the input parameter for the two-stage stochastic optimization. The two-stage stochastic optimization models the sequential bidding process of joint day-ahead and continuous intraday market participation. The proposed method is applied to a case study of a multi-energy system. In this case study, savings of 11 % are expected by the proposed method for a selected day. Overall, the method is an efficient decision-making tool for operational optimization one day ahead of delivery, incorporating the complex market structure of the day-ahead and continuous intraday market.

Acknowledgements

This study is funded by the German Federal Ministry of Economic Affairs and Energy (ref. no.: 03EI1015A). The support is gratefully acknowledged.

References

D.R. Alexander, M. Mo, A.F. Stent, 2012, Arithmetic Brownian motion and real options, European Journal of Operational Research, 219, 1, 114–122, 10.1016/j.ejor.2011.12.023.
N. Baumgärtner, R. Delorme, M. Hennen, A. Bardow, 2019, Design of low-carbon utility systems: Exploiting time-dependent grid emissions for climate-friendly demand-side management, Applied Energy, 247, 2, 755–765, 10.1016/j.apenergy.2019.04.029.
T. Björk, 2009, Arbitrage theory in continuous time, 3rd ed., Oxford University Press, Oxford.
Bundesnetzagentur | SMARD.de, 2021, Marktdaten, https://www.smard.de/home/downloadcenter/download-marktdaten.
C. Corinaldesi, D. Schwabeneder, G. Lettner, H. Auer, 2020, A rolling horizon approach for real-time trading and portfolio optimization of end-user flexibilities, Sustainable Energy, Grids and Networks, 24, 01, 100392, 10.1016/j.segan.2020.100392.
J.C. Cox, S.A. Ross, M. Rubinstein, 1979, Option Pricing: A Simplified Approach, Journal of Financial Economics, 7, 229–263.
ENTSO-E, 2021, ENTSO-E: Transparency Platform, https://transparency.entsoe.eu/.
EPEX SPOT, 2021, Data of the ID3 index of the continuous intraday market.
E. Garnier, R. Madlener, 2015, Balancing forecast errors in continuous-trade intraday markets, Energy Syst, 6, 3, 361–388, 10.1007/s12667-015-0143-y.
C. Koch, L. Hirth, 2019, Short-term electricity trading for system balancing: An empirical analysis of the role of intraday trading in balancing Germany's electricity system, Renewable and Sustainable Energy Reviews, 113, 109275, 10.1016/j.rser.2019.109275.
M. Kremer, R. Kiesel, F. Paraschiv, 2020, An Econometric Model for Intraday Electricity Trading, Philosophical Transactions of the Royal Society A., Forthcoming, 10.2139/ssrn.3489214.
N. Löhndorf, D. Wozabal, 2021, Gas storage valuation in incomplete markets, European Journal of Operational Research, 288, 1, 318–330, 10.1016/j.ejor.2020.05.044.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50328-6

Planetary boundaries analysis of Fischer-Tropsch Diesel for decarbonizing heavy-duty transport

Margarita A. Charalambous[a], Juan D. Medrano-Garcia[a], Gonzalo Guillén-Gosálbez[a*]

[a]*Institute for Chemical and Bioengineering, Department of Chemistry and Applied Biosciences, ETH Zürich, Vladimir-Prelog-Weg 1, 8093 Zürich, Switzerland*
**gonzalo.guillen.gosalbez@chem.ethz.ch*

Abstract

Here we evaluated Fischer Tropsch-diesel (FT-diesel) use in heavy-duty trucks based on various production pathways differing in the CO_2 and H_2 provenance. To better understand the global environmental implications of fuelling heavy-duty trucks (HD trucks) with FT-diesel, we quantified environmental impacts over the entire life cycle using seven Planetary Boundaries (PBs) regulating the Earth's resilience. Our environmental assessment follows a well-to-wheel scope with the functional unit based on the global annual freight demand. The baseline scenario corresponds to the conventional fossil fuel. Our results show that the fossil fuel alternative is unsustainable as it transgresses the climate change PBs. Using FT-diesel based on captured CO_2 could help operate within the safe operating space but it could induce critical burden-shifting if the CO_2 and H_2 sources are not adequately selected.
Keywords: Energy, Food and Environmental Systems

1. Introduction

In recent years, liquid fuels based on renewable carbon that can substitute conventional ones with minimal changes to current infrastructure have attracted increasing interest. Notably, fossil diesel can be replaced with "drop-in" fuels with similar or better characteristics. FT-diesel is a promising alternative to fossil diesel due to the high cetane number and improved properties with the potential to optimize combustion efficiency and decrease emissions.

So far, studies related to FT-diesel have focused mainly on the production process based on biomass as a raw material (Martín and Grossmann 2011), and very few on the CO_2-based production process (Al-Yaeeshi et al. 2019). Environmental assessments of FT-diesel often quantify impacts based on conventional life cycle assessment (LCA) metrics (Holmgren and Hagberg 2009; Wernet et al. 2016), which are hard to interpret due to the absence of thresholds that can classify the studied systems as environmentally unsustainable. Hence, the absolut environmental sustainability implications of this fuel remain unclear.

Here we evaluated FT-diesel use in HD trucks based on various production pathways differing in the provenance of the raw material using seven PBs. The PBs concepts developed initially by Rockström et al., 2009, provides a framework to carry out absolute environmental sustainability assessments considering the Earth's carrying capacity. The framework considers 11 control variables linked to nine Earths' biophysical subsystems or processes. These include climate change, stratospheric ozone depletion, ocean acidification, biogeochemical flows of nitrogen and phosphorus, land system change,

freshwater use, biosphere integrity, atmosphere aerosol loading, and introduction of novel entities. All the PBs jointly establish the so-called safe operating space (SOS) for humanity. Consequently, for a scenario to be regarded as sustainable, none of the planetary boundaries should be transgressed. In essence, referring the LCA results to the safe operating space (SOS) delimited by these environmental guardrails facilitates the interpretation phase, particularly when evaluating systems that can be potentially deployed at a large scale. Based on this concept, we conducted an absolute environmental assessment over the global annual freight demand (33 trillion tkm), covering eight FT-diesel HD truck scenarios while benchmarking them against the fossil diesel HD truck counterpart –business as usual (BAU scenario)–. Hence, going well beyond standard LCAs, we analyze whether the FT-diesel trucks would help humanity operate safely within the PBs.

2. Methodology

2.1. Life cycle assessment and planetary boundaries

Following the general life cycle assessment (LCA) methodology, we carried out an environmental assessment based on the ISO 14040/44 framework using the SimaPro 9.0 software. The environmental assessment aims to assess the absolute sustainability of fuelling the global freight activities with FT-diesel from various sources. The functional unit corresponds to the global annual tkm demand for on-road HD truck activities, estimated by the International Energy Agency to be around $33 \cdot 10^{13}$ tkm.

For our analysis, we adopted a well-to-wheel scope using an attributional approach. The system boundaries cover all the upstream activities, from the production of H_2 and CO_2, through the FT-diesel synthesis, to the fuel combustion in HD trucks. We estimated the life cycle inventory for FT-diesel by assuming that most of the emissions can be attributed to CO_2 and H_2, as shown in Galán-Martín et al. (2021) for various bulk chemicals.

In order to construct the life cycle inventory, we first simulated the production of FT-diesel. Our calculations are based on the works of Shafer et al. (2019) for the FT-reactor, and Tomasek et al. (2020), for the wax hydrocracker. The Anderson-Shulz-Flory distribution (a) of choice for maximum diesel production is 0.88, and H_2/CO equals 2, based on which we calculated the product distribution of the FT-reactor. To calculate the total CO_2 and H_2 needed for the process, we assume that all the CO is converted into C_1-C_{22} in the FT-reactor. Hydrocracking of the waxes was modelled according to Tomasek et al. (2020). The total production of diesel, gasoline, kerosene, and C_1-C_4, coming from the FT-reactor and the wax hydrocracker are summed in order to get the final products. Furthermore, the light ends are combusted to produce CO_2, which is then recycled to the water-gas shift reactor (WGSR). Since gasoline and kerosene are co-produced, our LCA considers a system expansion approach with avoided burdens. The final inputs of the FT-diesel life cycle inventory are presented in Table 1.

Table 1: Life cycle inventory for the production of 1 kg FT-Diesel from CO_2 and H_2

FT-diesel	1	kg
Avoided products		
Gasoline	0.89	kg
Kerosene	0.72	kg
Inputs		
H_2	0.29	kg
CO_2	1.42	kg

The foreground system, i.e., truck and road constructions, etc., is based on the "Lorry 16-32 metric ton, EURO6" of the Ecoinvent v3.5 database. In essence, we consider the LCI for the BAU scenario, replacing the fossil diesel's inventory with that of FT-diesel, and adjusting the direct emissions based on Schemme et al. (2017). The scenarios differ in the origin of the educts (Figure 1). For this study, CO_2 is captured either from point sources at coal power plants, or directly from air (Coal and DAC). H_2 is produced through an electrolytic or thermochemical route. Polymeric water electrolysis is considered as the electrolytic route, powered by different energy sources, i.e., onshore wind, nuclear, and bioenergy with CCS (BECCS). Furthermore, for the thermochemical route, the conversion of biomass to hydrogen with carbon capture and storage (BTH CCS) is considered.

Data for the production of electrolytic H_2 were taken from Bareiß et al. (2019), considering for wind power a capacity factor of 0.34, respectively. For the different electricity sources, we used data from Ecoinvent v1.03, except for the BECCS scenario, based on Oreggioni et al. (2017). For the thermochemical route, the inventory was retrieved from Susmozas et al. (2016). Concerning the capture of CO_2, the Coal scenario is based on Iribarren, Petrakopoulou, and Dufour (2013), and the DAC scenario on Keith et al. (2018).

The life cycle impact assessment (LCIA) quantifies the absolute environmental sustainability level of FT-diesel by converting the LCI elementary flows into impacts on the control variables of seven PBs. Consequently, for a scenario to be regarded as sustainable, none of the planetary boundaries should be transgressed Our study follows the characterization factors proposed by Ryberg et al. (2018) to quantify the impact on six PBs, together with the ones introduced by Galán-Martín et al. (2021) to evaluate the impact on biosphere integrity.

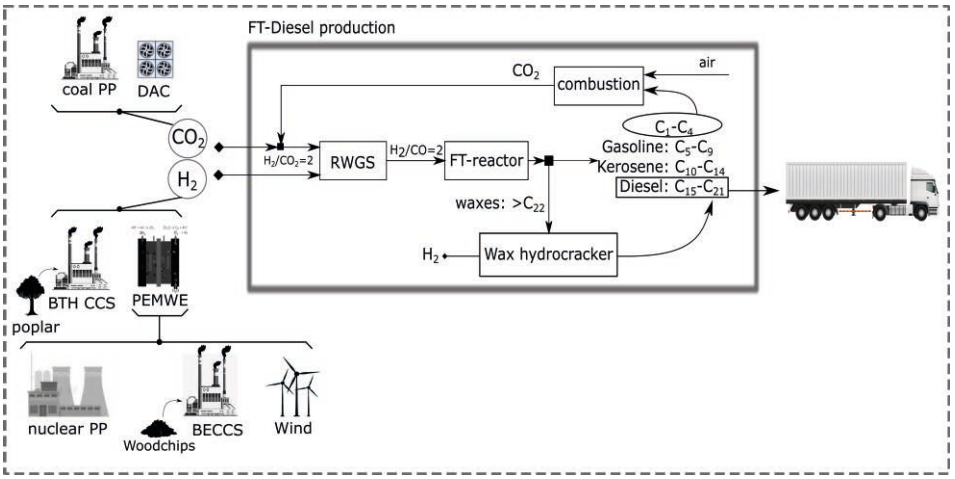


Figure 1: System boundaries of the different scenarios. From the production of CO_2 and H_2 from different sources to the production of FT-Diesel, and lastly, the end-use in HD trucks.

3. Results

3.1. Relative impact to the Safe Operating Space

Figure 2 shows that the current BAU scenario is unsustainable due to the transgressions of the climate change PB (CO_2, EI). Overall, all the FT-diesel scenarios have the potential to decrease the impacts of the BAU scenario. However, only two are sustainable, namely those based on DAC with electrolytic H_2 from nuclear, and H_2 from Biomass with CCS. These scenarios operate within the SOS for all th PBs. On the CO_2 boundary (75%, -140%, respectively), 1% and 2% in nitrogen flows (N-flows), and 7% and 25% in biosphere integrity (BII). All the remaining scenarios fail to be sustainable because they lead to burden-shifting to the N-flows and BII. Focusing on the scenarios with electrolysis routes powered with wind and nuclear electricity (DAC + Wind, DAC + Nuclear, coal + Wind, coal + Nuclear), undoubtedly, DAC scenarios would perform better than those relying on coal. The CO_2 coming from fossil resources was modelled as a positive emissions entry in contrast to the DAC scenario, where CO_2 is coming from the air, and hence is modeled as a negative emissions entry. Ultimately, these scenarios represent an interim solution as fossil fuels should be ultimately phased out. Scenarios that make use of CCS and biomass (DAC + BTH CCS, DAC + BECCS, coal + BTH CCS, coal + BECCS) show a great potential in the GHG-related PBs (CO_2, EI, OA), but lead to burden-shifting due to biomass growth. Notably, the N-flows are affected by the fertilisers, and the BII category by the use of land, e.g., DAC + BECCS and coal + BECCS take 30% of the SOS in the N flows.

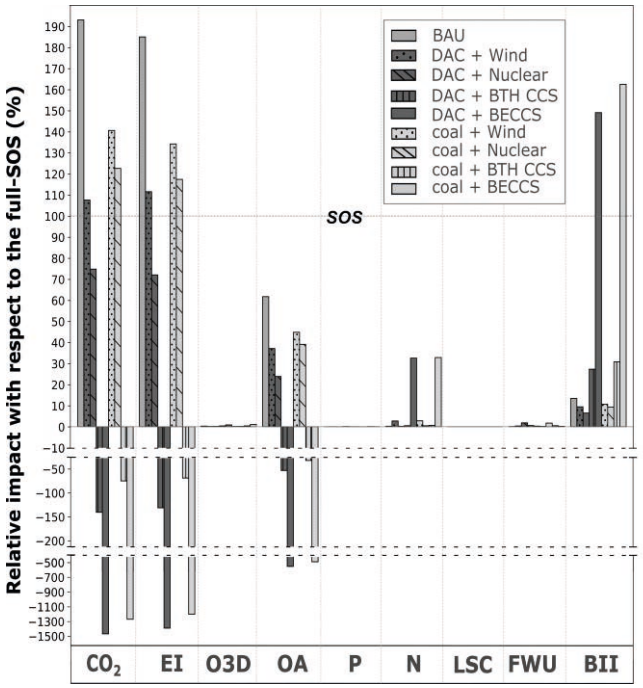


Figure 2: Relative impact in percentage of the safe operating space (SOS). The abbreviations of the PBs are: CO_2 (Climate change CO_2 concentration), EI (Energy imbalance), O3D (Stratospheric ozone depletion), P (Phosphorus flows), N (Nitrogen flows), LSC (Land system chage), FWU (Fresh water use), BII (Biosphere integrity).

3.2. Impact breakdown

The breakdown of impacts in Figure 3 (upper) shows that most carbon-positive impacts come from the combustion emissions, i.e., 45%-85% of the total positive contributions in all the scenarios. Carbon negative impacts come from the biomass based scenarios with CCS (BTH CCS, BECCS), and are linked to the H_2 production. Electrolysis powered with BECCS has the most negative impacts due to the carbon-negative nature of the electricity generated, which requires large amounts of biomass. Furthermore, H_2 from nuclear performs 1.2-fold better in the CO_2 PB compared to H_2 from Wind. Regarding the CO_2 capture technologies, DAC is the only technology that can provide negative impacts since CO_2 is modelled as a negative emission entry. With regard to the biosphere integrity (BII), Figure 3 (lower) shows that the biggest impacts come BTH CCS and BECCS, with the latter being the worst. These high impacts are linked to the extensive land use for biomass growth. All the other scenarios perform better comparing to the BAU, however, it is important to mention that 65% of the carbon positive impacts are coming from the combustion emissions and 25% from the construction of roads.

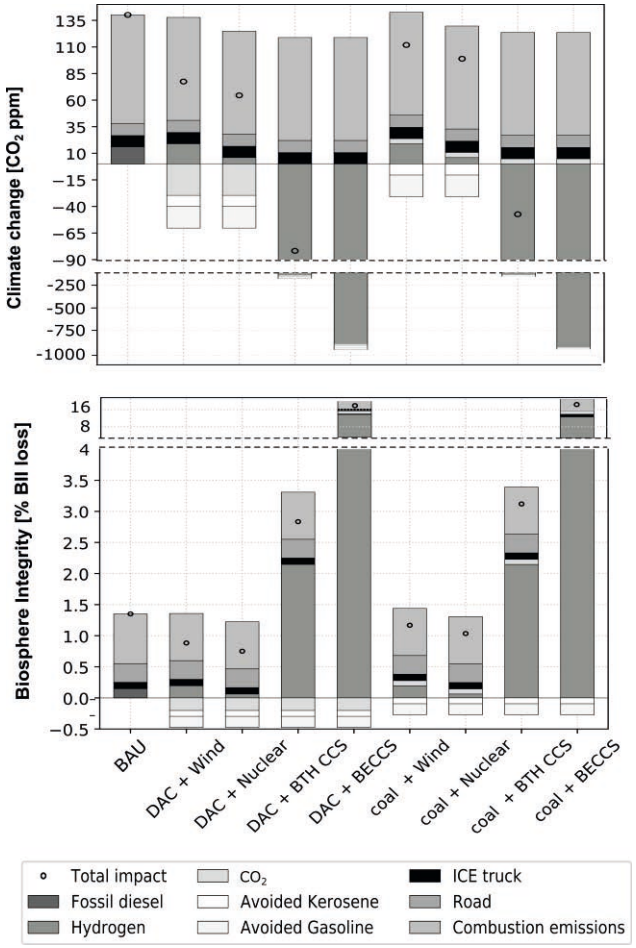


Figure 3: Breakdown of impacts in the CO_2 boundary (upper), nitrogen flows (middle), and biosphere integrity (lower).

4. Conclusions

Using CO_2-based fuels has attracted increasing attention, yet their broad sustainability implications remain unclear. Here we assessed the absolute environmental sustainability of FT-diesel from renewable carbon as an alternative fuel for HD trucks using seven planetary boundaries (PBs). We found that the current fossil-based fuel alternative transgresses the climate change-related PBs, while renewable-carbon fuels could help operate within these ecological limits. However, burden-shifting to BII and N-flows may occur. This collateral damage would be more critical in the biomass-related scenarios, which remove large amounts of CO_2 but need land and fertilisers for biomass growth. Hence, the CO_2 and H_2 sources for producing these fuels should be selected carefully to mitigate climate change without exacerbating the damage in other critical Earth-system processes, thereby preserving the planet's stability.

Acknowledgements: This work was created as part of the NCCR Catalysis, a National Centre of Competence in Research funded by the Swiss National Science Foundation.

References

Al-Yaeeshi, Ali Attiq, A. AlNouss, Gordon McKay, and T. Al-Ansari, 2019, "A model based analysis in applying Anderson–Schulz–Flory (ASF) equation with CO_2 utilisation on the Fischer Tropsch gas-to-liquid process.", Computer Aided Chemical Engineering, 46, 397–402.

Bareiß, Kay, Cristina de la Rua, Maximilian Möckl, and Thomas Hamacher, 2019, "Life cycle assessment of hydrogen from proton exchange membrane water electrolysis in future energy systems.", Applied Energy, 237, 862-72.

Galán-Martín, Ángel et al., 2021, "Sustainability footprints of a renewable carbon transition for the petrochemical sector within planetary boundaries." One Earth, 4, 565–83.

Goedkoop, Mark et al., 2016, "Introduction to LCA with SimaPro Colophon. "Introduction to LCA with SimaPro."

Holmgren, Kristina, and Linus Hagberg, 2009, "Life cycle assessment of climate impact of Fischer-Tropsch diesel based on peat and biomass."

Iribarren, Diego, Fontina Petrakopoulou, and Javier Dufour., 2013, "Environmental and thermodynamic evaluation of CO_2 capture, transport and sstorage with and without enhanced resource recovery." Energy, 50, 477–85.

Keith, David W., Geoffrey Holmes, David St. Angelo, and Kenton Heidel, 2018, "A process for capturing CO_2 from the atmosphere.", Joule, 2,1573–94.

Li, Pengcheng, Zhihong Yuan, and Mario R. Eden, 2016, "A comparative study of Fischer-Tropsch synthesis for liquid transportation fuels production from biomass.", Computer Aided Chemical Engineering, 38, 2025–30.

Martín, Mariano, and Ignacio E. Grossmann. 2011. "Process optimization of FT-diesel production from lignocellulosic switchgrass." Industrial and Engineering Chemistry Research 50(23): 13485–99.

Oreggioni, Gabriel D. et al. 2017. "Environmental assessment of biomass gasification combined heat and power plants with absorptive and adsorptive carbon capture units in Norway." International Journal of Greenhouse Gas Control, 57, 162–72.

Rockström, J. et al. 2009. "A Safe operation space for humanity." Nature, 46, 472–75.

Schemme, Steffen, Remzi Can Samsun, Ralf Peters, and Detlef Stolten. 2017. "Power-to-Fuel as a key to sustainable transport systems – An analysis of diesel fuels produced from CO_2 and renewable rlectricity.", Fuel, 205, 198–221.

Shafer, Wilson D. et al. 2019. "Fischer-Tropsch: Product selectivity-the fingerprint of synthetic fuels." Catalysts.

Susmozas, Ana et al., 2016, "Life-Cycle performance of hydrogen production via indirect biomass gasification with CO_2 Capture." International Journal of Hydrogen Energy, 41, 19484–91.

Szabina Tomasek, Ferenc Lonyi, Jozsef Valyon, Anett Wollmann, Jeno Hancsok., "Hydrocracking of Fischer−Tropsch paraffin mixtures over strong acid bifunctional catalysts to engine Fuels."

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50329-8

Renewable Power Systems Transition Planning using a Bottom-Up Multi-Scale Optimization Framework

Ning Zhao[a*], Yanqiu Tao[a], Fengqi You[a]

[a]*Cornell University, Ithaca, New York, 14853, USA*
nz225@cornell.edu

Abstract

In this work, we propose a novel multi-scale bottom-up optimization framework to address the decarbonization transition planning for power systems, which incorporates multiple types of information for each existing or new unit in the power systems, including its technology, capacity, and age. To reduce the computational challenge, a novel approach integrating Principal Component Analysis (PCA) with clustering techniques is proposed to obtain representative days. To illustrate the applicability of the proposed framework, a case study for New York State was presented. The proposed approach obtaining representative days using PCA coupled with K-means shows better performance than multiple state-of-the-art clustering approaches.

Keywords: decarbonization, renewable electricity transition, multi-scale optimization.

1. Introduction

Power systems decarbonization has been a priority topic for countries around the world (Gong et al., 2015). It facilitates the design of power systems decarbonization transition pathways to simultaneously optimize the systems' capacity changes and simulate the corresponding hourly operations, while considering each individual unit in the power systems (Zhao et al., 2020, 2021). Existing multi-scale energy transition optimization models typically include two time scales on yearly and hourly bases (Bennett et al., 2021). The yearly time scale accounts for the decisions of capacity changes to the power systems, while the operational decisions are made on an hourly basis in conjunction with the design decisions (Brown et al., 2018). To reduce the computational requirements associated with simultaneous planning for the energy transition pathways and simulating the hourly systems operations for the next multiple decades (Prina et al., 2020), the representative day approach has been widely applied in multi-scale energy transition optimization studies (Teichgraeber et al., 2019). Multiple approaches have been used to obtain the representative days, such as rule-based selection, agglomerative hierarchical clustering, and K-means clustering (Gabrielli et al., 2018). On the other hand, most of the existing multi-scale bottom-up energy transition models include only the capacity and technology information of a unit, while including the ages of both existing and future units in the framework is crucial for developing more reliable transition pathways, because existing units with large ages and new units with short facility lifetimes may retire during the transition period of decades owing to the lifespan limits. To the best of our knowledge, there is no existing research work on the multi-scale bottom-up renewable electricity transition optimization that incorporates multiple dimensions of information for each individual unit, including its technology, capacity,

and age. To fill the knowledge gap, we propose a multi-scale bottom-up optimization framework that incorporates multiple dimensions of information for a unit as well as a machine learning-based approach to construct the reduced model.

2. Multi-scale bottom-up energy transition optimization framework

The overview of the proposed framework is presented in Figure 1. The proposed framework consists of three steps: (1) data processing based on machine learning, (2) multi-scale transition optimization, and (3) detailed hourly operational simulation.

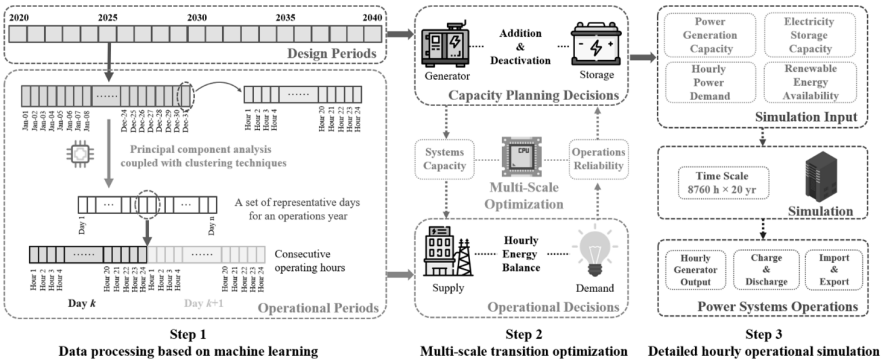


Figure 1. Overview of the proposed multi-scale bottom-up energy transition framework.

In the first step, machine learning techniques, which have been widely used in optimization (Shang et al., 2019), are applied to obtain the representative days based on the power load data from an entire year for developing a reduced optimization model, as it can be extremely computationally demanding for the multi-scale energy transition optimization to account for the yearly capacity planning and the hour-by-hour operational simulation simultaneously for the whole planning horizon of multiple coming decades. Specifically, a novel approach is proposed to obtain the representative days by coupling Principal Component Analysis (PCA) with clustering techniques that include agglomerative hierarchical clustering (AHC), Gaussian mixture model (GMM), Dirichlet process mixture model (DPMM), and K-means clustering. The data being clustered is the 24-dimension hourly power loads for all days in a year. We investigate the performances of using PCA coupled with each clustering approach. The clustering performances are evaluated by three metrics, namely intra-cluster variance, inter-cluster variance, and the Calinski-Harabasz index.

In the second step, multi-scale energy transition optimization is conducted based on a reduced model using representative days that are obtained from the first step. Two time scales are applied in the proposed optimization framework, namely the design periods and the operational periods. The planning horizon is equally partitioned on an annual basis, and the capacity planning decisions that include the additions and deactivations of generators and storage units should be determined for each year of the resulting design periods. On the other hand, to ensure the reliability and energy balance of the deep-decarbonized electric power systems with high penetration of renewable energy, hourly systems operations during the operational periods are incorporated in the proposed optimization model in conjunction with the changes to the electric power sector resulted from the capacity planning decisions.

In the third step, the hourly power systems operations, namely unit comment (Padhy, 2004), are simulated via optimization of each design period, based on the optimal energy transition results from the multi-scale optimization step, hourly power demand projections, and the hourly availability of renewable energy. Specifically, the generation and electricity storage capacities are fixed in the simulation according to the optimal capacity planning decisions in the second step, and the simulation aims to minimize the total operational cost by determining the outputs of each generator, the charging and discharging of energy storage units, and the importation and exportation of electricity on an hourly basis (Qiu et al., 2020), while ensuring the reliability, potential faults (Ajagekar and You, 2021), and balances of the electric power systems.

3. Case study for New York State

A case study on the renewable electricity transition for the New York State is presented to illustrate the applicability of the proposed multi-scale bottom-up optimization framework. The renewable electricity requirements and the climate targets for the New York State are set following the state legislation. The generation and storage capacity data, the annual electricity generation projections, and the scheduled power systems capacity changes for the New York State are obtained based on a report from the New York Independent System Operator (NYISO). In addition, the data on generation capacities for existing distributed solar PV in the state are collected following a study of the New York State Energy Research and Development Authority (NYSERDA). The technological and economic data projections for the power generation and electricity storage technologies are collected from a recent study (Tian and You, 2019). The hourly operations data for the state are obtained from the NYISO energy market and operation data, while the hourly availability of solar, on-land wind, and offshore wind are retrieved from the literature (Ning and You, 2019, 2022).

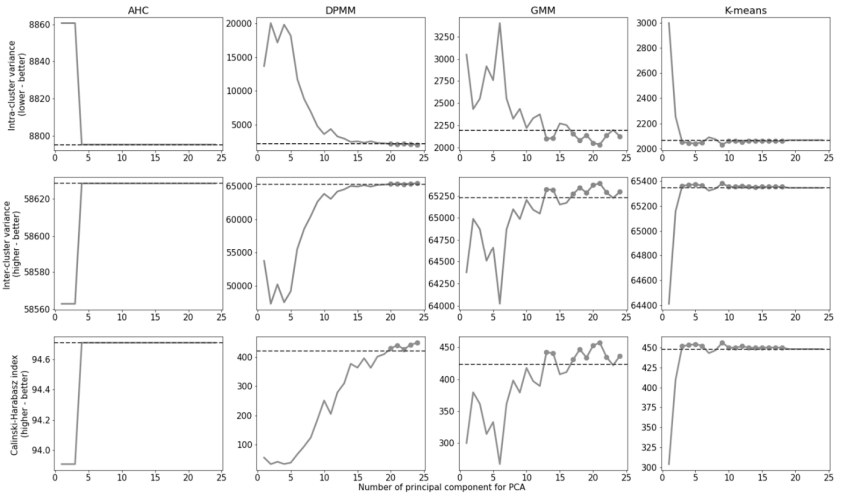


Figure 2. Intra-cluster variance, inter-cluster variance, and the Calinski-Harabasz index using PCA coupled with clustering approaches. Horizontal lines indicate performance without PCA.

To obtain the representative days, we investigate the performances of coupling PCA with multiple clustering approaches that include AHC, DPMM, GMM, and K-mean, as

well as the performances using each particular clustering technique without coupling with PCA, as shown in Figure 2. From the performance evaluation results, using PCA coupled with clustering techniques could provide more effective or at least the same clustering results compared with using these techniques individually without PCA. For AHC, all three metrics are not as good as the other types of clustering techniques, regardless of whether it couples with PCA or not, and the best clustering result using AHC coupled with PCA is the same as using AHC alone. On the other hand, coupling PCA with other clustering techniques could improve the data grouping performances compared to using these techniques without PCA, as shown by lower intra-cluster variances, higher inter-cluster variances, and higher Calinski-Harabasz indices when PCA is involved. The improvement is owing to the effectiveness of PCA in capturing the correlations of the high-dimensional input data.

The optimization programs of the energy systems transition problem are coded in GAMS 27.3 on a PC with an Intel Core i7-8700 @ 3.20 GHz and 32.00 GB RAM, running on a Windows 10 Enterprise, 64-bit operating system. The energy transition planning is solved using CPLEX 12.9.0.0 with an optimality tolerance of 1%. The problem has 4,889 integer variable, 2,294,943 continuous variables, and 1,596,333 constraints. The optimal objective value is $ 96,343MM, and it takes 12,294 CPUs to solve the problem using the proposed optimization framework. Furthermore, to obtain detailed optimization results, power systems operations are simulated on an hourly basis for the entire planning horizon, based on the optimal transition pathway for electric power systems. The total transition cost under detailed operational simulation is $97,729MM, indicating that the difference between the optimal costs from energy transition planning and detailed hourly simulation is less than 1.5%. The simulation time of less than 70 CPUs is significantly less than the optimization time for energy transition planning, because the capacities of generators and storage units in each year are fixed for the operational simulations. As a result, each year's simulations are independent of other years, leading to a substantial reduction of computational demand.

The power generation capacity and annual electricity generation by the source during the decarbonization transition are shown in Figure 3(a) and Figure 3(b), respectively. As for generation capacities, offshore wind starts to participate in power generation in 2024, and its total capacity remains relatively stable during 2025-2030. In the 2030s, offshore wind power capacity gradually increases until the end of the planning horizon. Regarding solar PV, the generation capacity of utility solar PV has no significant changes during the beginning years of the planning horizon, and it starts to increase after 2027. This is mainly because the annual electricity consumption in the New York State is expected to decrease at the beginning years, owing to efficiency improvements across the state, while the total annual power load is projected to increase after 2027. On the other hand, distributed solar PV has a stable capacity across the planning horizon, which is owing to two reasons: it has lower economic efficiency compared to utility solar PVs, and most of the existing ones will not retire by 2040. For annual electricity generation shown in Figure 3(b), offshore wind would generate the most electricity by the end of the planning horizon, while hydropower and utility solar PV are the other two primary generation technologies in 2040. Note that although the total capacity of offshore wind is less than that of utility solar PV in 2040, offshore wind turbines tend to have much higher average capacity factors than utility solar PV, which enables them to generate more electricity on an annual basis. Hydropower currently accounts for the majority of renewable electricity generation in the state, and it

continues to provide stable electricity on an annual basis across the planning horizon, owing to its relatively stable total generation capacity over the planning years.

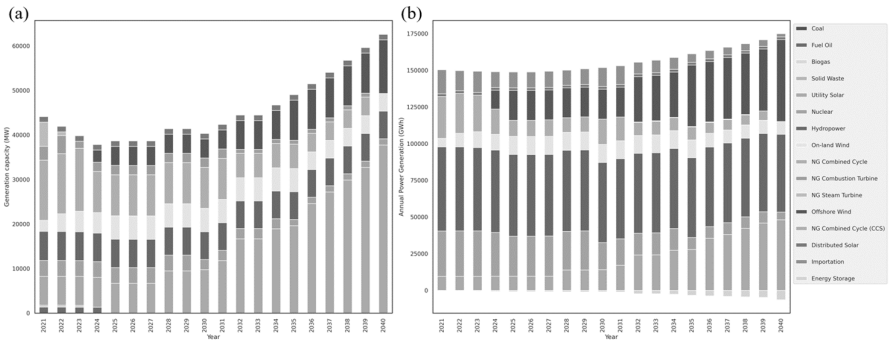


Figure 3. Electricity generation capacity and annual electricity generation by the source during the renewable electricity transition. (a) Electricity generation capacity according to the optimal transition pathway. (b) Annual electricity generation according to the operational simulations.

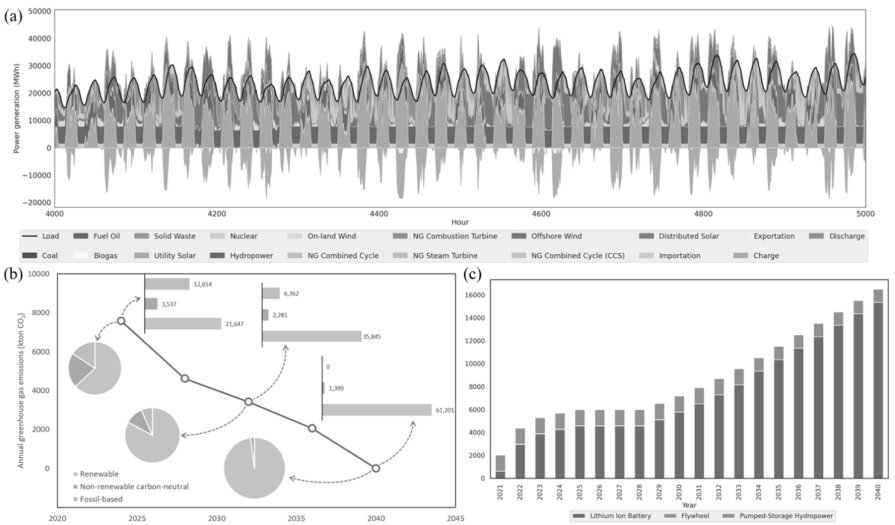


Figure 4. (a) Hourly power systems operations. (b) Annual greenhouse gas emissions. (c) Electricity storage capacities.

Figure 4 shows the hourly power systems operations, annual greenhouse gas emissions, and electricity storage capacities. Note that the fluctuation of electricity supply capacity is significant, owing to the high penetration level of variable renewable energy, such as solar and wind, and consequently, the charging and discharging activities are conducted frequently on a large scale. Solar PVs show clear periodic power outputs, because of the limited availability of solar energy during the evening, while the power outputs from offshore wind show no clear intra-day correlations. As for greenhouse gas emissions, they reduce decrease almost linearly across the planning horizon, while the reduction rate at the beginning years is slightly higher compared to the later periods, owing to more deactivated fossil-based power generation capacities in this period. Note that the

greenhouse gas emissions reach zero in 2040, indicating that the goal of 100% decarbonized power systems is achieved. For electricity storage changes, lithium-ion batteries have the highest storage capacity by the end of the transition process, because of their higher economic efficiency, and their increasing trend is consistent with the increasing power generation capacities from intermittent sources.

4. Conclusion

In this work, a novel multi-scale bottom-up optimization framework was proposed to address the decarbonization transition planning for power systems, which incorporated multiple dimensions of information for each existing or new unit in the power systems, including its technology, capacity, and age. To reduce the computational challenge, a novel approach integrating PCA with clustering techniques was proposed to obtain representative days. To illustrate the applicability of the proposed framework, a case study for New York State was presented. The proposed approach showed better performance than multiple state-of-the-art clustering approaches.

References

A. Ajagekar, F. You, 2021, Quantum computing based hybrid deep learning for fault diagnosis in electrical power systems. Applied Energy, 303, 117628.

J. A. Bennett et al., 2021, Extending energy system modelling to include extreme weather risks and application to hurricane events in Puerto Rico. Nature Energy, 6, 3, 240-249.

T. Brown, J. Hörsch, D. Schlachtberger, 2018, PyPSA: Python for Power System Analysis. Journal of Open Research Software, 6, 1, 4.

P. Gabrielli, M. Gazzani, E. Martelli, M. Mazzotti, 2018, Optimal design of multi-energy systems with seasonal storage. Applied Energy, 219, 408-424.

J. Gong, F. You, 2015, Sustainable design and synthesis of energy systems. Current Opinion in Chemical Engineering, 10, 77-86.

C. Ning, F. You, 2019, Data-Driven Adaptive Robust Unit Commitment Under Wind Power Uncertainty: A Bayesian Nonparametric Approach. IEEE Transactions on Power Systems, 34, 2409-2418.

C. Ning, F. You, 2022, Deep Learning Based Distributionally Robust Joint Chance Constrained Economic Dispatch Under Wind Power Uncertainty. IEEE Transactions on Power Systems, 37, 191-203.

N.P. Padhy, 2004, Unit commitment-a bibliographical survey. IEEE Transactions on Power Systems, 19, 1196-1205.

M. G. Prina, G. Manzolini, D. Moser, R. Vaccaro, W. Sparber, 2020, Multi-Objective Optimization Model EPLANopt for Energy Transition Analysis and Comparison with Climate-Change Scenarios. Energies, 13, 12.

H. Qiu, F. You, 2020, Decentralized-distributed robust electric power scheduling for multi-microgrid systems. Applied Energy, 269, 115146.

H. Teichgraeber, A. R. Brandt, 2019, Clustering methods to find representative periods for the optimization of energy systems. Applied Energy, 239, 1283-1293.

C. Shang, F. You, 2019, Data Analytics and Machine Learning for Smart Process Manufacturing: Recent Advances and Perspectives in the Big Data Era. Engineering, 5, 1010-1016.

X. Tian, F. You, 2019, Carbon-neutral hybrid energy systems with deep water source cooling, biomass heating, and geothermal heat and power. Applied Energy, 250, 413-432.

N. Zhao, F. You, 2020, Can renewable generation, energy storage and energy efficient technologies enable carbon neutral energy transition? Applied Energy, 279, 115889.

N. Zhao, F. You, 2021, New York State's 100% renewable electricity transition planning under uncertainty using a data-driven multistage adaptive robust optimization approach with machine-learning. Advances in Applied Energy, 2, 100019.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50330-4

Design and Operation of Urban Energy Network: Integration of Civic, Industrial, and Transportation Sectors

Ruonan Li[a], Vladimir Mahalec[a*]

[a]*Department of Chemical Engineering McMaster University, 1280 Main St. West, Hamilton ON L8S 1A8, Canada*
mahalec@mcmaster.ca

Abstract

Integration of distributed energy systems for entities can further reduce greenhouse (GHG) emissions beyond the minimum emissions achieved by the individually operated energy systems. This work introduces an optimization approach with relative sizes of integrated entities, design and operation of the integrated energy system equipment, and production rates of plants as decision variables to maximize GHG emissions reduction brought by the integrated operation. The approach also differentiates temperature levels of heating demands to ensure feasible heat transfer by formulating heat balance for each process that requires heating. Results from case studies on an integrated system with a residential building with electric vehicles, a supermarket, a confectionery plant, a bakery plant, and a brewery show that, when optimizing the size of entities, the maximum GHG emissions reduction achieved by the integrated system is relatively constant under the various sizes of the residential building.

Keywords: Distributed energy network; GHG emissions reduction; Light industry; Energy system integration; Energy, Food and Environmental Systems.

1. Introduction

The integration of energy sectors reduces GHG emissions of urban areas by combining the heating, cooling, and electricity demands of civic structures, industrial plants, and transportation sectors (Fichera et al., 2017). The combined cooling, heating, and power (CCHP) system provides a solution for integrating energy sectors, where the power generation unit (PGU) is the critical equipment. It combusts fuel to generate electricity and uses the waste heat for the heating and cooling demand of an entity. Generally, the electricity and heat generated cannot be entirely consumed at the same time. Besides implementing additional equipment and energy resources, the unbalanced energy load and supply can also be solved by integrating individually operated CCHP systems of entities through heat and electricity transfer. Thus, an energy network forms, where each entity performs both as an energy supplier and consumer.

Knowledge gaps in existing studies on the integration of energy systems are identified as the following: 1. Existing studies assume the heating demands of all entities are at a uniform temperature. The assumption is not valid when the integrated system includes industrial plants, which require utilities at different temperatures for production. 2. Existing studies are based on fixed industrial energy demand profiles. It can lead to all entities having high energy demands at the same time. 3. There lacks an approach that identifies the optimal relative size of entities being involved.

Thus, this work introduces a novel optimization approach for the integration of energy systems in residential buildings with electric vehicles, commercial buildings, and industrial plants. The approach addresses the following areas not previously explored: 1. differentiates temperature levels of heating demands to ensure feasible heat transfer among entities; 2. sets production rates of integrated industries as decision variables to adjust industrial energy demands and further increase GHG emissions reduction of the integrated system; 3. provides optimal relative sizes of integrated entities, design, and operation of energy system equipment that maximize GHG emissions reduction. The formulation of the optimization problem is presented in Section 2. The approach has been tested by the case study presented in Section 3, where the results are shown in Section 4. Section 5 summarizes the main findings.

2. Optimization problem formulation

The optimization approach intends to maximize GHG emissions reduction brought by the integrated operation, compared to the non-integrated system. It is achieved by finding the optimal design and operation of energy system equipment, the relative size of entities in the integrated network, and the production rates of integrated industries. This work assumes the energy system of an entity has the structure shown in Figure 1.

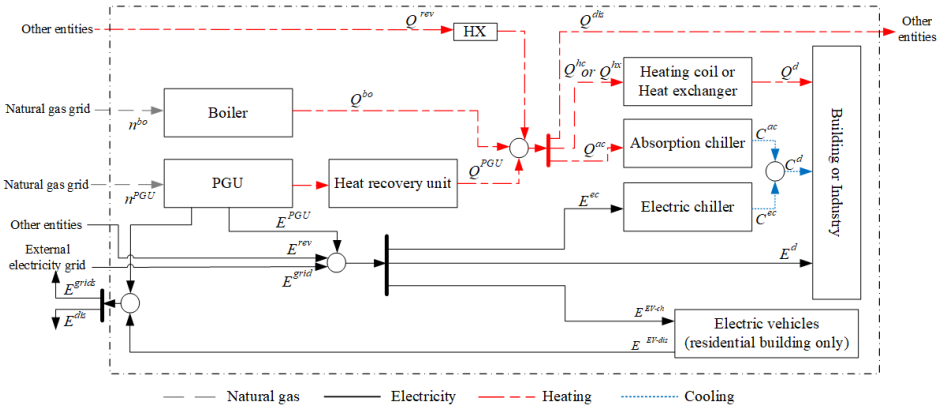


Figure 1 – A representation of the energy system in an entity.

2.1. Decision variables

Decision variables of the optimization approach can be classified as design decision variables, which are time-invariant, and operation decision variables, which can be manipulated during the operation.

Design decision variables include (i) size of energy system equipment: Cap_i^{eqp}; (ii) size of each entity: S_i. Operation decision variables include (i) amount of fuel used by PGU and boiler: $n_{i,t}^{PGU}$ and $n_{i,t}^{bo}$; (ii) amount of electricity an entity purchases and sells to the external grid: $E_{i,t}^{grid}$ and $E_{i,t}^{grids}$; (iii) whether there is heat and electricity transfer between two entities: $y_{i,i',t}$ and $z_{i,i',t}$ and the amount: $Q_{i,i',t}^{dis}$ and $E_{i,i',t}^{dis}$; (iv) production rate of the plants: $x_{i,t}$; (v) whether electric vehicles are charged or discharged: u_t and w_t and the corresponding amount: E_t^{EV-ch} and E_t^{EV-dis}.

2.2. Constraints

Constraints for the optimization problem can be further divided into four categories: energy generation of equipment, electric vehicles, energy transfer, and energy balances.

(1) Energy generation of equipment: Constraints under this category calculate the amount of heat ($Q_{i,t}^{eqp-out}$), electricity ($E_{i,t}^{eqp-out}$), or cooling ($C_{i,t}^{eqp-out}$) generated by a piece of equipment ($Egy_{i,t}^{eqp-out}$) based on its energy consumption.

$$Egy_{i,t}^{eqp-out} = \eta_i^{eqp} Egy_{i,t}^{eqp-in} \tag{1}$$

η_i^{eqp} is the efficiency of the equipment. Depending on the equipment, $Egy_{i,t}^{eqp-in}$ can represent fuel, heat, or electricity used by the equipment. Taking PGU and boiler as examples, $Egy_{i,t}^{eqp-in}$ represents the amount of fuel combusted ($n_{i,t}^{PGU-in}$ and $n_{i,t}^{bo-in}$) in the equipment. Besides energy consumption, heat, electricity, or cooling generated by an entity should be less than or equal to the equipment capacity (Cap_i^{eqp}).

(2) Electric vehicles: This work assumes all electric vehicles (EVs) can only be charged or discharged in the residential building and investigates all EVs as an aggregated subsystem to simplify the formulation. The amount of electricity in the battery (E_t^{EV}) equals the amount of electricity at the previous time (E_{t-1}^{EV}), plus the charged electricity (E_t^{EV-ch}), and minus the electricity discharged (E_t^{EV-dis}).

Since the charging and discharging behavior cannot occur at the same time, binary variables - u_t and w_t are used, as formulated in Eq. (2). ER^{EV} represents the maximum charging and discharging rate of EVs. Eq. (3) restricts EVs to be fully charged when leaving the building (t^l) and defines the amount of electricity in EVs when returning home (t^r). E^{EV-con} is the amount of electricity consumed by EVs outside.

$$E_t^{EV-ch} \le ER^{EV} u_t, E_t^{EV-dis} \le ER^{EV} w_t, u_t + w_t \le 1, u_t, w_t \in \{0,1\} \tag{2}$$

$$E_{t=t^l}^{EV} = Cap^{EV}, E_{t=t^r}^{EV} = Cap^{EV} - E^{EV-con} \tag{3}$$

(3) Energy transfer: Constraints associated with energy transfer of the integrated system (i) ensure energy transfer between two entities at a period is unidirectional by using the binary variables $y_{i,i',t}$ and $z_{i,i',t}$, which indicate whether heat and electricity are being dispatched from entity *i* to *i'*, respectively; (ii) restrict the amount of energy dispatched based on the amount of heat and electricity generated in the entity and the heat transfer pipe size; (iii) calculate the available heat and electricity an entity received by excluding energy loss during the transfer from the dispatched heat.

(4) Energy balance: Constraints under this category ensure the energy demand of an entity can be fully satisfied by all available energy resources. Since the proposed optimization approach differentiates temperatures of heating demands, the heat balance of each process is developed individually to reflect whether the process can use the transferred heat or not. For processes that cannot use the transferred heat, Eq. (4) is applied without the received heat item – $Q_{i,i',p,t}^{rev}$.

$$\frac{Q_{i,p,t}^{d}}{\eta_{i,p}^{hx}} = Q_{i,p,t}^{PGU-out} + Q_{i,p,t}^{bo-out} + \sum_{i} Q_{i,i',p,t}^{rev} \tag{4}$$

$Q^{d}_{i,p,t}$ represents the heating demand of a process p. For industrial plants, the energy demand of a process is calculated based on the production rate – x_i (decision variable) and the energy used to produce a unit of product. $Q^{PGU-out}_{i,p,t}$ and $Q^{bo-out}_{i,p,t}$ are heat generated by the PGU and boiler. For residential buildings and commercial buildings whose heating demands are at a uniform temperature, p equals one.

2.3. Objective function

As mentioned, the objective function of the problem is maximizing GHG emissions reduction led by the integrated operation. It is based on the minimum GHG emissions of the integrated system and the non-integrated system, as shown in Eq. (5).

$$GHGD\ \% = \left(GHG_{non-integrated} - GHG_{integrated}\right)/GHG_{non-integrated} \tag{5}$$

The minimum GHG emissions of the integrated ($GHG_{integrated}$) and non-integrated system ($GHG_{non-integrated}$) are calculated based on the amount of fuel ($n^{PGU-in}_{i,t}$ and $n^{bo-in}_{i,t}$), electricity ($E^{grid}_{i,t}$) used by the system, and electricity sold to the external grid ($E^{grids}_{i,t}$). σ_{NG} and σ_E are GHG emissions coefficients associated with using fuel and grid electricity. Eq. (6) shows an example of GHG emissions for the integrated system.

$$GHG_{integrated} = \sum_i \sum_t \sigma_{NG}\left(n^{PGU-in}_{i,t} + n^{bo-in}_{i,t}\right) + \sigma_E E^{grid}_{i,t} - 0.5\sigma_E E^{grids}_{i,t} \tag{6}$$

The minimum GHG emissions of the non-integrated system can be simplified as a linear equation related to entity sizes. The linear relationship holds because there exist optimal operation patterns for the energy system of each non-integrated entity, which minimizes GHG emissions of the non-integrated system. When the sizes of the entities change, the optimal operation patterns of equipment do not change; however, the equipment sizes increase or decrease correspondingly to maintain the minimum GHG emissions.

3. Case study description

An integrated system with a residential building, a supermarket, a confectionery plant, a brewery, a bakery plant, and EVs has been used for case studies. The residential building and supermarket are assumed to have fixed energy profiles based on published information from Sullivan (2020) and Ghorab (2019), respectively. The sizes of the two entities are relative sizes, which represent the number of buildings having the base-case energy demands. Sizes of the confectionery plant, brewery, and bakery plant are the maximum production rates of the plants. Energy used to make a unit of product for the three plants is obtained based on the process studied by Singh (1986) and Therkelsen et al. (2014). Since the supermarket requires a large amount of energy for low-temperature refrigeration, the supermarket has been assumed does not have the PGU, heat recovery unit, or absorption chiller. For reducing the computation time, the relative size of the supermarket is set to be equal to the size of the residential building. Additionally, each unit of the relative size of residential building has been assumed to have 870 EVs.

4. Results and discussion

It has been found that the integrated operation can reduce GHG emissions of the system by a maximum of 17.5 %. The reduction is achieved by integrating a residential building,

a supermarket, 870 EVs, a brewery with a capacity of 2,720 kg/hr, and a 5,000 kg/day bakery plant. There is no confectionery plant in the system. The reduction is due to, with energy transfer among entities, the integrated system purchasing 65.0 % less electricity from the external grid and operates boilers 79.6 % less.

As shown in Figure 2, although the maximum GHGD% (Case 1) is achieved when the relative size of the residential building is one, under larger relative sizes of the residential building, the highest GHGD % (Case 2, Case 3, and Case 4) at lower values can be achieved when optimizing sizes of the other entities. The values are 17.2 %, 16.7 %, and 15.9 % when relative sizes of the building are three, five, and ten, respectively. Table 1 presents the corresponding plant sizes. The case studies reflect the situation when there are specific requirements on entity sizes of the integrated system.

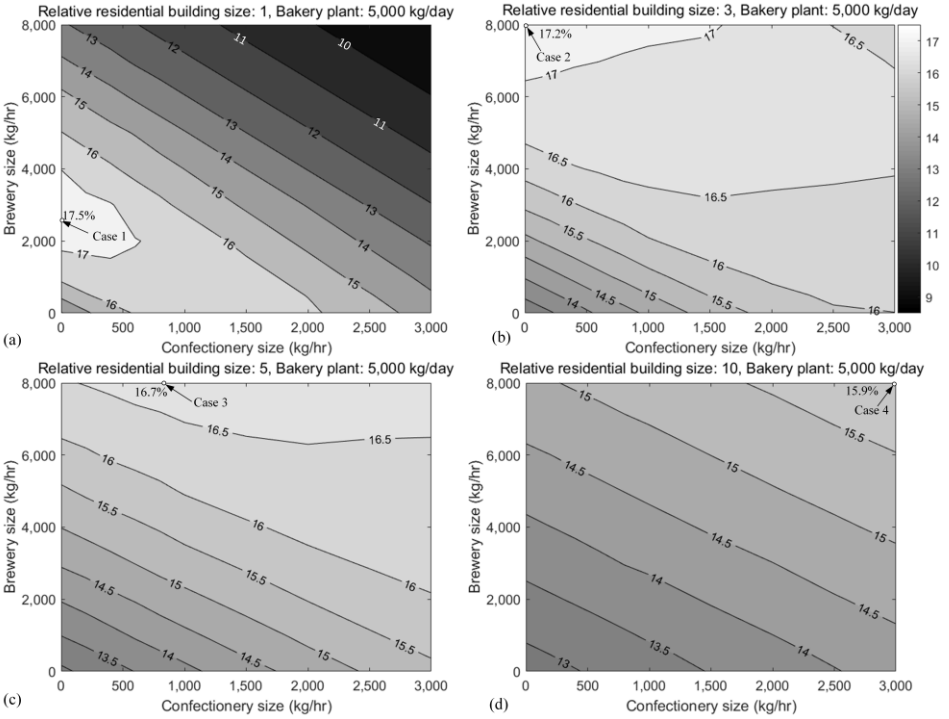


Figure 2 – GHGD % of the integrated system under various entity sizes.

According to the results shown in Figure 2 and Table 1, there is a small difference between the maximum GHGD % and the highest GHGD % when increasing the relative size of the residential building. The results indicate requirements on entity size do not significantly impact GHGD % of the overall integrated system. With requirements on entity sizes, the optimization problem can be formulated by adding additional constraints to define the desired entity sizes and find the optimal sizes of other entities.

Results also show that as the relative size of the residential building increases, the highest GHGD % decreases. It is due to upper bounds on industrial capacities force the integrated system to operate in a way, which deviates from its optimal relative entity sizes and optimal operation pattern. Thus, the maximum GHGD % cannot be held.

With an increase in the size of the residential building, the size of the bakery plant reaches its upper bound first (5,000 kg/day), followed by the brewery (8,000 kg/hr) and the

confectionery plant (3,000 kg/hr). It is due to the bakery plant performing as a critical energy supplier in the integrated system, where all heat transfer and 20.2 % electricity transfer is dispatched by the bakery plant under the optimal entity sizes (Case 1). The brewery reaches its maximum size before the confectionery plant, which indicates the brewery is more suitable for the integrated operation. Compared to the confectionery plant, the brewery requires heating at lower temperatures. Thus, the brewery has a better ability to use the transferred heat.

Table 1 – Highest GHGD % under different relative sizes of the residential building.

	Relative size of residential building	Bakery plant (kg/day)	Brewery (kg/hr)	Confectionery plant (kg/hr)	GHGD %
Case 1	1	5,000	2,720	0	17.5 %
Case 2	3	5,000	8,000	0	17.2 %
Case 3	5	5,000	8,000	800	16.7 %
Case 4	10	5,000	8,000	3,000	15.9 %

5. Conclusions

This work quantifies reductions in GHG emissions that can be achieved by cross-sector integration of energy systems. Even if the energy systems within each sector are optimized for the lowest GHG emissions within that sector, further reduction in GHG emissions can be accomplished by integration between the sectors (residential buildings, commercial buildings, light industries, and electric vehicles). Each entity in the integrated system implements an independently operating combined cooling, heating, and power (CCHP) system, where there are heat and electricity transfers among entities. The optimal design and operation of energy systems are determined for equipment in each entity, the optimal production rate of plants, and the optimal relative size of entities, considering temperatures of heating demands.

Results from case studies on an integrated system with a residential building, a supermarket, a confectionery plant, a bakery plant, and a brewery show the integrated operation can lead to a maximum GHGD % of 17.5 %. If optimizing sizes of entities, the highest GHGD % can be maintained between 15.9 % and 17.5 %, even when there are requirements on sizes of specific entities and the integrated system deviates from its optimal relative entity sizes and operation. Future studies, which include more types of entities and consider partial load effects on equipment efficiency are worth investigating.

References

Fichera, Alberto, Mattia Frasca, and Rosaria Volpe. 2017. "Complex Networks for the Integration of Distributed Energy Systems in Urban Areas." *Applied Energy* 193: 336–45.

Ghorab, Mohamed. 2019. "Energy Hubs Optimization for Smart Energy Network System to Minimize Economic and Environmental Impact at Canadian Community." *Applied Thermal Engineering* 151(June 2018): 214–30.

Singh, R.P. 1986. "Energy Accounting of Food Processing Operations." In *Energy in Food Processing*, ed. R.P. Singh. New York: Elsevier Science Publishing Company Inc., 19–68.

Sullivan, Brendan. 2020. "A Comparison of Different Heating and Cooling Energy Delivery Systems and the Integrated Community Energy and Harvesting System in Heating Dominant Communities." McMaster.

Therkelsen, Peter, Eric Masanet, and Ernst Worrell. 2014. "Energy Efficiency Opportunities in the U.S. Commercial Baking Industry." *Journal of Food Engineering* 130: 14–22.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50331-6

Sustainable Design of Hybrid Energy Systems for Net Zero Carbon Emission

Xueyu Tian[a*], Fengqi You[a]
[a] *Cornell University, Ithaca, New York, 14853, USA*
xt93@cornell.edu

Abstract

This article addressed the sustainable design of carbon-neutral energy systems with earth source heat, lake source cooling, on-site electricity generation, and peak heating systems. A multi-period optimization model given time horizon and temporal resolution is built based on the proposed superstructure of carbon-neutral energy systems to minimize the total annualized cost. The aim is to determine the optimal design of the carbon-neutral energy systems in the target region, seasonal operations, energy mix, and corresponding capacity of each base-load and peak-load technology involved while fulfilling the seasonal demand for electricity, heat, and cooling. The applicability of the proposed modeling framework is illustrated through case studies using Cornell University as the living laboratory.

Keywords: carbon neutrality; energy systems; renewables; decarbonization.

1. Introduction

The Paris Agreement sets a goal to curb global greenhouse gas (GHG) emissions, driving vast penetration of renewable energy worldwide. Extensive research on deep decarbonization of energy systems is conducted at the city-level (Wiryadinata et al., 2019), state-level (Zhao and You, 2020), and country-level (Vaillancourt et al., 2017). Electrification of heat and cooling generation and decarbonization of electricity generation is identified as a promising lever to address the ambitious climate goals (de Chalendar et al., 2019). However, heat and cooling generation stand a chance to destroy the stability of the power system due to the surge in electric load involved if they are electrified in an uncontrolled way (Sánchez-Bautista et al., 2017). Therefore, it seems to be a reliable and promising decarbonization option by exploring renewable heat and cooling generation technologies rather than simply using electrified counterparts (Gong and You, 2015). Among the vast array of renewable heat and cooling generation technologies, geothermal energy and deep water source cooling system show great potentials for the decarbonization transition of energy systems (Lee et al., 2019). Recent research efforts have also identified the values of green hydrogen (Dodds et al., 2015), large-scale heat pumps (Bach et al., 2016), biomass and biogas (Kassem et al., 2020), and thermal energy storage (Ochs et al., 2020) for decarbonizing the heating system. There is a lack of studies addressing the sustainable design of energy systems toward carbon neutrality by simultaneously exploring renewable electricity, heat and cooling generation, and electrified heating and cooling options in the region with a humid continental climate, such as New York State (NYS) (Zhao and You, 2021). In this paper, a multi-period optimization model, given time horizon and temporal resolution for total annualized cost (TAC) minimization, is built. The aim is to determine the optimal design of the carbon-neutral energy systems in the investigated region, seasonal operations, energy mix, and corresponding capacity of each base-load and peak-load

technology involved. The applicability of the proposed modeling framework is illustrated through case studies developed using the data from the Cornell campus.

2. Problem Statement and Model Summary

We are given a superstructure of carbon-neutral energy systems, including a set of renewable electricity generation technologies and a set of renewable and electrified heating and cooling options, as shown in Figure 1. To capture the optimal design, seasonal operations, energy mix, and corresponding capacity of each base-load and peak-load technology in the carbon-neutral energy systems, a time horizon and a set of time periods are specified to improve the temporal resolution of the model. The multi-period optimization problem of the proposed carbon-neutral energy system with earth source heat, lake source cooling (LSC), on-site electricity generation, and peak heating options for the total annualized cost minimization is formally defined in this section. The aim is to determine the optimal design of the carbon-neutral energy systems in the target region, seasonal operations, energy mix, and the corresponding capacity of each base-load and peak-load technology involved while fulfilling the seasonal demand for electricity, heat, and cooling.

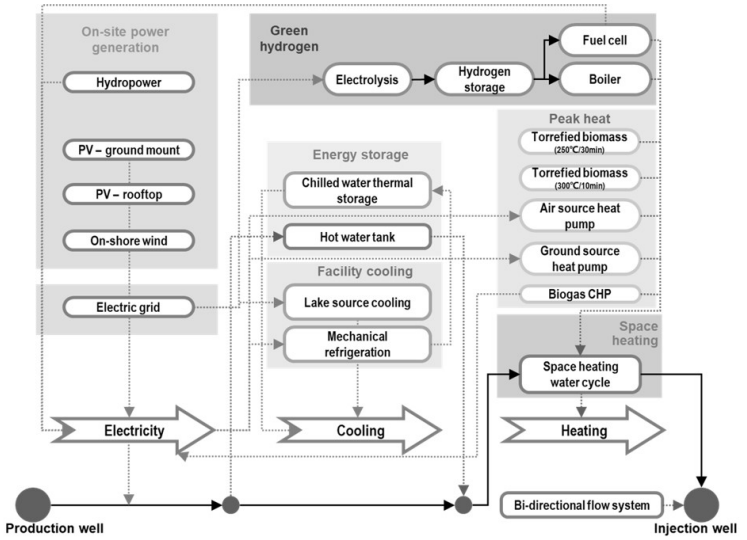


Figure 1. Superstructure of the proposed carbon-neutral energy systems.

The general multi-period optimization model is subjected to the mass balance and configuration constraints, energy balance constraints, logic constraints, and techno-economic evaluation constraints. The integer decision variables represent the selection of technologies. The number of geothermal well-pairs is an integer decision variable (Tian et al., 2019). Other essential decision variables such as the mass flow rates, energy flows, and capacities are continuous variables. The objective function, total annualized cost, includes integer variables such as the numbers of production wells and injection wells and thus is a mixed-integer function. The nonlinear terms mainly come from the separable concave terms induced by the economy of scale. Therefore, the resulting problem is a mixed-integer nonlinear programming (MINLP) problem. The general form of this MINLP problem is summarized as follows.

$$\min \quad TAC = AIC + AOC + RE \tag{1}$$

s.t. mass balance and configuration constraints;
energy balance constraints;
logic constraints;
techno-economic evaluation constraints;

where *AIC*, *AOC*, and *RE* refer to the annualized investment cost, annual operating cost, and replacement cost, respectively.

3. Global Optimization Strategy

The resulting MINLP problem embraces both integer and continuous variables, along with nonlinear functions, so the global optimization of this problem is likely to be computationally challenging for general-purpose global optimization solvers. A global optimization strategy is utilized to solve the proposed MINLP problem efficiently (Gong and You, 2018). Specifically, we substitute the separable concave functions induced by the economy of scale for capital investment estimation with successive piecewise linear relaxations. The resulting MINLP problem is solved iteratively following the tenet of the branch-and-refine algorithm (You and Grossmann, 2011). The pseudocode of the global optimization algorithm is presented in Figure 2. *ub* and *lb* stand for the upper and lower bound, respectively.

Global Optimization Algorithm

$lb \leftarrow 0$, $ub \leftarrow +\infty$, $gap \leftarrow +\infty$, $iter \leftarrow 0$;
Initialize the problem with one-piece approximation functions;
while $gap \geq tol$ **do**
  $iter \leftarrow iter + 1$;
  Solve the problem for the optimal solution m^* and objective function value OBJ^*;
  Estimate the objective function value OBJ' of the original problem based on m^* and concave functions;
  $lb \leftarrow \max\{lb, OBJ'\}$, $ub \leftarrow \min\{ub, OBJ^*\}$, $gap \leftarrow (ub - lb) / ub$;
  Update the problem by adding a new partition point m^*;
end while
return m^*

Figure 2. The pseudocode of the global optimization algorithm.

4. Application to Cornell University Campus Energy Systems

The proposed multi-period optimization modeling framework for energy systems decarbonization is applied to address the optimal design of the carbon-neutral energy systems using the main campus of Cornell University located in Ithaca as the living laboratory. Based on the optimization results, the optimal configuration of the carbon-neutral energy system in the target region, seasonal operations, energy mix, and corresponding capacity of each base-load and peak-load technology involved are determined while accommodating the seasonal demand for electricity, heat, and cooling across the main campus of Cornell University located in Ithaca, NYS. Three case studies are developed based on the real data from the main campus of Cornell University located in Ithaca with a consideration of different scopes of peak-load technologies (Tian et al., 2019). The first case study aims to obtain the global optimal solution of the multi-period optimization problem with a monthly model resolution for the proposed carbon-neutral energy system with earth source heat, LSC, on-site electricity generation, and peak heating options, including biomass or biogas heating, heat pumps, green hydrogen, and thermal storage. The second one explicitly excludes

biomass or biogas as the peak heating options to evaluate the economic potential of electrified peak heating systems based on heat pumps.

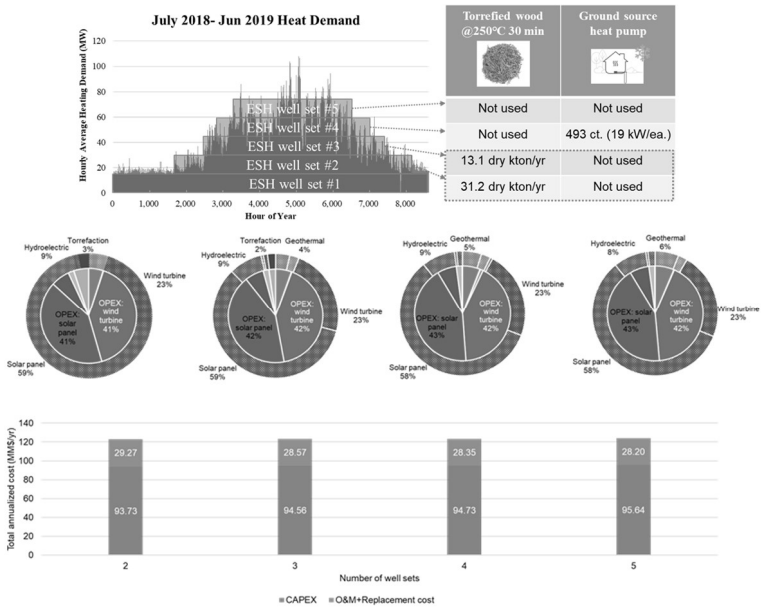


Figure 3. Selection of technologies and economic performance for case study 1.

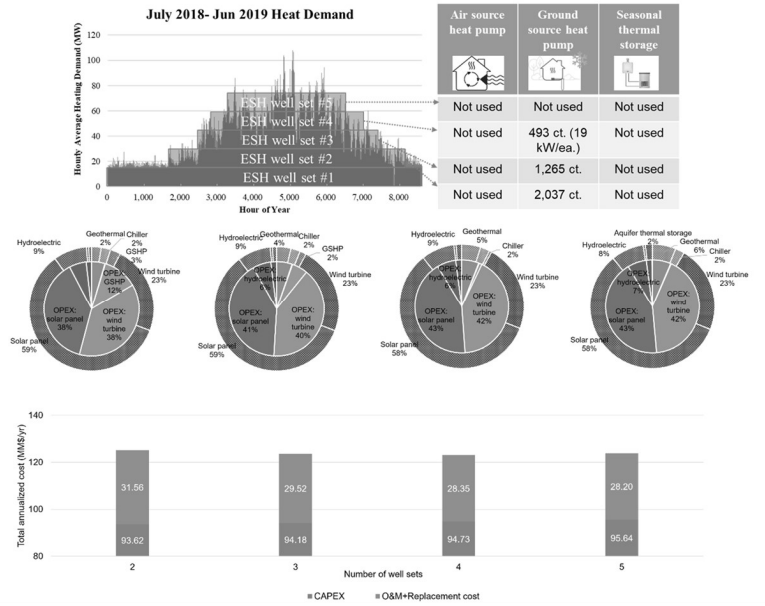


Figure 4. Selection of technologies and economic performance for case study 2.

Figure 3 shows the selection of technologies of the global optimal solution and the corresponding economic performance. When the capacity of base-load earth source heat

is low, burning torrefied biomass is selected as the optimal peak heating technology through optimization. However, as the number of the geothermal well-pairs attains four, the ground source heat pump outperforms the torrefied biomass from the economic perspective. No peak heating technologies are needed when five geothermal well-pairs serve as the base-load heat supplier based on a monthly model resolution. Specifically, the annual consumption of torrefied biomass is 31.2 dry kton and 13.1 dry kton for the two-well-set case and three-well-set case, respectively. Overall, torrefied biomass outperforms the ground source heat pumps when the base-load capacity is low. We note that as the number of geothermal well-pairs increases from two to five, the annualized investment cost increases from \$93.73 MM/yr to \$95.64 MM/yr. The annual operating costs corresponding to the four cases are \$29.27 MM/yr, \$28.57 MM/yr, \$28.35 MM/yr, and \$28.20 MM/yr, respectively. When the number of base-load geothermal well-pairs equals two and three, torrefied biomass treated at 250 ℃ for 30 minutes is selected, accounting for 3% and 2% of the annualized investment cost, respectively. When the number of geothermal well-pairs attains four, 493 ground source heat pumps with a typical capacity of 19 kW in the North American region are employed to address the peak-load heat demand.

Figure 4 demonstrates the peak heating options with different base-load capacities, where only ground source heat pumps are employed with no thermal energy storage. Specifically, 2,037, 1,265, and 493 ground source heat pumps with a typical capacity of 19 kW for each in the North American region are deployed to handle the peak-load heat demand for the cases with two, three, and four base-load earth source heat pumps, respectively. The number of geothermal well-pairs is chosen as the investigated input parameter, ranging from two to five. We note that as the number of geothermal well-pairs increases from two to five, the annualized investment cost increases from \$93.62 MM/yr to \$95.64 MM/yr, while the annual operating cost decreases from \$31.56 MM/yr to \$28.20 MM/yr. In terms of capital investment, solar panels (58%-59%), wind turbines (23%), and hydroelectric power plant (8%-9%) are the major contributors. When the base-load earth source heat capacity is low, the operating cost associated with the ground source heat pump is more pronounced. The remaining annual operating cost is mainly sourced from the operations of solar and wind farms.

When the capacity of base-load earth source heat is low, i.e., the number of geothermal well-pairs equalling two, both green hydrogen and hot water tanks are needed through optimization with capacities of 53.5 GWh and 35.5 GWh, respectively. As the geothermal well-pairs attains three and four, green hydrogen is no longer needed to pursue the lowest total annualized cost. When five geothermal well-pairs serve as the base-load heat supplier based on a monthly model resolution, i.e., some short-term peak-load demands are neglected by considering a monthly average, no peak heating technologies are needed. Specifically, the capacity of hot water tanks for the three-well-set case and four-well-set cases are 41.4 GWh and 7.0 GWh, respectively. We note that as the number of geothermal well-pairs increases from two to five, the annualized investment cost decreases from \$191.33 MM/yr to \$95.64 MM/yr, while the annual operating costs decrease from \$57.85 MM/yr to \$28.20 MM/yr. When the number of base-load geothermal well-pairs equals two, both green hydrogen and hot water tank are in need to manage the peak-load heat demand, which accounts for 42% and 11% of the annualized investment cost, respectively. In addition to the operating cost of solar and wind farms, other operating costs are associated with green hydrogen and the replacement of hot water tanks. The total annualized cost corresponding to the two-

well-set case is substantially higher than the other cases by a factor of 1.6 - 2.0 due to the high capital investment of hydrogen generation through electrolysis and storage.

5. Conclusion

A multi-period optimization model, given time horizon and temporal resolution for total annualized cost minimization, was built. The aim is to simultaneously determine the optimal design of the carbon-neutral energy systems in the investigated region, seasonal operations, energy mix, and corresponding capacity of each base-load and peak-load technology involved while fulfilling the seasonal demand for electricity, heat, and cooling. The applicability of the proposed modeling framework was illustrated through three case studies developed by leveraging the real-world data from the main campus of Cornell University, located in Ithaca, NYS.

References

B. Bach, J. Werling, T. Ommen, et al., 2016, Integration of large-scale heat pumps in the district heating systems of Greater Copenhagen. Energy, 107, 321-334.

J.A. de Chalendar, P.W. Glynn, S.M. Benson, 2019, City-scale decarbonization experiments with integrated energy systems. Energy & Environmental Science, 12(5), 1695-1707.

P.E. Dodds, I. Staffell, A.D. Hawkes, et al., 2015, Hydrogen and fuel cell technologies for heating: A review. International Journal of Hydrogen Energy, 40(5), 2065-2083.

J. Gong, F. You, 2015, Sustainable design and synthesis of energy systems. Current Opinion in Chemical Engineering, 10, 77-86.

J. Gong, F. You, 2018, A new superstructure optimization paradigm for process synthesis with product distribution optimization: Application to an integrated shale gas processing and chemical manufacturing process. AIChE Journal, 64, 123-143.

N. Kassem, J. Hockey, C. Lopez, et al., 2020, Integrating anaerobic digestion, hydrothermal liquefaction, and biomethanation within a power-to-gas framework for dairy waste management and grid decarbonization. Sustainable Energy & Fuels, 4(9), 4644-4661.

I. Lee, J.W. Tester, F. You, 2019, Systems analysis, design, and optimization of geothermal energy systems for power production and polygeneration: State-of-the-art and future challenges. Renewable and Sustainable Energy Reviews, 109, 551-577.

F. Ochs, A. Dahash, A. Tosatto, M.B. Janetti, 2020, Techno-economic planning and construction of cost-effective large-scale hot water thermal energy storage for Renewable District heating systems. Renewable Energy, 150, 1165-1177.

A.d.F. Sánchez-Bautista, J.E. Santibañez-Aguilar, et al., 2017, Optimal Design of Energy Systems Involving Pollution Trading through Forest Plantations. ACS Sustainable Chemistry & Engineering, 5, 2585-2604.

X. Tian, T. Meyer, H. Lee, et al., 2020, Sustainable design of geothermal energy systems for electric power generation using life cycle optimization. AIChE Journal, 66, e16898.

X. Tian, F. You, 2019, Carbon-neutral hybrid energy systems with deep water source cooling, biomass heating, and geothermal heat and power. Applied Energy, 250, 413-432.

K. Vaillancourt, O. Bahn, E. Frenette, et al., 2017, Exploring deep decarbonization pathways to 2050 for Canada using an optimization energy model framework. Applied Energy, 195, 774.

S. Wiryadinata, J. Morejohn, K. Kornbluth, 2019, Pathways to carbon neutral energy systems at the University of California, Davis. Renewable Energy, 130, 853-866.

F. You, I.E. Grossmann, 2011, Stochastic inventory management for tactical process planning under uncertainties: MINLP models and algorithms. AIChE Journal, 57(5), 1250-1277.

N. Zhao, F. You, 2020, Can renewable generation, energy storage and energy efficient technologies enable carbon neutral energy transition? Applied Energy, 279, 115889.

N. Zhao, F. You, 2021, New York State's 100% renewable electricity transition planning under uncertainty using a data-driven multistage adaptive robust optimization approach with machine-learning. Advances in Applied Energy, 2, 100019.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50332-8

Prediction of Charge / Discharge Behavior of Tri-Electrode Zinc-air Flow Battery Using Linear Parameter Varying Model

Woranunt Lao-atiman[a], Amornchai Arpornwichanop[a,b] and Soorathep Kheawhom[a,b*]

[a] *Department of Chemical Engineering, Faculty of Engineering, Chulalongkorn University, Thailand*
[b] *Center of Excellence in Process and Energy Systems Engineering, Chulalongkorn University, Bangkok, Thailand*
**soorathep.k@chula.ac.th*

Abstract

Currently, energy storage systems are in the research spotlight as they can support the application of renewable energy. Owing to their high energy density and low cost, zinc-air flow batteries (ZAFBs) are seen to have great potential for use as renewable energy storage devices. However, the battery management system (BMS) for ZAFBs is still underdeveloped as a precise prediction of their nonlinear behavior is required. To overcome this drawback, a linear parameter varying (LPV) model is established via a multiple linear time-invariant (LTI) model with charge/discharge current and state of charge (SOC) as scheduling parameters. Validation of the developed model is carried out using various battery data from different experimental batches. According to the discharge current and SOC, results demonstrate that the nonlinear behavior of the ZAFBs can be predicted by the LPV model developed. Thus, the LPV model is found to be comparable with the linear model for local accuracy. In the case of global accuracy, it is seen that the LPV model outperformed the linear model. Such a result reveals the ability of the LPV model to predict the dynamics of the ZAFBs and their feasibility for use in the BMS.

Keywords: Zinc-air battery; battery modeling; Linear-parameter varying model.

1. Introduction

Nowadays, renewable energy technologies have attracted widespread interest due to the increase in energy consumption and the critical environmental crisis. Renewable energy sources such as solar and wind energy are intermittent by nature. Therefore, the power generated from these energy sources is found to be inconsistent. To address this issue, energy storage systems (ESSs) exhibit great potential because they can provide the stability for energy utilization. Zinc-air batteries (ZABs) are promising candidates for ESSs due to their high energy density and low cost (Lao-atiman et al., 2019, Radenahmad et al., 2021, Khezri et al., 2020).

In battery research, BMS is generally studied as it can enhance the safety and operability of the battery (Pop et al., 2008). One of the research aspects involving the development of BMS is modeling. Battery modeling can be done in a variety of ways. For example, simulation using a theoretical continuum model has been conducted to analyze the phenomena occurring inside the battery (Schröder and Krewer, 2014, Maia et al., 2017). The empirical model has also been studied for use in BMS in various types of batteries because this type of model is suitable for online prediction due to its speed

and simplicity of calculation. Recently, a LPV model has been proposed to predict the discharge behavior of ZABs (Lao-atiman et al., 2020). Another example is the data-driven model developed for lithium-ion battery health monitoring (Sukanya et al., 2021). One important feature of BMS, which must be established, is battery state estimation, especially SOC estimation (Chang, 2013). As regards SOC estimation, an empirical model has also been used with various adaptive filters such as state observer (Hu and Yurkovich, 2012) or Kalman filter-based estimator (Wassiliadis et al., 2018). Nevertheless, studies of empirical models involving SOC changes are still inadequate in ZAB research.

This work aims to develop the LPV model and use it to predict the dynamic behavior of tri-electrode ZAFBs including the influence of charge/discharge current and SOC. Battery data used for dynamic identification and validation have been measured from the laboratory-made tri-electrode ZAFB. Linear state-space models identified from different conditions have been used to create the LPV model, which has discharge current level and SOC as scheduling parameters. After that, the LPV model was tested for validity using battery data from different experimental batches.

2. Description of tri-electrode ZAFB and experimental data

As shown in Figure 1A, the laboratory made ZAFB in this work was designed as a tubular cylinder cell. The cylinder support structure was made of poly vinyl chloride (PVC). The cell is circulated by an electrolyte, which is an 8 M solution of potassium hydroxide (KOH) having 0.5 M zinc oxide (ZnO). The anode active material is Zn electroplated onto the current collector, which is nickel (Ni) foam. The cathode active material is oxygen from the ambient air. The oxygen reduction reaction (ORR) occurred at the cathode current collector, which is Ni foam coated with a catalytic layer and gas diffusion layer. The charging electrode is made of Ni foam.

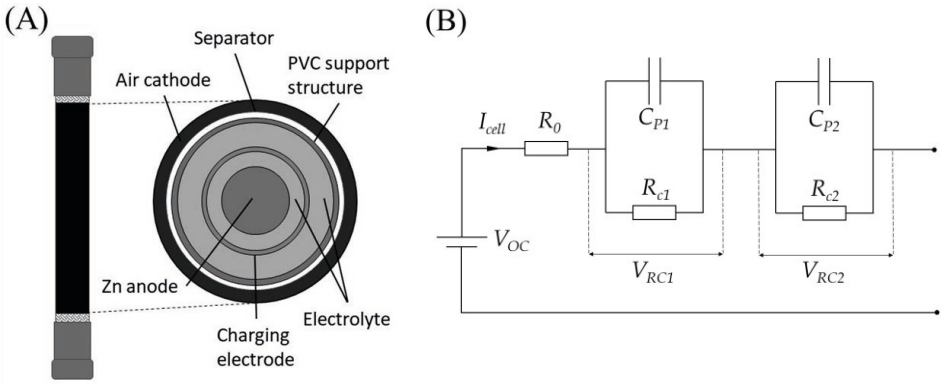


Figure 1 (A) Schema of laboratory-made tri-electrode ZAFB and (B) Electrical equivalent circuit diagram of battery: second-order RC model.

To obtain the experimental data, battery testing equipment (NEWARE, CT-4008-5V20mA, Neware Technology Ltd., Shenzhen, China) was used. Sampling time was 1 s. The used data are battery response data with current as input and voltage as output. For the dynamic identification, the pattern of the used profile was charge / discharge for 50 mAh alternating with a rest period of 5 min. The procedure was repeated until the voltage cutoff of 0.5 V was met for discharging or the capacity reached 500 mAh for charging. The applied current values were 500, 1000, 1500 and

2000 mA. As regards model validation, different sets of random charging and discharging data were used.

For the nomenclature of data, the data starting with DSOC is the discharge steps data and the data starting with CSOC is the charge steps data. The following numbers define the discharge current, for example, the data DSOC500A contain discharge steps with a current of 500 mA for 50 mAh alternating with a rest for 5 min. The data named MULTI A and MULTI B are discharge steps data with multiple current levels. Likewise, CMULTI refers to the charge steps data with multiple discharge currents.

3. LPV modeling

The LPV model is a time-variant model whereby parameters are varied as a function of scheduling parameters (p). The LPV model was constructed from a set of discrete state space model. The model contains 3 states: V_{RC1}, V_{RC2} and SOC. Both V_{RC1} and V_{RC2} represent the overpotential of the zinc electrode and counter electrode (air or charging), respectively. SOC represents the state of charge of the battery. SOC is included as a state because the model is used with a state estimator to estimate SOC in the next part. To make the LPV model more practical, the model was established having an equivalent circuit second-order RC model, as illustrated in Figure 1B. The LPV model can be written as follows:

$$\begin{bmatrix} V_{RC1}(k+1) \\ V_{RC2}(k+1) \\ SOC(k+1) \end{bmatrix} = \begin{bmatrix} A_1 & 0 & 0 \\ 0 & A_2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} V_{RC1}(k) \\ V_{RC2}(k) \\ SOC(k) \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \\ \frac{\Delta t}{3600 \cdot C_n} \end{bmatrix} I_{cell}(k) \quad (1)$$

$$V_{cell}(k) = V_{oc} + \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} V_{RC1}(k) \\ V_{RC2}(k) \\ SOC(k) \end{bmatrix} + D \cdot I_{cell}(k) \quad (2)$$

where A_1, A_2, B_1, B_2 and D are state space parameters. V_{cell}, V_{OC} and I_{cell} are cell voltage, open circuit voltage and cell current, respectively. Δt and C_n are sampling time and nominal capacity, respectively.

The calculation of SOC in Eq.(1) is based on the coulomb counting (CC) method. Eq.(2), which includes V_{OC}, ensures that the model is able to calculate cell voltage; V_{cell}. I_{cell} and SOC have been selected as scheduling parameters for A_1, A_2, B_1, B_2 and D. However, some arbitrary assumptions have been made in order to adapt the model with the scenarios of battery data. Firstly, for discharging, when SOC decreased, Zn at the electrode fully depleted. Thus, V_{RC1} is affected but V_{RC2} is not affected by the SOC change. This outcome enabled both A_2 and B_2 to become functions of only the current level for discharging. The next assumption is that the SOC effect on the overpotential is less significant for both electrodes when the battery is charged. Hence, B_1 and B_2 become functions of the current level for charging. The last assumption made infers that the internal resistance of the system is independent of current (Zhong et al., 2021). As seen in Figure 1B, D is equivalent to R_0 which is related to ohmic resistance. D is, therefore, assumed to be a function of SOC for both discharging, i.e., $A_2(I)$, $B_2(I)$ and $D(SOC)$ and for charging, i.e., $B_1(I)$, $B_2(I)$ and $D(SOC)$.

Regarding varying model parameters, the correlations between model parameters and scheduling parameters were constructed from the identified model parameters. For instance: model parameters as identified from the discharge step data with a discharge current of 1000 mA and SOC of 0.5 demonstrated scheduling current

levels and SOCs of 1 A and 0.5, respectively. After the correlations were accounted for, the LPV model was validated according to the various dataset, including data obtained via a different batch of the experiment. For comparison purposes, the linear state space model was introduced. The expression of the linear model is the same as Eqs. 1 and 2 but the model parameters are not varied with scheduling parameters.

4. Results and discussion

As mentioned previously, the LPV model for the tri-electrode ZAFB was developed from the state space model as a based linear time-invariant (LTI) model. The state space parameters of the identified model were fitted to make the correlation with the scheduling parameters. From Eqs.(1) and (2), five parameters enabled the correlations to be set up viz. A_1, A_2, B_1, B_2 and D. There were also two scheduling parameters, including current level and SOC; the fitted correlations became surface functions. Nevertheless, some model parameters are functions of only one scheduling parameter according to the assumption made in the previous section. Additionally, V_{OC} has also varied with SOC. Therefore, a correlation with V_{OC} was also made. In Table 1, a list of the functions used to fit the correlations for both discharging and charging are tabulated.

Table 1 Functions used to fit the model parameter correlations.

Fitting function	parameter
Discharging	
$\alpha exp(\beta(SOC) + \gamma(I_{cell})) + \delta exp(\varepsilon(I_{cell}) + \theta(SOC)) + \vartheta$	A_1
$\alpha exp(\beta(SOC) + \gamma(I_{cell})) + \delta exp(\varepsilon(I_{cell}) + \theta(SOC))$	B_1
$\alpha exp(\beta(I_{cell})) + \gamma exp(\delta(I_{cell}))$	A_2, B_2
$\alpha exp(\beta(SOC)) + \gamma exp(\delta(SOC))$	D
Charging	
$\mu_{00} + \mu_{10}(SOC) + \mu_{01}(I_{cell}) + \mu_{11}(SOC)(I_{cell}) + \mu_{02}(I_{cell})^2$	A_1, A_2
$\alpha exp(\beta(I_{cell})) + \gamma exp(\delta(I_{cell}))$	B_1, B_2
$\alpha exp(\beta(SOC)) + \gamma exp(\delta(SOC))$	D
Open Circuit Voltage	
$\alpha exp(\beta(SOC)) + \gamma exp(\delta(SOC))$	V_{OC}

Regarding validation of the LPV model, the model was tested by predicting various response data. Testing data included the same data used to identify the LTI model and the data obtained from a different batch of the experiment. In Figure 2, the fit percentages of the prediction of the LPV model and linear model are shown. Results revealed that the LPV model was more accurate than the linear model in most cases for both discharging and charging. Such an outcome occurred due to the effect of SOC change in the LPV model. Thus, the LPV model proved to be more accurate than the linear model over a wider range of SOC. For further validation, the model was tested with data obtained from a different batch of the experiment including the data named MULTI A and MULTI B for discharging and CMULTI for charging. In Figure 2, it was observed that the fit percentage of LPV model prediction was acceptably high although the data were obtained from a different experimental batch.

In Figure 3, graphical prediction results for data MULTI A and CMULTI are displayed. In Figure 3A, for data MULTI A, prediction errors were still observed at some current levels. The highest error occurred upon discharging near the battery

depletion zone. Besides, the LPV model performed adequately in predicting the response at the resting zone. As regards CMULTI data, a high error occurred at the beginning range of the predicted response. Such an error may have arisen from the mismatch between the V_{OC} correlation of the model and the resting voltage of this data. It is evident that the LPV model was quite accurate as it was able to address the effect of different currents and SOC changes. This outcome demonstrated the potential of the LPV model and its feasibility for use in SOC estimation with a model-based state estimation algorithm.

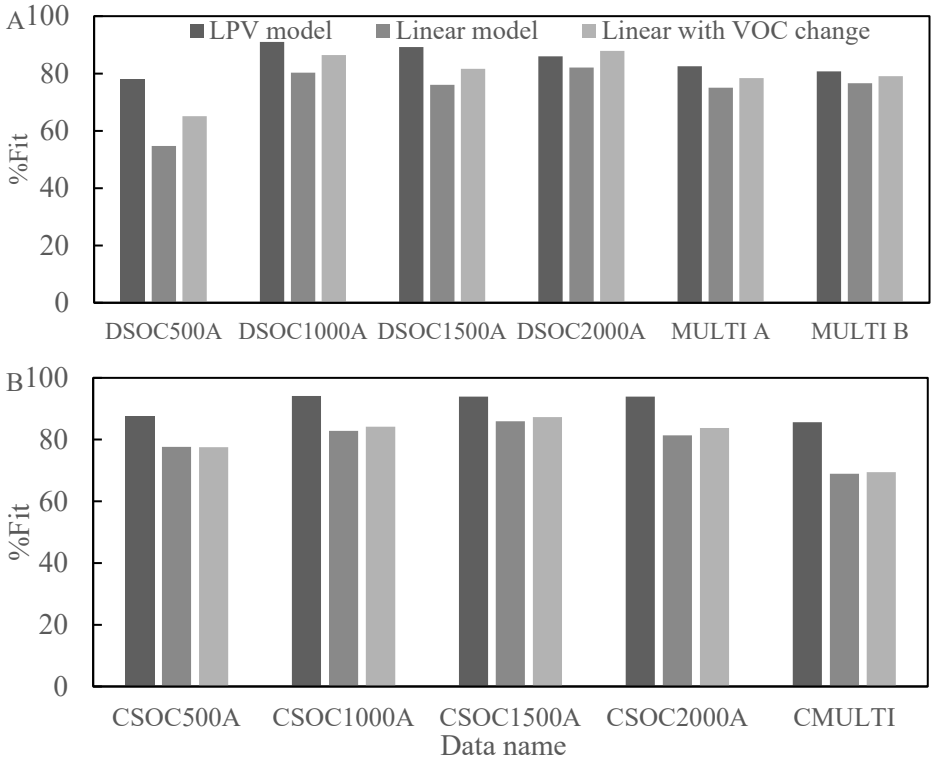


Figure 2 Comparison of fit percentage of model prediction between various models and data for (A) discharging and (B) charging. Fit % can be expressed as: $100 \times \left(1\text{-}\frac{mean|y\text{-}\hat{y}|}{mean|y\text{-}mean(y)|}\right)$.

5. Conclusion

An LPV model for a tri-electrode ZAFB was proposed. Thus, parameters of the LTI model at various current and SOC values were used to construct the correlations for the LPV model. The experimental data used for dynamic identification and validation were obtained from the laboratory-made tri-electrode ZAFB. Results of the validation confirmed that the LPV model was able to predict the behavior of the ZAFB. Moreover, it was found that the LPV model was more accurate than the linear model based on the comparison using normalized mean absolute error. The proposed LPV model was also able to handle the effect of SOC changes as SOC was one of the scheduling parameters of the model. Overall, results show that the LPV model is a promising dynamic model for predicting the dynamic behavior of a tri-electrode ZAFB.

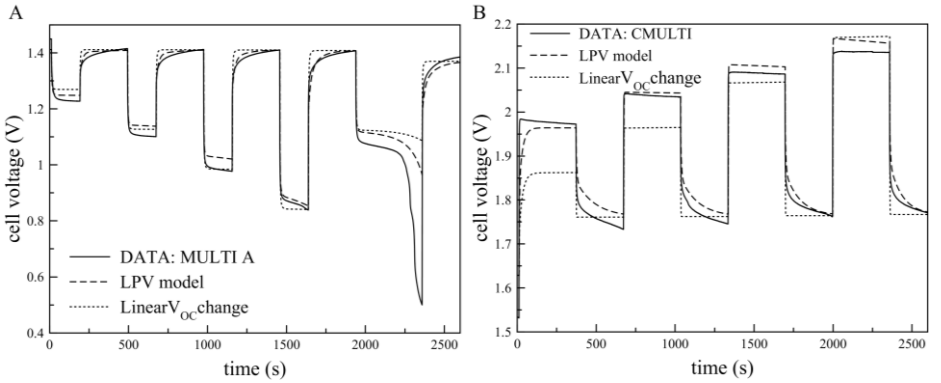


Figure 3 Comparison of response between measured data, predicted data from LPV model and linear model: (A) multiple discharge current steps and (B) multiple charge current steps.

References

W.-Y. Chang, 2013, The State of Charge Estimating Methods for Battery: A Review, ISRN Applied Mathematics, 2013**,** 7.

Y. Hu & S. Yurkovich, 2012, Battery cell state-of-charge estimation using linear parameter varying system techniques, Journal of Power Sources, 198**,** 338-350.

R. Khezri, S. Hosseini, A. Lahiri, S. R. Motlagh, M. T. Nguyen, T. Yonezawa & S. Kheawhom, 2020, Enhanced Cycling Performance of Rechargeable Zinc–Air Flow Batteries Using Potassium Persulfate as Electrolyte Additive, International Journal of Molecular Sciences, 21**,** 7303.

W. Lao-Atiman, S. Olaru, A. Arpornwichanop & S. Kheawhom, 2019, Discharge performance and dynamic behavior of refuellable zinc-air battery, Scientific Data, 6**,** 168.

W. Lao-Atiman, S. Olaru, S. Diop, S. Skogestad, A. Arpornwichanop, R. Cheacharoen & S. Kheawhom, 2020, Linear parameter-varying model for a refuellable zinc-air battery, Royal Society Open Science, 7**,** 201107.

L. K. K. Maia, Z. Güven, F. La Mantia & E. Zondervan, 2017, Model-based Optimization of Battery Energy Storage Systems, In: ESPUÑA, A., GRAELLS, M. & PUIGJANER, L. (eds.) Computer Aided Chemical Engineering: Elsevier.

V. Pop, H. Bergveld, D. Danilov, P. Regtien & P. Notten, 2008, Battery Management Systems: Accurate State-of-Charge Indication for Battery-Powered Applications.

N. Radenahmad, R. Khezri, A. A. Mohamad, M. T. Nguyen, T. Yonezawa, A. Somwangthanaroj & S. Kheawhom, 2021, A durable rechargeable zinc-air battery via self-supported MnOx-S air electrode, Journal of Alloys and Compounds, 883**,** 160935.

D. Schröder & U. Krewer, 2014, Model based quantification of air-composition impact on secondary zinc air batteries, Electrochimica Acta, 117**,** 541-553.

G. Sukanya, R. Suresh & R. Rengaswamy, 2021, Data-driven prognostics for Lithium-ion battery health monitoring, In: TÜRKAY, M. & GANI, R. (eds.) Computer Aided Chemical Engineering: Elsevier.

N. Wassiliadis, J. Adermann, A. Frericks, M. Pak, C. Reiter, B. Lohmann & M. Lienkamp, 2018, Revisiting the dual extended Kalman filter for battery state-of-charge and state-of-health estimation: A use-case life cycle analysis, Journal of Energy Storage, 19**,** 73-87.

Y. Zhong, B. Liu, Z. Zhao, Y. Shen, X. Liu & C. Zhong, 2021, Influencing Factors of Performance Degradation of Zinc–Air Batteries Exposed to Air, Energies, 14**,** 2607.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50333-X

An optimized resource supply network for sustainable agricultural greenhouses: A circular economy approach

Sarah Namany, Ikhlas Ghiat, Fatima-Zahra Lahlou, Tareq Al-Ansari*
[a]*College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar*
**talansari@hbku.edu.qa*

Abstract

In light of the ever-increasing demand on food products and the associated intensification of agricultural activities, the environment and natural capital are witnessing unprecedented pressures. Alleviating these stresses will require the deployment of more sustainable and resilient food systems that offer a cost efficient and environmentally friendly alternatives to the conventional energy and water intensive food technologies. Greenhouses represent a promising solution to accommodate sustainable food supplies despite uncooperative external climate conditions. However, conventional greenhouses continue to rely on resource inflows originating from unstainable practices and systems such as groundwater abstraction and energy intensive production of fertilisers. Circular economy represents an opportunity to enhance resources utilisation and mitigate environmental burdens associated with their intensive exploitation at competitive costs through deploying novel solutions. In this study, waste integration involving CO_2 enrichment and sequestration along with wastewater reuse in irrigation are investigated for tomato cultivation in a greenhouse system in the State of Qatar. In this regard, a multi-objective optimisation model based on a mixed-integer linear program (MILP) is proposed to determine the optimal technology configuration and supply network that delivers resources for a greenhouse operation. The purpose of this framework is to minimise water and carbon footprints in addition to the total costs associated with resource provision to the greenhouse. Considering results from the pareto front, the optimal distribution requires 54,3579 $/year and generates around 20,043 kg of CO_{2eq}/year. This option suggests a sustainable cultivation alternative attributed to its environmental efficiency with regards to water savings, CO_2 offsetting, and fertiliser use reduction.
Keywords: Circular economy, greenhouse, wastewater, CO_2 sequestration, optimisation, MILP.

1. Introduction

One of the most pressing global challenges faced nowadays is ensuring food security under restrained resource availability and a continuously growing population. Demand for food is prognosticated to increase by 60% by 2050, which will further increase the consumption of resources, mainly water and energy (Godfray et al., 2010). The agricultural sector already accounts for 70% of the total freshwater withdrawals, and 3.5% to 4.8% of the total energy consumption. As such, there is an impetus to shift agricultural systems to more sustainable practices that will not only produce more food, but also enable an efficient use of resources and reduce the associated environmental impact (Ghiat et al., 2021a, 2020). A key aspect of sustainable agriculture is water irrigation management. Irrigation water is traditionally supplied from unstainable sources, mainly

groundwater or desalinated seawater which is an energy-intensive process. The reuse of wastewater for irrigation requirements represents an opportunity for the agricultural sector, whereby it can alleviate stresses on freshwater resources and curb environmental emissions related to desalination processes as well as the production of fertilisers. The use of treated wastewater for irrigation also provides a solution for fertilisation in which nutrient requirements such as nitrogen (N), phosphorus (P) and potassium (K) can be met. The levels of nutrients and contaminants present in the treated wastewater can vary depending on the sector or type of wastewater (e.g. municipal wastewater, industrial wastewater) and can also vary within the same sector (Lahlou et al., 2021). Several studies investigated the use of treated wastewater to meet plant nutrient requirements and concluded that it can be a valuable solution to substitute industrial fertilisers either partially or completely. The reuse of wastewater instead of conventional discharge can also save water resources and reduce the greywater footprint, because it eliminates the need of diluting wastewater with freshwater resources before sea discharge. In fact, the agricultural sector can tolerate higher levels of nutrients and contaminants, especially N, P, and K, as compared to wastewater discharge in the sea, which renders its use in agriculture not only meet plant water requirements but also their nutrient requirements (García-Delgado et al., 2012; Lahlou et al., 2020; Musazura et al., 2019; Raju and Byju, 2019). The need for sustainable intensification of food systems has also led researchers to investigate novel techniques to enhance production and reduce resource consumption such as CO_2 enrichment. CO_2 levels between 800-1200 ppm in the air have been linked to enhanced yields and reduced water consumption due to improved photosynthesis and lower transpiration rates. In this case, CO_2 is perceived as a commodity which can be supplied from industrial carbon capture systems, compressed, and transported either via pipeline or trucks to agricultural production sites (Ghiat et al., 2021b; Nederhoff, 1994). The expensive costs and complexities related to CO_2 capture and transportation are still a major challenge in the implementation of this practice. Many studies have investigated the supply chain of CO_2 enrichment in agricultural greenhouses in efforts to optimise the economic and environmental costs (Govindan and Al-Ansari, 2019). Similarly, this paper contributes to solving logistical complexities in the supply of novel and sustainable agricultural methods.

Previous studies have not considered the optimised allocation of both treated wastewater and CO_2 captured from wastewater treatment plants (WWTPs) to agricultural production systems. Thus, the objective of this study is to develop an optimisation model for the sustainable supply of water, fertilisers, and CO_2 to agricultural greenhouses all from WWTPs. The presented model can be applied to a network of multiple WWTPs and greenhouses. The optimisation model is applied to a case study in the State of Qatar, comprising of two existing WTTPs and an agricultural greenhouse producing tomatoes. The aim is to supply the greenhouse with the necessary water, nutrients, and CO_2 requirements while; 1) minimising the greywater footprint of wastewater, 2) minimising the carbon footprint, and 3) minimising the economic cost. A mixed integer linear program (MILP) model is presented to allocate the necessary resources to the greenhouse from the WWTPs at a reduced greywater footprint, environmental and economic costs.

2. Methodology

2.1. Model Formulation

The methodology proposed in this study consists of designing a multi-objective optimisation model that investigates the benefits of supplying an agricultural greenhouse

with the required water, fertilisers, and CO_2 for an enhanced crop yield. The model fosters the concept of circular economy by introducing the reuse of treated wastewater for irrigation and fertilisation purposes, in addition to the enrichment of the greenhouse using CO_2 as means to improve the photosynthesis of the grown crops. In order to conduct this optimisation, a multi-objective mixed-integer linear program (MILP) is developed, wherein the first objective is to maximise the amount of greywater saved by the process of reusing wastewater instead of discharging it (equation 1). The second objective consists of minimising the environmental impact associated with the allocation of water and CO_2 to the greenhouse (equation 2). As for the third objective, it aims to minimise the economic costs associated with the reuse of treated wastewater and the CO_2 generated from different sources (equation 3). The mathematical formulation of the model is presented in the following section.

Maximise: $GW = Q_w \sum_{i=1}^{m} \frac{n_i - N_{max}}{(N_{max} - N_w)} x_i^w 10^{-3}$ (1)

Minimise: $EC = \sum_{i=1}^{m} x_i^w c_i^w + \sum_{i=1}^{m} x_i^{CC} c_i^{CC} - Q_w \sum_i^m x_i^w n_i c^N 10^{-3}$ (2)

Minimise: $Env = \sum_{i=1}^{m} x_i^w e_i^w + \sum_{i=1}^{m} x_i^{CC} e_i^{CC} - Q_w \sum_i^m x_i^w n_i CF^N 10^{-3}$ (3)

The decision variables are x_i^w and x_i^{CC} which represent the percentage contribution of different water sources and CO_2 origins to the total requirements of the greenhouse, respectively.

Subject to the following constraints:

$x_i^w;\ x_i^{CC} > 0$ (4)

$\sum_{i=1}^{m} x_i^w = 100\%$ (5)

$\sum_{i=1}^{m} x_i^{CC} = 100\%$ (6)

Such that:

GW is the grey water footprint in m^3;

EC is the total cost in $;

Env is the total environmental impact represented by Global Warming Potential (GWP) emissions in kg of CO_{2eq};

m is the total number of treated wastewater sources;

i is the index of the treated wastewater source;

Q_w is the total quantity of water required for growing the crop inside the greenhouse in m^3;

n_i is the amount of nitrogen content in wastewater coming from the different sources in mg/L;

N_{max} is the maximum allowable concentration of nitrogen in the treated wastewater mg/L;

N_w is the concentration of nitrogen in the freshwater mg/L;

c^N is the cost of 1kg of nitrogen fertilizer in $/kg$_N$;

CF^N is the carbon footprint associated with the production, packaging, and transportation of 1 kg of nitrogen fertilizer in kg-$CO_{2\text{-eq}}$/kg$_N$;

c_i^w and c_i^{CC} are the total economic costs associated with each source of treated wastewater and CO_2 capture, respectively. They include the cost of transportation of the treated water or CO_2 by means of pipelines, CO_2 compression in addition to the operational cost of the treatment or the capture.

They can be defined using the following equations:

$c_i^w = (c_i^{wT} d_i + c_i^{wO}) Q_w$ (7)

$c_i^{CC} = (c_i^{CCT} d_i + c_i^{CCO}) Q_{CC}$ (8)

where c_i^{wT} and c_i^{CCT} are the unit costs of transporting water and CO_2. d_i is the distance between the greenhouse and the wastewater treatment plant, which is also to location of

the captured CO_2. c_i^{wo} and c_i^{cco} are the unit environmental impacts associated with transporting and generating water and CO_2. As for Q_{cc} is the total quantity of CO_2 to be injected in the greenhouse.

2.1. Illustrative example and available data

Increasing water scarcity levels, limited land available for agricultural activities and harsh climatic conditions have rendered the satisfaction of local demands for food and water a challenging target. To overcome this problem, the State of Qatar has shifted the focus to some alternative water and food systems that can meet the local need while preserving the environment. In fact, greenhouses were adopted to support the local food production by providing an adequate environment for crops cultivation. As part of the strategies deployed to improve the efficiency of greenhouses and enhance their yields, CO_2 enrichment represents a sustainable technique to increase the productivity while offsetting the emissions generated from energy intensive technologies such as wastewater treatment. The reuse of wastewater also provides an opportunity to lift the burden on groundwater and represents a cleaner alternative in terms of the engendered emissions in comparison with desalination. In this study, the potential of utilising treated wastewater and CO_2 from two water treatment plants in a greenhouse producing tomatoes. The example is formulated as a planning problem where the allocation of resources from the sources (wastewater treatment plants (WWTP) to the sink (greenhouse) is optimised following the formulation presented in section 2.1. Tables 1 and 2 presents the input data used in the model.

Table 1. Water and carbon requirements of the greenhouse and the distance to the different sources.

Sink	**Water Requirements**	**CO_2 Requirements**	**Distances from**	
			Doha North WWTP	Shahaniyah
	Q_w (m^3/year)	Q_{cc} (kg/year)	d_i (km)	
Greenhouse	1,664	14,300	23	8

Table 2. Economic and environmental costs input data.

		Operations (production or treatment)	**Transportation and CO2 compression**
Unit economic cost	Water treatment	0.11 \$/ m^3	0.000671 \$/ m^3/km
	CO_2 Capture	0.000241 \$/kg	0.0000172 \$/ kg of CO_2/km
Unit Environmental cost	Water treatment	0.49 kg of CO2eq/ m3	0.00162 kg of CO_{2eq}/ m^3/km
	CO2 Capture	0 kg of CO_{2eq}/ kg	0.036 kg of CO_{2eq}/ kg of CO_2/km

3. Results

Results of the multi objective optimisation are summarised in Figure's 1 and 2. The optimal solution suggests that 60% of the amount of water supplied to the greenhouse can be sourced from Doha North treatment plant, while the remaining 40% can be supplied by Shahaniyah. As for the CO_2 requirements, Shahaniyah contributes with 96% of total

needed quantity while only 4% is taken from Doha North. Considering the costs and savings associated with the optimal solution, the Pareto Front demonstrates that an amount of 2,629 m^3/year could be achieved in terms of greywater footprint. As for the economic and environmental costs, the optimal distribution requires 54,3579 $/year and generates around 20,043 kg of CO_{2eq}/year. As for the nutrient's intake, the wastewater used for irrigation supplies 20 kg/year of Nitrogen, which can substitute 15% of the nitrogen supplied from industrial fertilisers that are produced using energy-intensive technologies and processes.

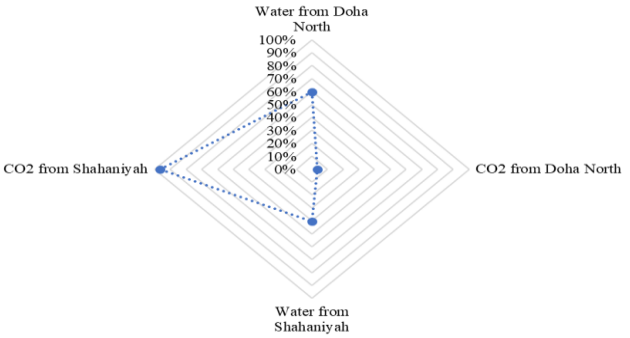


Figure 1: The optimal contribution of each resource source to the sink.

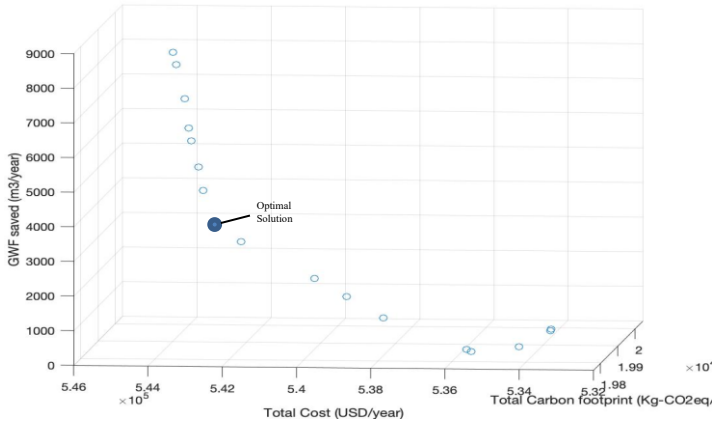


Figure 2: The Pareto Front.

4. Conclusion

Shifting to sustainable agricultural practices is essential in the light of the increasing pressures on the natural resource base and the environment. Adopting alternative water sources and deploying efficient cultivation techniques is deemed beneficial in alleviating the environmental burden that energy-intensive systems are inflicting. In this paper, the potential of using treated wastewater and CO_2 carbon sequestration in a greenhouse is investigated through a multi-objective optimisation model aiming to minimise environmental impacts and economic costs. Results of the study assert that requirements

of the greenhouse can be fulfilled from two wastewater treatment plants that can deliver the necessary water, CO_2, and nitrogen. The optimal solution implies an investment of 54,3579 $/year and produces 20,043 kg of CO_{2eq}/year with a significant reduction of greywater footprint amounting to 2,629 m^3/year. In addition, the use of wastewater instead of the conventional aquifer or desalinated water offsets almost 15% of the nitrogen required by the plant, reducing the dependency on industrial fertilisers.

5. References

C. García-Delgado; E. Eymar; J.I. Contreras; M.L. Segura, 2012, Effects of fertigation with purified urban wastewater on soil and pepper plant (Capsicum annuum L.) production, fruit quality and pollutant contents. Spanish J. Agric. Res., 10, 209.

I. Ghiat; R. Govindan; S. Namany; T.Al-Ansari, 2020, Network optimization model for a sustainable supply network for greenhouses, Computer Aided Chemical Engineering 48, 1885-1890.

I. Ghiat; H.R. Mackey; T. Al-Ansari, 2021, A Review of Evapotranspiration Measurement Models, Techniques and Methods for Open and Closed Agricultural Field Applications. Water, 13, 2523.

I. Ghiat; F. Mahmood; R. Govindan; T. Al-Ansari, 2021, CO2 utilisation in agricultural greenhouses: A novel 'plant to plant' approach driven by bioenergy with carbon capture systems within the energy, water and food Nexus. Energy Convers. Manag., 228, 113668.

H.C.J. Godfray; J.R. Beddington;I.R. Crute; L. Haddad;D. Lawrence; J.F. Muir; J. Pretty; S.Robinson;S.M. Thomas; C. Toulmin, 2010, Food Security: The Challenge of Feeding 9 Billion People. Science (80) 327, 812–818.

F.-Z. Lahlou; H.R. Mackey; T. Al-Ansari, 2021, Wastewater reuse for livestock feed irrigation as a sustainable practice: A socio-environmental-economic review. J. Clean. Prod., 294, 126331.

F. zahra. Lahlou; S. Namany; H.R. Mackey; T. Al-Ansari, 2020, Treated Industrial Wastewater as a Water and Nutrients Source for Tomatoes Cultivation: an Optimisation Approach. Comput. Aided Chem. Eng., 48, 1819–1824.

W. Musazura; A.O. Odindo; E.H. Tesfamariam; J.C. Hughes; C.A. Buckley, 2019, Nitrogen and phosphorus dynamics in plants and soil fertigated with decentralised wastewater treatment effluent. Agric. Water Manag. , 215, 55–62.

E.M. Nederhoff, Effects of CO2 concentration on photosynthesis, transpiration and production of greenhouse fruit vegetable crops.

R. Govindan; T. Al-Ansari, 2019, Simulation-based reinforcement learning for delivery fleet optimisation in CO2 fertilisation networks to enhance food production systems. In Proceedings of the Proceedings of the 29th European Symposium on Computer Aided Process Engineering.

J. Raju; G. Byju, 2019 Quantitative determination of NPK uptake requirements of taro (Colocasia esculenta (L.) Schott). J. Plant Nutr., 42, 203–217

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50334-1

Exergoeconomic optimization of a double effect evaporation process of coffee extract

Tinoco-Caicedo D.L.,[a,b,*] Calle-Murillo J.,[a] Feijoó-Villa E.[a]

[a] *Facultad de Ciencias Naturales y Matemáticas, Escuela Superior Politécnica del Litoral, 090903 Guayaquil, Ecuador*
[b] *Department of Process Engineering, Universidad de las Palmas de Gran Canaria, 35017 Gran Canaria, Spain*
**dtinoco@espol.edu.ec*

Abstract

The exergoeconomic optimization performed on different industrial processes is used to reduce investment costs, operational costs and the exergy destruction rate in order to increase at the same time the rentability and sustainability of a factory. This study focuses on exergoeconomic optimization of a double effect evaporation process of coffee extract in a factory located in Ecuador. The specific product cost was minimized, and the exergetic efficiency was maximized by a parametric study with an iterative methodology and the integration of a sub-optimized steam recompression system. The parametric study showed that the exergetic efficiency could be increased by 5% and the exergy destruction cost cost rate reduced by $1143/h by changing the concentration of the feed extract, the pressure of the 1st effect and the outlet pressure in the expansion valve to 26 w/w%, 35 kPa and 300 kPa, respectively. However, by integrating the mono-objective optimized steam recompression system in the process, the exergy destruction cost rate could be reduced up to 73%, the exergetic efficiency could be increased by 13% and the specific product cost could be reduced by $225/t.

Keywords: Exergoeconomic Optimization; Double Effect Evaporation; Exergetic efficiency

1. Introduction

The economic growth in the world requires an increase of the energy consumption. The World Council of Energy estimated that by 2030 there will be an increase of approximately 35% in the energy consumption (Tvaronavičienė & Ślusarczyk, 2019). The food factories, such as the plants that produces instant coffee, have one of the highest consumptions of energy (around 60.2 MJ/kg of product) using as energy source fossil fuels (95%), and electricity (5%) (Maroulis & Saravacos, 2007). An exergetic optimization is necessary to improve the sustainability of these industrial processes, by the reduction of the exergy destruction foot print (Romero & Linares, 2014). Also exergoeconomic optimization could be used as a tool to simultaneously improve the thermodynamic and economic performance of a system (Abusoglu & Kanoglu, 2009).

Many methods for the optimization of process have been developed, but they possess certain strengths and weaknesses. The optimization by GA (genetic algorithm) allows for a rapid search of solutions, however they could be lost in a sub-optimal solution (Ding et al., 2020). Utilizing a parametric study results in a better solution when the studied system is complex and modeled by a high resolution simulator or neural networks (Alirahmi & Assareh, 2020). The exergoeconomic optimization has mainly developed for power plants

(Alharbi et al., 2020), trigeneration systems (Gholizadeh et al., 2020), fuel cells (Feili et al., 2020) and similar systems, but for food factories few studies exist

This study proposes a re-engineering of a double effect evaporation process (DEEP) of coffee extract, by integrating steam recompression system. The aim of the present work is minimize the specific product cost by a mono-objective optimization of the recompressor in the evaporation process. Also a parametric study is done to the process in order to balance the investment cost and the exergy destruction cost rates of the system. This study is based on the results of an advanced exergoeconomic analysis of the DEEP of coffee extract presented in a previous study by the authors (Tinoco-Caicedo et al., 2021).

2. Materials and Methods

2.1. System Description

The DEEP of coffee extract integrated to a steam recompressor is presented on Figure 1. The coffee extract (stream 1) is an aqueous solution which is heated in the E-102 with steam (stream S1) generated by a boiler (B-101). The heated coffee extract (stream 3) is pumped to enter to a double effect evaporation system (D-101 and D-102) in order to increase the concentration of soluble solids by using steam as the heating medium. The concentrated coffee extract (stream 9) is cooled with chilled water (stream W3) to prevent development of microorganism, and then it (stream 10) is sent to a drying process in order to produce instant coffee powder. The evaporate water obtained in the evaporators (stream 11) is sent to a mechanical steam recompression (C-101) in order to increase its pressure (stream 12) so it can be reused for heat exchange. For the process simulation, the SRK thermodynamic model is used for high pressure steam and the ideal model for low pressure steam and flue gas because they are at low pressure. The liquid phase is assumed ideal because it is a dilute solution.

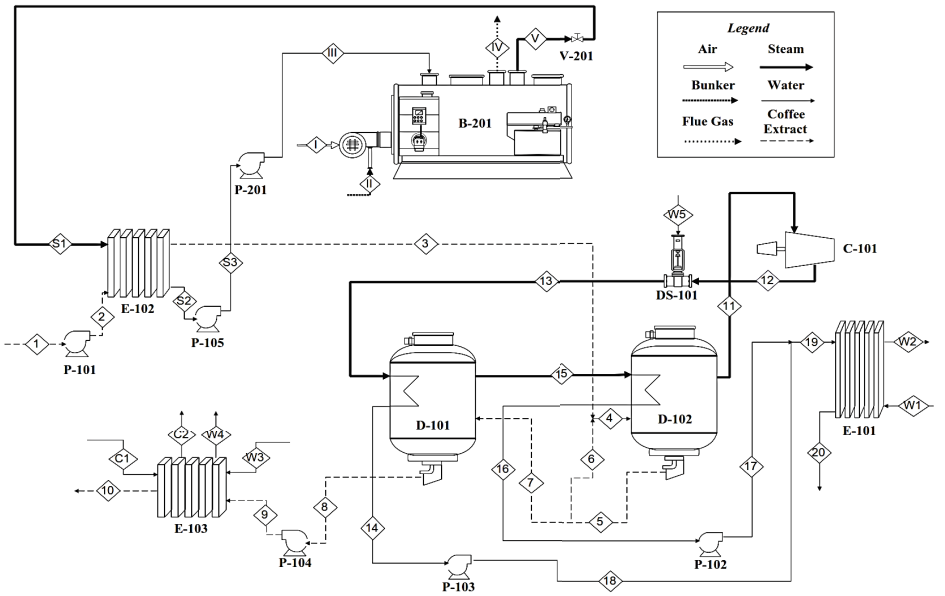


Figure 1. Process flow diagram of the double effect evaporation of coffee extract with a mechanical steam recompression.

2.2. Exergoeconomic Optimization

A mono-objective optimization was performed to the recompressor in order to minimize the specific product cost. The mathematical formulation of the optimization problem is shown in Eq (1)-(4) and based in the literature (Bejan et al., 1996).

$$Z_k^{CI} = \beta(TCI_k) \quad (1) \qquad \min_{\varepsilon_k} \dot{C}_{P,k} = \dot{C}_{F,k} + \dot{Z}_k^{CI} + \dot{Z}_k^{OM} \quad (2)$$

$$TCI_k = B_k \left(\frac{\varepsilon_k}{1-\varepsilon_k}\right)^{n_k} \dot{E}x_{p,k}^{m_k} \quad (3) \qquad Z_k^{OM} = \gamma_k(TCI_k) + \omega_k \tau \dot{E}x_{p,k} + R_k \quad (4)$$

Where the subscript k is the k-component, CI the Capital Investment, TCI the Total Capital Investment, β the Capital Recovery Factor, τ is the average annual plant operation hours, $c_{F,k}$ the specific fuel cost which is the electrical energy price, γ_k is a coefficient that takes into account the Operational and Maintenance Cost (OM) and is assumed as 1.06. The ω_k and R_k constants have not be taken into consideration because it is assumed that the variability in OM are negligible, as suggested by Liu & He, 2020. For determining the non-linear regression constants B_k, n_k, and m_k, the TCI was fitted as a function of the exergetic efficiency (ε_k) and product exergy ($\dot{E}x_{P,k}$) by using the software Wolfram Mathematica®. The isentropic efficiencies for the compressor were varied between 50% to 90%. With this assumption the Eqs (5)-(7) are used in order to find the optimal exergetic efficiency (ε_k^{OPT}), and exergoeconomic factor (f_k^{OPT}), for the compressor C-101.

$$\varepsilon_k^{OPT} = \frac{1}{1+F_k} \quad (5) \qquad F_k = \left(\frac{(\beta+\gamma_k)B_k n_k}{\tau c_{F,k} \dot{E}x_{P,k}^{(1-m_k)}}\right)^{\frac{1}{n_k+1}} \quad (6) \qquad f_k^{OPT} = \frac{1}{1+n_k} \quad (6)$$

Furthermore an iterative methodology proposed by (Hamdy et al., 2019) was followed for the exergoeconomic optimization, the routine is presented on Figure 2. For the parametric study, the operational parameters were selected in function of the most critical components presented in the previous exergoeconomic analysis of the base case (Tinoco-Caicedo et al., 2021). In order to balance the overall exergetic destruction cost rate ($\dot{C}_{D,k}$) and investment cost rate ($\dot{Z}_k$) of the process, a value of 50% was considered as the optimum exergoeconomic factor (f_k).

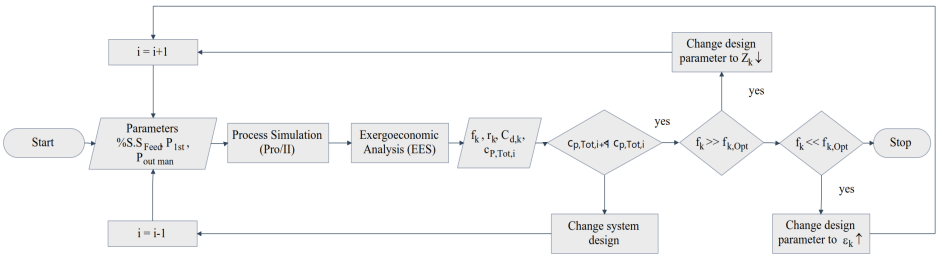


Figure 2. Algorithm used for the exergoeconomic optimization

3. Results and Discussion

The parametric study was performed first by analysing the effect of some important operational parameters presented on Table 1. These parameters were selected because they affect significantly the components which have the highest exergy destruction cost rate ($\dot{C}_{D,k}$) and investment cost rates ($\dot{Z}_k$).

Table 1. Recommended changes for the main components of the process.

Component	f_k(%)	ε_k(%)	Goal	$S.S_{Feed}$	P_{1st}	$P_{out\ man}$
D-101	85.8%	52%	$\dot{Z}_k \downarrow$	↑	↓	↑
D-102	54.8%	60%	$\dot{Z}_k \downarrow$	↑	↑	-
E-101	2.0%	19%	$\dot{C}_{D,k} \downarrow$	↑	-	-

The table shows some recommended changes in order to adjust the exergoeconomic factor to 50%. In order to achieve a reduction in the investment cost rate of the evaporators (D-101 y D-102), and the exergy destruction cost rate in the heat exchanger E-101, an increment in the concentration of soluble solids ($S.S_{Feed}$) in the feed extract is required.

On Figure 3. a) it is shown that an increment of the concentration of the feed extract to 26% allows the reduction of the overall exergy destruction cost rate of the process.

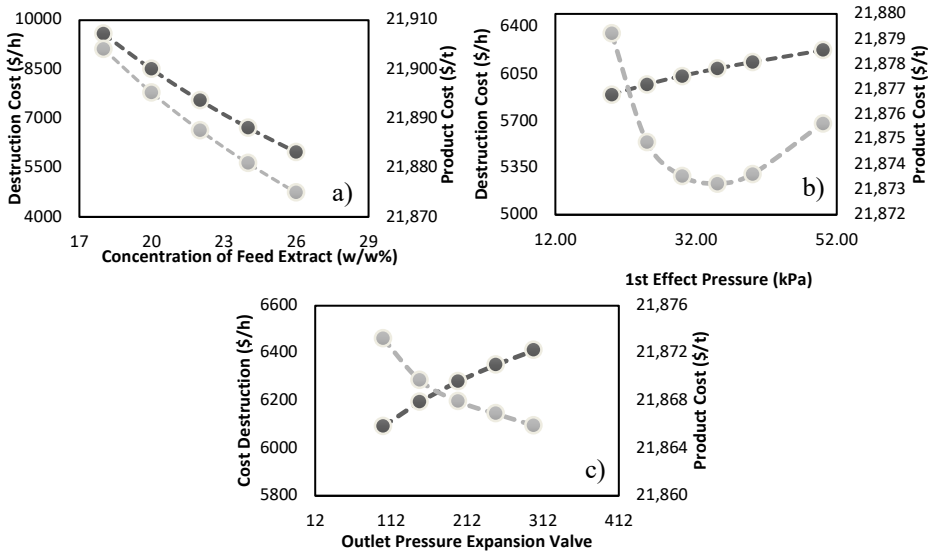


Figure 3. Effect of a) the initial concentration of soluble solids in coffee extract; b) the 1st effect pressure; c) the outlet pressure expansion valve; on the overall exergy destruction cost rate (•) and the product cost (•).

Furthermore, on Figure 3.b) it is shown that a reduction in the 1st effect pressure (P_{1st}) causes a reduction of the exergy destruction cost rate and a negligible increase of the product cost. An intermediate pressure of 33 kPa could minimize the product cost.

Additionally, an increment of the outlet pressure in the expansion valve ($P_{out\ man}$) causes a negligible reduction of the specific product cost as is shown in Figure 5. This is caused because the purchased equipment cost of the first effect evaporator is slightly reduced with the increment of pressure of the steam used for the heat exchanging.

For the mono-objective optimization of the recompression system, the constants for the non-linear regression model (B_k:275.834, n_k:0.85488 and m_k:0.9564) were estimated, with a R^2 of 0.9972. The optimum values of f_k y ε_k were 53,9% and 64% respectively. This result could be obtained when the compressor C-101 has an isentropic efficiency of 65%.

The results of the parametric study and the structural change are summarized on Table 2. It is shown that the steam recompression (structural change) allows for a more significant reduction of the specific product cost and an increase of 10% in the exergetic efficiency. This is caused because most of the exergy destruction rate is due to the condensation of the steam after the evaporation process. When this steam is recompressed and reused in the process, the exergy destruction rate is significantly reduced and therefore the specific product cost is also reduced.

Table 2. Results of the exergoeconomic optimization

Parameter	Units	Base Case	1st	2nd	3rd	Optimized Case
			Parametric Study			Structural Change
$S.S_{Feed}$	%	22	26	26	26	26
P_{1st}	kPa	25	25	35	35	35
$P_{out\ man}$	kPa	101.3	101.3	101.3	300	300
$\dot{C}_{D,Tot}$	$/h	7555	5973	6092	6412	2034
$c_{P,Tot}$	$/t	21888	21875	21873	21866	21663
$\dot{Z}_{Tot}$	$/h	199	182	179	163	203
f_{Tot}	%	2.57	2.96	2.85	2.48	9.09
ε_{Tot}	%	80	84	84	83	93

4. Conclusions

The purpose of this study was to perform an exergoeconomic optimization of the recompressor of the DEEP of coffee extract to minimize de product cost and a parametric study to balance the investment costs and the exergy destruction cost rate. The results of this research show that the optimized recompressor has an isentropic efficiency of 65% and an exergetic efficiency of 64%. Furthermore the structural change had a more significant effect because the specific product cost was reduced by $225/t and the exergy destruction cost rate was reduced by 73%, while the exergetic efficiency increase was 13%. The parametric study allows to reduce the exergy cost rate and increase the exergetic efficiency by 15% and 5% respectively, while the specific product cost was not affected significantly. The initial concentration of soluble solids proved to be a significant parameter for the process, given that an 8 w/w% increase of the initial concentration of soluble solids reduced the avoidable exergy destruction cost by 15%. Finally, the results suggest that a reduction of the exergy destruction rate in the system can be achieved and

it is possible to have annual savings of $\$8.37x10^5$ in the overall operating costs. Further research is needed to optimize the solid-liquid extraction of coffee, in order to achieve a higher initial concentration of soluble solids in the coffee extract. Also, the recompression system has to be proved in a pilot scale in order to confirm the results of the simulation.

References

Abusoglu, A., & Kanoglu, M. (2009). Exergoeconomic analysis and optimization of combined heat and power production: A review. *Renewable and Sustainable Energy Reviews*, *13*(9), 2295–2308.

Alharbi, S., Elsayed, M. L., & Chow, L. C. (2020). Exergoeconomic analysis and optimization of an integrated system of supercritical CO2 Brayton cycle and multi-effect desalination. *Energy*, *197*, 117225.

Alirahmi, S. M., & Assareh, E. (2020). Energy, exergy, and exergoeconomics (3E) analysis and multi-objective optimization of a multi-generation energy system for day and night time power generation - Case study: Dezful city. *International Journal of Hydrogen Energy*.

Bejan, A., Tsatsaronis, G., & Moran, M. (1996). *Thermal Design and Optimization*. Wiley.

Ding, P., Yuan, Z., Shen, H., Qi, H., Yuan, Y., Wang, X., Jia, S., Xiao, Y., & Sobhani, B. (2020). Exergoeconomic analysis and optimization of a hybrid Kalina and humidification-dehumidification system for waste heat recovery of low-temperature Diesel engine. *Desalination*.

Feili, M., Ghaebi, H., Parikhani, T., & Rostamzadeh, H. (2020). Exergoeconomic analysis and optimization of a new combined power and freshwater system driven by waste heat of a marine diesel engine. *Thermal Science and Engineering Progress*, *18*, 100513.

Gholizadeh, T., Vajdi, M., & Rostamzadeh, H. (2020). Exergoeconomic optimization of a new trigeneration system driven by biogas for power, cooling, and freshwater production. *Energy Conversion and Management*, *205*(August 2019), 112417.

Hamdy, S., Morosuk, T., & Tsatsaronis, G. (2019). Exergoeconomic optimization of an adiabatic cryogenics-based energy storage system. *Energy*, *183*, 812–824.

Liu, Z., & He, T. (2020). Exergoeconomic analysis and optimization of a Gas Turbine-Modular Helium Reactor with new organic Rankine cycle for efficient design and operation. *Energy Conversion and Management*, *204*(November), 112311.

Maroulis, Z. B., & Saravacos, G. D. (2007). Food Plant Economics. In *Food Plant Economics*.

Romero, J. C., & Linares, P. (2014). Exergy as a global energy sustainability indicator. A review of the state of the art. *Renewable and Sustainable Energy Reviews*, *33*, 427–442.

Tinoco-Caicedo, Feijoó-Villa, E., Calle-Murillo, J., Lozano-Medina, A., & Blanco-Marigorta, A. M. (2021). Advanced exergoeconomic analysis of a double effect evaporation process in an instant coffee plant. In *Computer Aided Chemical Engineering*.

Tvaronavičienė, M., & Ślusarczyk, B. (2019). *Energy Transformation Towards Sustainability*. Elsevier Ltd.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50335-3

Ecohydrological modeling and dynamic optimization for water management in an integrated aquatic and agricultural livestock system

A. Siniscalchi [a,b], R. J. Lara [c], M. S. Diaz[a,b,*]

[a]*Planta Piloto de Ingeniería Química (PLAPIQUI CONICET-UNS), Camino La Carrindanga km. 7, Bahía Blanca, Argentina*
[b]*Departamento de Ingeniería Química, Universidad Nacional del Sur (UNS), Bahía Blanca, Argentina*
[c]*Instituto Argentino de Oceanografía (IADO-CONICET), Camino La Carrrindanga 7 Km7, Bahía Blanca, Argentina*
sdiaz@plapiqui.edu.ar

Abstract

In this work, we address the water-food-energy nexus by formulating an optimal control problem to mitigate extreme events, such as droughts and floods, by taking into account aquatic, agricultural livestock systems with associated carbon capture, as well as the required energy for irrigation.

The resulting dynamic optimization problem is subject to a differential algebraic equation system with four modules that include ecohydrological, agricultural and livestock models, as well as the associated carbon dioxide capture. Energy is calculated to pump water in aqueducts, if necessary, and for drip irrigation. Numerical results show that the integrated management of the aquatic and agricultural livestock system in a semiarid region can effectively mitigate, even avoid, the effects of droughts and floods, while improving the economic incomes in the region.

Keywords: optimal control, ecohydrological model, dynamic optimization, water-food-energy-health nexus.

1. Introduction

According to the UN Food and Agriculture Organization (FAO), water footprint associated to food production is around 70% of global water consumption. Water footprint is an indicator of the amount of water required for a production process and is used as a basis for more efficient management of this valuable resource (FAO, 1998). In arid and semi-arid regions, water is a resource of highly variable availability, where management becomes fundamental to reduce uncertainty in predicting mid- and long-term oscillations and thus supporting sustainable ecosystem services, particularly food safety food production. Droughts cause death of livestock, since animals cannot get access to drinking water, do not fatten and do not produce milk due to high temperatures associated to the lack of water and shade. Further, the uncontrolled variation of hydrochemical properties of water bodies can produce deleterious effects on reproduction and growth of valuable fish species. In Argentina, floods cause economic losses that can represent 1.1% of the Gross Domestic Product (World Bank, 2000), including crops, rural and urban inundation They increase uncertainty in investment strategies of touristic developments related to water bodies. The formation of transient, slow-moving, shallow

water surfaces can promote the explosive proliferation of disease vectors such as mosquitoes, snails, parasites, etc. Similarly, migration of wild birds that use wetlands for breeding and resting can exchange viruses with domestic poultry. In the last years, the water-food-energy (WFE) nexus concept has been promoted as a tool for achieving sustainable development accounting the relationship among these three resources. From a nexus perspective, integrated aquatic and agricultural livestock systems approaches, can help to solve some water-food-energy issues. Therefore, to achieve sustainability of hydrographic basins, interactions among the WFE fluxes and the ecosystems fluxes. For this reason, mathematical models that allow simulation and optimization representing environmental extreme events, such as droughts and floods, become fundamental tools for decision-making in the socio-economic development of a region (Siniscalchi et al, 2018; 2019).

In this work, we propose the formulation of an optimal control problem to address mitigation of extreme environmental events such as floods and droughts. The preservation of a valuable fish species within a salt lake is also addressed, as well the use of water as a resource in a productive livestock-agricultural system located in a semi-arid region. The optimization problem is constrained with a system of differential and algebraic equations representing ecohydrological and agricultural livestock models of a salt lake and its basin. Within this framework, four main objectives are addressed: a) to prevent flooding of a nearby village and its touristic areas during a wet period by diverting part of the flow from a Chasicó Lake tributary into a constructed reservoir (the diversion flowrate is a control variable); b) to optimize management of the constructed reservoir to keep salinity in the lake within a desired value for fish species (silverside) during drought periods; c) to include restoration strategies for native species that comprise a xerophilic woodland currently existing within the salt lake basin, combining new plantations of Prosopis sp, with drought resistant crops (*Chenopodium quinoa*) and pasture (*Eragrostis curvula*) ,irrigated with freshwater taken from the proposed constructed artificial reservoir and d) to provide drinking water and shade to cattle. For the last three objectives, the outlet freshwater flowrate from the constructed reservoir is a control variable. Numerical results show that if water is accumulated in an artificial reservoir during wet periods (six-year period, with average annual precipitations of 650 mm), a subsequent ten-year drought period (average annual precipitations 250 mm) can be mitigated, while maintaining the salinity of Chasicó Lake for the conservation of silverside fishing. In this way, during the dry period, quinoa and pasture can be sown and Prosopis sp. can provide shade and fodder to cattle and long-term ecosystem benefits. The proposed dynamic optimization model has proven to be a powerful tool for water management in an integral way.

2. Methodology

We propose an optimal control problem for medium term planning of management strategies in integrated aquatic, agricultural livestock systems in semiarid regions under extreme events, such as droughts and floods. The main concept is to address this problem under the water-food-energy (WFE) nexus. The dynamic optimization problem is subject to a differential algebraic equations system that represents the integrated model. The objective function is an integral one that aims to keep a salt lake volume (and its associated salinity, as it is an endorheic basin), at a desired value, to avoid flooding of the nearby village and fields and to keep salinity at optimal values for reproduction of a valuable fish species, silverside (*Odontheses bonariensis*), which has been exported to several countries, as Japan. There are two control variables: the diverted stream flow to a

nearby constructed reservoir to avoid flooding while keeping salinity at 23 kg/m^3 in a wet scenario and the stream flow that is fed to the salt lake from the constructed reservoir to keep lake salinity at the desired value in dry periods $Q_{Res}(t)$. When solving the dynamic optimization problem, the definition of a "wet" or "dry" period is not required, since this is determined by the input profiles provided for precipitations, temperature, wind and current conditions at daily time intervals.

The agroecohydrological model has four integrated submodels:

a) Ecohydrological model. It includes dynamic mass balances for water and salt within a salt lake and an artificial fresh water reservoir that is constructed to derive water during wet periods and to provide water for irrigation and to keep lake salinity within low values in dry periods.

$$\frac{dm}{dt} = \left[Q_{pp}(t)\left(\frac{V(t)}{h(t)}\right) + Q_{river}(t) + Q_{gw}(t) - \text{Evap}\,(t)\left(\frac{V(t)}{h(t)}\right)\right]\delta_w/1000 \quad (1)$$

here m is the total water mass in the lake (kg); δ_w, is water density (kg.m^{-3}), which is assumed constant; V corresponds to salt lake volume (m^3) and h is average depth (m). Q_{pp} (L.day^{-1}.m^{-2}) corresponds to precipitations, Q_{gw} (L.day^{-1}) is groundwater flowrate. Q_{river} L.day^{-1}) is the tributary discharge to Chasicó Lake, which is calculated as

$$Q_{river}(t) = Q_{cr}(t) - Q_{divert}(t) + Q_{res}(t) \quad (2)$$

where Q_{divert} (L.day-1) is the stream flowrate that could be diverted when necessary from Chasicó River (Q_{cr}) (L.day-1) to the reservoir during wet periods, to keep Chasicó Lake salinity within desired values (it is a control variable). Q_{res} (L.day-1) is the daily water amount that could be diverted from the reservoir to Chasicó River and, subsequently to the salt lake for salinity and volume control (this is also a control variable) in drought periods, if required. Evaporation per unit area ($Evap$, L.day-1.m-2) is calculated taking into account energy and momentum balances (Penman, 1948; Siniscalchi et al., 2018a). As salt concentration in both groundwater and the tributary is negligible, we assume that salt mass is constant within the salt lake and salt concentration (Cs) is calculated as:

$$\frac{dCs}{dt} = -\frac{Cs(t)}{V(t)}\left[Q_{pp}(t)\left(\frac{V(t)}{h(t)}\right) + Q_{river}(t) + Q_{gw}(t) - Evap(t)\left(\frac{V(t)}{h(t)}\right)\right]\delta_w/(\delta_w 1000) \quad (3)$$

The water mass balance in the artificial reservoir is as follows

$$\frac{dVres}{dt} = Q_{divert}(t) + Q_{well}(t) + Q_{pp}(t)A(t)\delta_{H_2O} - Evap(t)A(t)\delta_{H_2O} - Q_{res}(t) - \sum_j Q_j(t) \quad , \quad j = \textit{Prosopis sp,}\ \text{quinoa,}\ \textit{E. curvula.,}\ \text{cattle} \quad (4)$$

Where $Q_{well}(t)$ corresponds to a stream from a flowing well; $Q_j(t)$ is the daily water requirement for *Prosopis sp* (once seedlings are transplanted to field), quinoa, *Eragrostis curvula* irrigation and water requirement for cattle, respectively. Water requirement for trees and crops is calculated as a function of crop evapotranspiration $ETq(t)$ and precipitations.

b) Agricultural model. It includes grow models and evapotranspiration calculations for *Prosopis sp.*, quinoa (a high value crop), and a native pasture species, *Eragrostis curvula*. Biomass for Prosopis is calculated as follows.

$$BP = e^{(-2.14+2.530*ln\,(DBH))} \quad (5)$$

$$DBH = 6.1\ 10^{-4}(t) + 2.43 \quad (6)$$

where DBH stands for diameter breast height of Prosopis stem
Actual evapotranspiration is $ETi(t)$ is calculated as:

$$ETi(t) = Ki(t) * ET0(t) \quad \textit{,i=Prosopis,} \text{ quinoa, } \textit{E. curvula.} \quad (7)$$

where *Ki(t)* correspond to the cultivation "coefficient" for each species, which is represented with Fourier series, which were adjusted based on data on their water requirement through the different growth stages in a semiarid region, based on data from from the literature. *ET*0 is calculated as function of air temperature, solar radiation, wind, vapor pressure, etc (Siniscalchi et al., 2019).
c) Livestock model. It includes water requirement calculation for both breeding (270kg) and fattening (450kg) cattle, which is calculated at each time interval, as follows:

$$Q_{cattle270}(t) = [16.765 * e^{0.0295*T_{air}}] * N_1 \quad (8)$$

$$Q_{cattle455}(t) = [27.64 * e^{0.0288*T_{air}}] * N_2 \quad (9)$$

where N_1 and N_2 is the number of cows from each size. Correlations are based on data collected from the semiarid region that constitutes our case study.

d) Carbon capture model. Carbon capture (kg C/d) is calculated as function of tree and crop biomass as follows,

$$CC_{Eragrostis}(t) = 0.3 * Biomass_{Ec} \quad (10)$$

$$CC_{Prosopis}(t) = 0.5 * Biomass_{Psp} \quad (11)$$

Finally, energy calculations are carried out for pumps used in drip irrigation for crop and pasture.

3. Numerical results

In this work, the case study is Chasicó Lake and its endorheic basin, located in a semiarid region in Argentina. It is a salt lake. The optimal control problem described in Section 2 has been implemented and solved with a control vector parameterization methodology within gPROMS (Siemens Process Systems Engineering, 2020), for a time horizon of ten years of drought, considering an initial lake salinity of 22.9 kg/m^3 and an initial reservoir volume of 4.48E8 m^3. Freshwater flowrate from a 700 m depth flowing well is 348 m^3/h ($Q_{well}(t)$). Precipitation, temperature, wind profiles are represented by Fourier series whose parameters have been estimated based on historical data for dry periods (Siniscalchi et al., 2019). Average total annual precipitations are 256 mm and average calculated evaporation in both the salt lake (6000 ha surface area) and the artificial reservoir (2800 ha surface area) is 1104 and 770 mm, respectively. The objective is to keep the salt lake volume within 4.8E8 m^3 or, equivalently, lake salinity within 23 kg/m^3, which is optimal for silverside fish reproduction. The dynamic optimization problem has 13 differential equations and 42 algebraic ones. Total CPU time was 7168 s, considering time intervals of 15 days within a time horizon of 10 years.

Figure 1 shows salinity profiles without management; i.e., simulation results. It also shows profiles obtained with the proposed management strategy; i.e., solving the optimal control problem (by deriving freshwater from the artificial reservoir). It can be seen that, without management, salinity values increase to 73 kg/m^3, deeply affecting fish reproduction and survival.

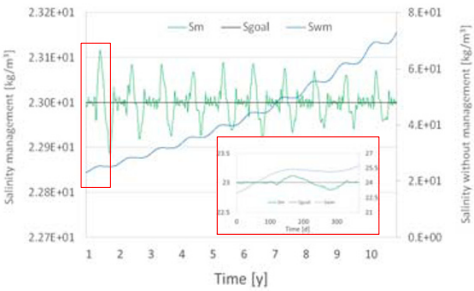


Figure 1: Salinity profiles with (green lines) and without management (blue line)

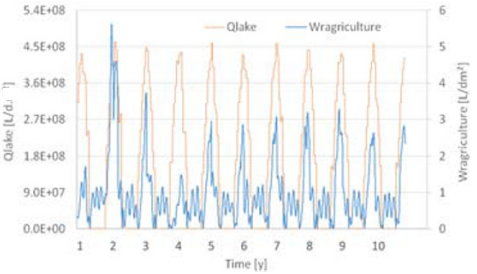


Figure 2: Water requirement for agriculture (blue line) and optimal profile for control variable (Q_{lake}) (orange line)

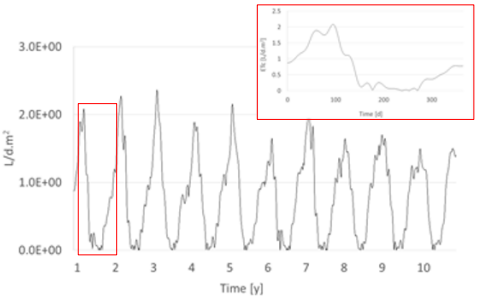


Figure 3: Evapotranspiration profile for *Eragrostis curvula*

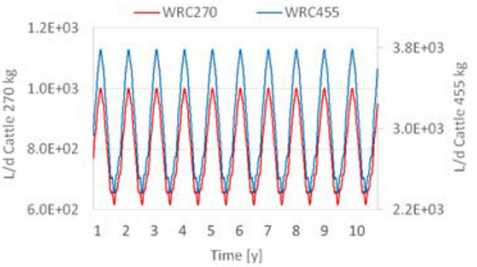


Figure 4: Water requeriment for breeding (red line) and fattening (blue line) cattle

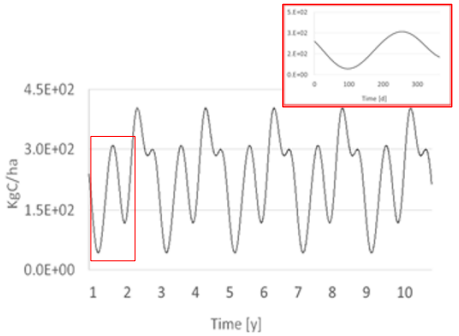


Figure 5: Carbon storage by *E. curvula*

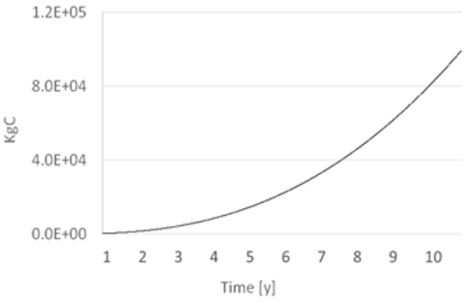


Figure 6: Carbon storage in *Prosopis* wood

When the optimal control problem is solved, salinity is kept around 23 kg/m^3 throughout the 10-year dry period, by deriving freshwater, Q_{lake}, from the artificial reservoir. Q_{lake} profiles (orange lines) are shown in Fig. 2, together with other water stream flows taken from the artificial reservoir and used for trees, crop and pasture irrigation. Figures 3 to 6 show variable profiles in the optimal solution, i.e., obtained by applying the proposed management strategy. Figure 3 shows *Eragrostis c.* evapotranspiration profile, with a detail of the first year. Figure 4 shows water requirement for cattle, which includes 30 breeding and 70 fattening cattle.
Figure 5 shows carbon storage in *Eragrostis c.*, with a detail of the first year profile, while Fig. 6 shows carbon storage in *Prosopis sp.* (a native tree species) wood, during the 10-year period. The function is a sigmoidal one for a 30-year life cycle, but the studied period corresponds to the exponential growth phase. It can be noted that *Eragrostis c.* carbon capture is two orders of magnitude lower, as it is pasture for cattle.

4. Conclusions

In this work we have formulated an integrated model for planning management strategies under extreme weather events in aquatic, agricultural and livestock systems, by formulating an optimal control problem. Dynamic models are formulated for each system, rendering dynamics for water flowrates, salinity within the lake, volume, crop, pasture and trees biomass and carbon capture, among others.

The model has proven to be an efficient tool to plan management within the concept of water-food-energy nexus. It can be demonstrated that the effects of droughts can be effectively mitigated, even avoided, while improving the economic incomes in the region preserving valuable fish population, allowing for the cultivation of high value-added crops (quinoa), pasture and even growing cattle in dry periods.

Current work deals with the inclusion of detailed energy calculation, as well as the study of the effects of severe droughts on local population health.

References

FAO - Food and Agriculture Organization of the United Nations, Allen, R., Pereira, L., Raes, D., Smith, M. Crop evapotranspiration - Guidelines for computing crop water requirements (1998)

Siemens Process Systems Engineering, gPROMS User Manual (2020)

Siniscalchi A.G., Kopprio G., Raniolo L.A., Gomez E.A., Diaz M.S., Lara R.J. Mathematical modelling for ecohydrological management of an endangered endorheic salt lake in the semiarid Pampean region, Argentina. Journal of Hydrology(2018) 563: 778–789

Siniscalchi, A.G., García Prieto, VC., Raniolo, A., Gomez, E., Lara, R.J., Diaz, M.S. Ecosystem Services Valuation and Ecohydrological Management in Salt Lakes with Advanced Dynamic Optimisation Strategies.Computer Aided Chemical Engineering (2019) 46:1579-1584

World Bank. (2000) Argentina - Water Resources Management: Policy Elements for Sustainable Development in the 21st Century, Main Report. Washington, DC. https://openknowledge.worldbank.org/handle/10986/14980.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50336-5

Parametric Analysis of Ortho-to-Para Conversion in Hydrogen Liquefaction

Amjad Riaz[a], Muhammad Abdul Qyyum[b], Arif Hussain[c], Muhammad Islam[a], Hansol Choe[a], Moonyong Lee[a*]

[a]*Process Systems Design & Control Lab (PSDC), School of Chemical Engineering, Yeungnam University, Gyeongsan-si, Gyeongsangbuk-do 38541, Republic of KOREA*
[b]*Department of Petroleum & Chemical Engineering, Sultan Qaboos University, Muscat, OMAN*
[c]*Department of Chemical Engineering, COMSATS University Islamabad, Lahore Campus, 54000, Lahore, PAKISTAN*
mynlee@yu.ac.kr

Abstract

Hydrogen is an energy carrier and is produced just like electricity. Hydrogen is liquefied for storage and transportation purposes to overcome the shortcomings of its low molecular weight and energy density per unit volume. The liquefaction of hydrogen is different from that of other substances as it involves a reactive transformation of its isomers: ortho-hydrogen and para-hydrogen. As the temperature decreases, the equilibrium concentration shifts toward a higher para- content from the normal concentration of 25 % at 25 °C. Para-hydrogen is preferred because of its lower boil-off rate, which is a major challenge at cryogenic temperatures. Ortho-para conversion, heat leak, sloshing, and flashing are considered as the reasons for such losses. The self-conversion rate of hydrogen in a non-equilibrium state is extremely slow; however, at cryogenic temperatures, o-p conversion is an exothermic affair. From the liquefaction point of view, this exothermic heat of conversion is an added work, increasing he liquefaction energy requirement by about 15 %. Catalysts are used to achieve the equilibrium concentration of p-H_2 at a finite rate. Little work has been done from the process systems point of view regarding o-p H_2 conversion. Therefore, parametric analysis of this vital conversion reaction, the spatial distribution of intermediate heat exchangers, and impact on the energy efficiency of the liquefaction process have been studied and partially presented here.
Keywords: Hydrogen Economy; Hydrogen Liquefaction; Ortho-Para Conversion; Hydrogen Energy Network.

1. Introduction

Hydrogen (H_2) is considered a fuel that may revolutionize the future energy mix. It is the only zero-emission, sustainable, and flexible energy carrier with the potential to overcome multifaceted problems and serve as a comprehensive solution to the carbon footprint. High purity, high energy mass density, and high-power density are the hallmarks of hydrogen. Whereas, apart from its production woes, H_2 does face some storage and transportation issues. However, the low molecular weight of H_2, which results in a very low volumetric energy density, impedes its adoption as an energy vector, especially from the perspective of storage and bulk transport (Abdalla et al., 2018; Zheng et al., 2019). H_2 is usually stored and transported as compressed gas or

cryogenic liquid. Cryogenic liquid hydrogen (LH_2) has three to four times higher volumetric energy density than compressed H_2 (International Energy Agency (IEA), 2019). LH_2 is usually preferred over the gaseous form for longer distances or high volumes. With the advent and commercial success of liquified natural gas (LNG), H_2 liquefaction is drawing more interest and attention ever.

LH_2 has a very low boiling point (-253 °C). Commercially, H_2 is liquefied in a three-stage process: precooling (up to -193 °C), cooling (up to -243 °C), and liquefaction (up to -253 °C). Liquid nitrogen is used as a refrigerant in the precooling phase, while a combination of JT valves or expanders and H_2 itself is used in the subsequent stages (Riaz et al., 2021). The specific energy consumption (SEC) is the benchmark used for comparing commercial and conceptual process designs; the ideal value for a feed at 25 bar is approximately 2.7 kWh/kg_{LH2} (Aasadnia & Mehrpooya, 2018; Stolzenburg et al., 2013). However, the commercial plants operate at SEC in the range of 12.5–15.0 kWh/kg_{LH2} (Yin & Ju, 2020).

Moreover, there are two variants of molecular H_2 depending upon the relative orientation of the nuclei of its constituent atoms: ortho-H_2 (nuclei spin in the same direction) and para-H_2 (nuclei spin in the opposite direction). The mixture composed of 75 % ortho-H_2 (o-H_2) and 25 % para-H_2 (p-H_2) at 25 °C is called Normal H_2 (n-H_2). The two isomers are not at the same energy level with o-H_2 being the excited state. The temperature dependent equilibrium between the two isomers shifts towards a more stable form with the decrease in temperature. This shift is called ortho-to-para conversion (OPC), and has a significant contribution to the SEC.

The OPC is critical from the storage and transportation point of view. If n-H_2 is liquefied without OPC, a portion of liquid will boil and trigger mass vaporization and energy losses. This happening is called boil-off and may vaporize 50% of the storage (Sherif et al., 2014) starting from 1% per hour (Baker & Shaner, 1978; Weitzel et al., 1958). The best way to mitigate these boil-off losses is to convert and store as p-H_2 (Ghorbani et al., 2019). As the initial concentration of o-H_2 decreases, there are lesser chances of boil-off losses, as depicted by Figure 1. Therefore, OPC is not neglected commercially, and a p-H_2 concentration >95 % is ensured (Stolten & Scherer, 2013). The best way is to conduct OPC simultaneously with the liquefaction. Therefore, the commercial processes use the approach of converting ortho- to para-hydrogen together with liquefaction.

2. Ortho–to–Para Conversion

The impact of this spin on the properties of matter is so weak that the spin isomers almost remain conserved (Zhang et al., 2021). The fast cooling and instant liquefaction does not alter the molecular composition of H_2 because of the slow OPC. It has been reported that the reaction is second-order (reaction rate = 0.0114 h^{-1}) and takes a long time to establish equilibrium in the absence of a catalyst. This mode of OPC is called self-conversion, which is also governed by temperature.

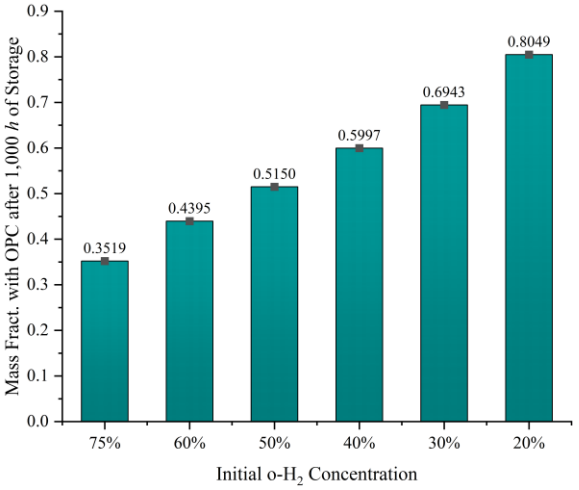


Figure 1 Effect of OPC on the liquid hydrogen boil-off rate as a function of storage time (McCarty et al., 1981)

The introduction of a suitable catalyst enhances the rate of reaction manifolds, magnetic materials, and radiation fields are often employed for the purpose. Generally, catalysts like iron oxide and nickel silica are pakced on the hydrogen side of the exchangers in an arrangement similar to a shell and tube heat exchanger (Jacob H. Stang et al., 2006). The following equation describes the chemical reaction inside the catalytic heat exchanger reactor:

$$o - H_2 \quad \rightarrow \quad p - \mathrm{H}_2 \quad + \quad \Delta H \tag{1}$$

It is said that the catalysis increases the energy requirements by 15 %, but it is the OPC that increases the heat load by around 15–20 % (Stetson et al., 2016; Stolten & Scherer, 2013). The additional heat duty is the exothermic enthalpy of conversion, i.e., 527 kJ/kg, higher than the latent heat of H_2 (~447 kJ/kg), resulting in vaporizing the liquid or heating the already cooled part.

The equilibrium constants of the ortho-para conversion of H_2 in the ideal gas state are independent of pressure. Accordingly, pressure does not appreciably change the ortho-para ratio under equilibrium conditions. Although the lowest rotation levels of the ortho and para varieties differ, ΔE (internal energy change) for the reaction is zero.

3. Research Methods

In recent times, most studies related to H_2 liquefaction are theoretical, i.e., modeling and simulation. The fact that the properties of spin isomers of H_2 differ, especially those related to temperature. Estimation of thermodynamic properties is at the core of process design, either modeling or simulation. Simulation is preferred because of the availability of more reliable components database and precise thermodynamic models. Peng-Robinson is the most widely used model for H_2 liquefaction processes; however, recent studies have preferred the modified version of the original Benedict-Webb-Rubin equation of state. At present, simulating such a reactive exchanger with the catalyst filled in the H_2 side of the exchanger is not possible in commercial simulation software.

To mimic the actual scenario, a reactor is installed right after the exchanger in simulation.

Another important aspect is related to the kinetics of the OPC reaction. Although there are many experimental studies about the kinetics of this conversion, very few are related to kinetic modeling, and even fewer are in the form acceptable to simulation software. Therefore, the most widely used approach is to utilize a conversion reactor. The overall percent conversion is calculated using the equilibrium data reported in the literature (Harkness & Deming, 1932; Scott et al., 1964; Woolley et al., 1948), as presented in Figure 2 by using the following correlation.

$$Conversion\,(\%) = C_o + C_1 \times T + C_2 \times T^2 \quad (2)$$

where C_o, C_1, and C_2 are conversion coefficients, and T is the H_2 temperature in Kelvin.

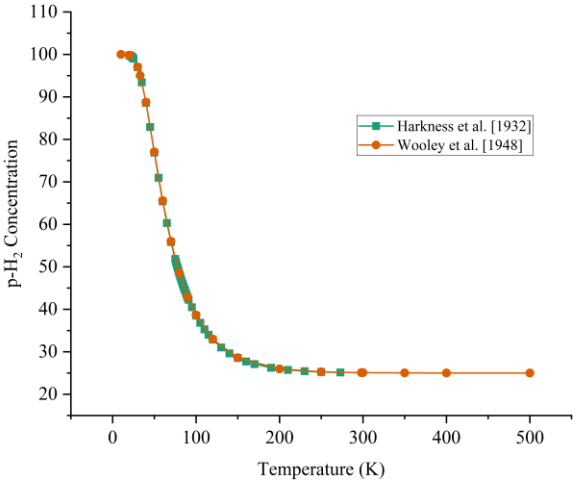


Figure 2 Change in equilibrium p-H_2 concentration with temperature (Harkness & Deming, 1932; Woolley et al., 1948)

4. Further Discussion

Aspen Hysys® has been used in this study which uses Eq. (2) to calculate the percent total conversion. Depending on the process configuration and the number of reactors used, the values reported in the literature vary. The values calculated for the conversion coefficients of each reactor in our recent contribution (Qyyum et al., 2021) are presented in Table 1 as an example. The conversion coefficients were obtained via fitting such that the p-H_2 concnetraion and temperature at each reactor's outlet coincided with the reported data. A similar approach was adopted by Ghorbani et al. (2019), Hammad & Dincer (2018), and Sadaghiani & Mehrpooya (2017) to name a few. The problem with this approach is that the operating conditions, especially temperature, have to be fixed at the reactor outlet; otherwise, the coefficients must be calculated time and again. Also, the heat of conversion is not entirely reflected in the reactor outlet stream leading to errors in the SEC and exergy efficiency calculations. A more appropriate way is to use the equilibrium H_2 data so that the temperature dependency and heat of conversion are accounted for.

Another aspect is the mode of operation for these reactors. The commercial processes use liquid nitrogen/hydrogen baths to carry out the OPC in an isothermal environment, while adiabatic reactors are also used in between. The current process schemes do not

consider this very important factor. The conceptual studies' specific energy and exergy calculations err by approximately 20 %, to say the least. The temperature thresholds for each exchanger/reactor shall not remain the same once an adiabatic reactor is considered, whose impact on the overall energy scenario is challenging to comprehend in the simulation environment.

Table 1 Coefficients for percentage conversion to be used in the conversion reactor model in Aspen Hysys® (Qyyum et al., 2021)

Reactor ID	Conversion Reactor Coefficients		
	C_o	C_1	C_2
R-1	66.12	−0.4125	1.168×10^{-3}
R-2	85.35	−0.3325	1.118×10^{-2}

5. Conclusions

The present study has considered a brief analysis of OPC in the broader context of H_2 liquefaction. The limitations of the current simulation approaches have been highlighted, ranging from property estimation techniques to the use of conversion reactors. The mode of operation of these reactors results in incorrect energy consumption estimates. Therefore, a more thorough analytical analysis may pave the way to develop an understanding of the complex H_2 liquefaction process, which not only saves energy but also help solve the process design issues.

Acknowledgment
This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (2021R1A2C1092152); Priority Research Centers Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2014R1A6A1031189); "Human Resources Program in Energy Technology" of the Korea Institute of Energy Technology Evaluation and Planning (KETEP), granted financial resource from the Ministry of Trade, Industry & Energy, Republic of Korea. (*No. 20204010600100*).

References

Aasadnia, M., & Mehrpooya, M. (2018). Large-scale liquid hydrogen production methods and approaches: A review. In *Applied Energy*. https://doi.org/10.1016/j.apenergy.2017.12.033

Abdalla, A. M., Hossain, S., Nisfindy, O. B., Azad, A. T., Dawood, M., & Azad, A. K. (2018). Hydrogen production, storage, transportation and key challenges with applications: A review. In *Energy Conversion and Management* (Vol. 165, pp. 602–627). Elsevier Ltd. https://doi.org/10.1016/j.enconman.2018.03.088

Asadnia, M., & Mehrpooya, M. (2017). A novel hydrogen liquefaction process configuration with combined mixed refrigerant systems. *International Journal of Hydrogen Energy*, *42*(23), 15564–15585. https://doi.org/10.1016/j.ijhydene.2017.04.260

Baker, C. R., & Shaner, R. L. (1978). A study of the efficiency of hydrogen liquefaction. *International Journal of Hydrogen Energy*. https://doi.org/10.1016/0360-3199(78)90037-X

Ghorbani, B., Mehrpooya, M., Aasadnia, M., & Niasar, M. S. (2019). Hydrogen liquefaction process using solar energy and organic Rankine cycle power system. *Journal of Cleaner Production*. https://doi.org/10.1016/j.jclepro.2019.06.227

Hammad, A., & Dincer, I. (2018). Analysis and assessment of an advanced hydrogen liquefaction system. *International Journal of Hydrogen Energy*. https://doi.org/10.1016/j.ijhydene.2017.10.158

Harkness, R. W., & Deming, W. E. (1932). The Equilibrium of Para and Ortho Hydrogen.

Journal of the American Chemical Society, *54*(7), 2850–2852. https://doi.org/10.1021/ja01346a503

International Energy Agency (IEA). (2019). The Future of Hydrogen - Seizing today's opportunities. In *The Future of Hydrogen - Seizing today's opportunities* (Issue June). https://doi.org/10.1787/1e0514c4-en

Jacob H. Stang, Nekså, P., & E. Brendeng. (2006). On the design of an effient hydrogen liquefaction process. *16th World Hydrogen Energy Conference 2006, WHEC 2006*.

McCarty, R. D., Hord, J., & Roder, H. M. (1981). *Selected Properties of Hydrogen (Engineering Design Data), National Bureau of Standards Monograph 168*. https://doi.org/Library of Congress Catalog Card Number: 80-600195

Qyyum, M. A., Riaz, A., Naquash, A., Haider, J., Qadeer, K., Nawaz, A., Lee, H., & Lee, M. (2021). 100% saturated liquid hydrogen production: Mixed-refrigerant cascaded process with two-stage ortho-to-para hydrogen conversion. *Energy Conversion and Management*, *246*, 114659. https://doi.org/10.1016/J.ENCONMAN.2021.114659

Riaz, A., Qyyum, M. A., Min, S., Lee, S., & Lee, M. (2021). Performance improvement potential of harnessing LNG regasification for hydrogen liquefaction process: Energy and exergy perspectives. *Applied Energy*, *301*(April), 117471. https://doi.org/10.1016/j.apenergy.2021.117471

Sadaghiani, M. S., & Mehrpooya, M. (2017). Introducing and energy analysis of a novel cryogenic hydrogen liquefaction process configuration. *International Journal of Hydrogen Energy*. https://doi.org/10.1016/j.ijhydene.2017.01.136

Scott, R. B., Denton, W. H., & Nicholls, C. M. (Eds.). (1964). *Technology and Uses of Liquid Hydrogen*. Pergamon Press.

Sherif, S. A., Goswami, D. Y., Stefanakos, E. K., & Steinfeld, A. (2014). *Handbook of Hydrogen Energy*. Taylor & Francis. https://books.google.co.kr/books?id=jGkLBAAAQBAJ

Stetson, N. T., McWhorter, S., & Ahn, C. C. (2016). Introduction to hydrogen storage. In *Compendium of Hydrogen Energy* (pp. 3–25). Woodhead Publishing. https://doi.org/10.1016/b978-1-78242-362-1.00001-8

Stolten, D., & Scherer, V. (2013). Transition to Renewable Energy Systems. In *Transition to Renewable Energy Systems*. https://doi.org/10.1002/9783527673872

Stolzenburg, K., Berstad, D., Decker, L., Elliott, A., Haberstroh, C., Hatto, C., Klaus, M., Mortimer, N. D., Mubbala, R., Mwabonje, O., Nekså, P., Quack, H., Rix, J. H. R., Seemann, I., & Walnum, H. T. (2013). *Efficient Liquefaction of Hydrogen: Results of the IDEALHY Project*. *November*, 7–9. https://www.idealhy.eu/uploads/documents/IDEALHY_XX_Energie-Symposium_2013_web.pdf

Weitzel, D. H., Loebenstein, W. V., Draper, J. W., & Park, O. E. (1958). Ortho-para catalysis in liquid-hydrogen production. *Journal of Research of the National Bureau of Standards*, *60*(3), 221–227. https://doi.org/10.6028/jres.060.026

Woolley, H. W., Scott, R. B., & Brickwedde, F. G. (1948). Compilation of thermal properties of hydrogen in its various isotopic and ortho-para modifications. *Journal of Research of the National Bureau of Standards*, *41*(5), 379–475. https://doi.org/10.6028/jres.041.037

Yin, L., & Ju, Y. (2020). Review on the design and optimization of hydrogen liquefaction processes. *Frontiers in Energy*, *14*(3), 530–544. https://doi.org/10.1007/s11708-019-0657-4

Zhang, X., Karman, T., Groenenboom, G. C., & Avoird, A. van der. (2021). Para-ortho hydrogen conversion: Solving a 90-year old mystery. *Natural Sciences*, *1*(1), e10002. https://doi.org/10.1002/NTLS.10002

Zheng, J., Chen, L., Wang, J., Xi, X., Zhu, H., Zhou, Y., & Wang, J. (2019). Thermodynamic analysis and comparison of four insulation schemes for liquid hydrogen storage tank. *Energy Conversion and Management*, *186*, 526–534. https://doi.org/10.1016/j.enconman.2019.02.073

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50337-7

Model agnostic framework for analyzing rainwater harvesting system behaviors

Qiao Yan Soh[a], Edward O'Dwyer[a], Salvador Acha[a], Nilay Shah[a]

[a]*Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, London SW7 2AZ, United Kingdom*
qiaoyan.soh13@imperial.ac.uk

Abstract

To evaluate risks and characterise the responses of a rainwater harvesting system under different rainfall types, this paper presents a model agnostic evaluation framework where a k-means clustering approach is supplemented with a statistical Partial Least Squares model. Four response modes were identified for a studied system. Using these response modes, a higher risk of system overflow was found in 4.5% of simulated scenarios with inadequate water supplies found in 48.2% scenarios. The rainfall distribution in time was found to be crucial in determining the response mode of the system, with sporadic high intensity events or consistent, high total volume events allowing the system to operate in a response mode corresponding to lower system stresses, but with reduced provision of rainwater.

Keywords: Rainwater harvesting and detention; Environmental Systems; Modelling, Analysis and Simulation

1. Introduction

Rainwater harvesting (RWH) systems have shown considerable promise as an alternative or addition to existing urban water management systems as a strategy for ensuring long-term sustainability. Acting as a temporary detention system and a secondary water source, an RWH system facilitates achieving long-term sustainability by reducing stresses in centralised wastewater management systems, improving efficiencies in local water reuse, and minimizing the need for potable water supply.

In the current climate of increasing pressures on urban water supply, however, the adoption rates for RWH systems have been relatively low. Cities require assurances of the long-term effectiveness and sustainability of new infrastructural investment, but the process of evidencing the efficiency of RWH systems can be complex and time consuming. Modelling these systems provides researchers and system managers a cost-effective method of analysis that would otherwise be capital intensive. Model accuracy is therefore a key challenge in the implementation of RWH systems.

In addressing these barriers, much of the existing research has focused on showcasing the effectiveness of RWH systems operating under an extensive list of climatic conditions (Jing et al., 2018), demand scenarios (Nnaji et al., 2017) and operational strategies (Soh et al., 2020). With the intention of ensuring model accuracy or improving the accessibility of modelling methods to non-researchers, a plethora of modelling and tank design approaches have been developed around the world. Quinn et al. (2021) studied how best to characterise and evaluate the performance of RWH systems in terms of meeting both the objectives of supplying water and in stormwater management. In comparing model

outputs with empirical data, Ward et al. (2012) showed that detailed methods provided the best approximation of real RWH performance, but require detailed, long-term data to achieve their full potential. This is complicated by the fact that RWH systems are sensitive towards their local environments and in-depth modelling and evaluation is typically required for each implementation, which can become impractical for water managers around the world. In this paper, a model agnostic statistical framework is presented for evaluating system risk and performance as a practical strategy for evidencing the long-term sustainability and efficiency of RWH systems.

2. Methods

To map RWH system behaviors with rainfall types and characteristics, system response modes were derived from k-means clustering of performance indicators. This was coupled with a Partial Least Squares (PLS) model which relates the rainfall statistical markers with their corresponding system performance indicators in latent space as a fast and holistic method for evidencing RWH system performance and sustainability.

2.1. System Modelling

A digital twin which utilised mass balances for modelling storage tanks and orifice flow equations for flows between tanks within the system was previously developed for a three-tank urban RWH system (Soh et al., 2020). With historical rainfall data available at five-minute intervals, the rainfall volumes were uniformly disaggregated to match the one-second time resolution used in the digital twin model.

Using historical rainfall data divided into 24-hour windows, 1496 samples of simulated system behaviors were recorded with a focus on the system's ability in controlling water quantities under a passive operational model. The performance indicators were developed accordingly with the purposes of maximizing the water availability provided by the RWH system, protecting against surface inundation, and the prospects of doing so with minimal spatial costs. More specifically, these are:

- **Harvested Volumes:** Total volume of water collected at the end of the simulation, preferably maximized to establish the water availability levels achievable by the system.
- **Unused Capacities:** Tank capacities that have not been used throughout entire simulation period, if possible, maximized to determine the minimal tank sizes required to handle different rainfall patterns.
- **Overflow:** Maximum amount of water in a single one-second timestep that overflows from the system, which should be minimized to ensure that excess water is adequately removed from the urban catchment in the maintenance of the health and safety objectives of implementing a RWH system.
- **Maximum Discharge Rates:** Discharge rates out of the system into the larger wastewater network, measuring stresses the system may provide to its counterparts downstream and should be minimized.

2.2. K-means Clustering

The k-means clustering works to separate the system responses into groups of equal variances through minimizing the within-cluster-sum-of-squares, in turn identifying the main response modes of the RWH system. The number of significantly different response modes is determined through the optimal number of clusters for a k-means algorithm, established empirically using the Python Scikit-Learn library (Pedregosa et al., 2012) and selecting the number of clusters that demonstrated the highest average silhouette score.

Table 1: List of rainfall statistical markers used in PLS model.

Marker	Description	Representation
Maximum observed rain rate	Highest volume of water observed in a single timestep in the timeseries.	Volume, Intensity
Rain-only mean	Mean of all non-zero timesteps in the timeseries	Volume, Intensity
Rain-only standard deviation	Standard deviation of all non-zero timesteps in rainfall timeseries	Volume, Intensity
Non-zero timesteps	Number of wet timesteps in the timeseries	Duration, Distribution
Maximum continuous rainfall	Maximum number of continuous timesteps with non-zero rainfall in the timeseries.	Duration, Distribution
Adjusted precipitation concentration index (PCI)	Measures the spread of rainfall within the timeseries. Defined as a function of timeseries with N timesteps with mean μ and standard deviation σ: $$PCI = \frac{1}{N}\left[1 + \left(\frac{\sigma}{\mu}\right)^2\right] \times 100(\%)$$	Duration, Distribution

2.3. PLS Model

The PLS model works to relate model inputs with its outputs though their latent spaces and was built using the Python toolbox PyPhi (García-Muñoz, 2019). As the PLS is a projection-based method, the selection of input variables should be independent. From the set of rainfall timeseries data, a set of statistical markers, presented in Table 1, were extracted such that the volumes, intensities, duration, and distribution of rainfall in a single 24-hour timeseries is represented as much as possible.

2.4. Supplementing clustering results with PLS scores

Scores, or projections of the statistical markers into latent space, provided by the PLS model were coupled with the results derived from the k-means clustering. With each score labelled by their corresponding cluster label, clusters in the latent space, as well as their corresponding driving factors can be identified. The relative positions between input samples in latent space highlights the differences in the original input variables for these samples. This can be used to indicate the properties that change between two given sample points, and by extension, used to identify the rainfall properties that distinguish two given response modes.

3. Results and discussion

3.1. Clustering and the classification of response modes

Through the clustering of the performance indicators, four main response modes for the system were identified. Boxplots visualising the distribution of the performance indicators in each cluster were used to characterise each response mode, where:

- **Cluster 0** contains system responses that have low water availability, but a smaller tank would suffice and has minimal discharge rates for high integrability.
- **Cluster 1** represents system behaviors that demonstrate high water availability and discharge rates, but without utilizing much of the existing tank capacity.
- **Cluster 2** responses show low discharge rates out of the system, with high water availability and little need for a large tank capacity.
- **Cluster 3** includes responses that have the highest discharge rates and water availability out of all the response modes. However, this is also accompanied by the need for a larger tank capacity.

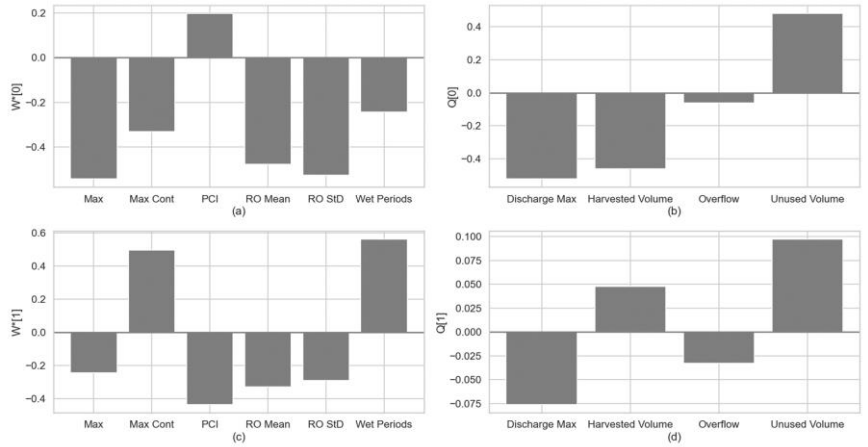


Figure 1: Projection coefficients used in PLS model for (a) input vectors onto latent variable 1, (b) output vectors onto latent variable 1, (c) input vectors for latent variable 2, and (d) output vectors for latent variable 2.

An analysis of the rainfall profiles contained in each cluster showed that rainfall volume is a significant driver of the response modes of the system. This follows with the rainfall profiles in Cluster 3 with the highest rainfall peaks, and profiles in Cluster 0 containing rainfall data with low intensities and volumes. Both clusters have sub-optimal performance, where Cluster 3 demonstrates a risk of overflow occurring, and Cluster 0 providing low water yields. The cluster sizes for these modes quantified the risk of undesirable system performances, with 4.5% and 48.2% of the total rainfall scenarios demonstrating possible overflow and low water availability respectively. The responses in Clusters 1 and 2 are much more desirable, and the rainfall characteristics distinguishing between these response modes and the sub-optimal modes would be identified alongside the statistical PLS model.

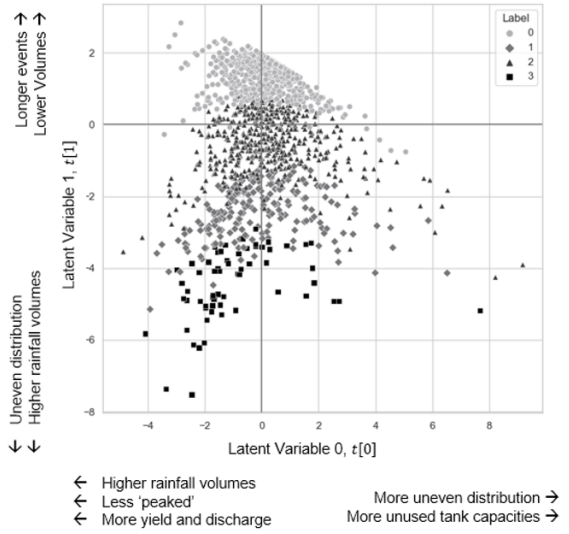


Figure 2: Score scatter of rainfall markers in latent space and main variable shifts along axes.

3.2. PLS model and the identification of main system driving forces

The PLS model identified two main driving forces in the RWH system, which are represented by the two latent variables found to best represent the input and output vectors. The latent variables were found to be able to explain the variation in 87.5% of the input, and 59.3% of the output. Whilst this study focuses on characterizing the impact of rainfall on the system performance, the behavior of a RWH system is expected to also be highly dependent on its internal design and is assumed, in this case, to account for the other 40.7% of variation observed in the outputs. Figure 1 shows the projection coefficients derived from the PLS model for both latent variables in the input and output vectors. The first input latent variable has coefficients for all inputs variables except the PCI marker positively correlated with each other and hence it is associated generally with the input rainfall volumes. The second input latent variable is associated with the rainfall duration and distribution, demonstrating that longer rainfall events are typically correlated with lower rainfall volumes and intensity.

In examining the relationship between the input and output variables in latent space, the model suggests that a higher rainfall volume, regardless of its distribution in time, is positively correlated with a higher discharge rate and improved water availability. Higher rainfall volumes also tend to require a larger tank, as observed in the first latent variable, the dimension associated with the overall rainfall volumes. Coefficient values in the second output latent variable are very low, hence correlations shown between the input and output vectors in this dimension are not significant.

3.3. Distinguishing rainfall properties between response modes

The input latent variables were plotted in latent space and coupled with response mode labels derived from the k-means clustering discussed in Section 3.1. Figure 2 shows the input variables represented in latent space and the associated changes in input variables in each axis direction. There are clear distinctions between each response mode in the input latent variable space. With Cluster 3 lying mostly within the high rainfall volume quarters and Cluster 0 in the low rainfall quarters, this confirms the initial analysis about rainfall volumes playing a key role in influencing the system's response mode.

For a given value for the second latent variable, the extreme values of the first latent variable typically belong to a response mode different from that of the central values, giving a parabolic structure to the clusters. In conjunction with the loadings plot shown in Figure 1(a), these extreme values in the first latent variables would correspond to either, in the negative direction, rainfall with more consistent and high rainfall volumes, or in the positive direction, more unevenly distributed rainfall patterns.

Rainfall days either with consistent but low intensity rainfall delivering a high total rainfall volume, or sporadic but high-intensity rainfall delivering lower total rainfall volumes would, therefore, drive the system towards a response mode with lower system stresses. This is characterised by smaller tank capacities, better integrability with its downstream systems, but accompanied by a reduced ability in ensuring water availability. This is evident in how the samples belonging to Cluster 2, which is associated with lower discharge rates and water availability than in Cluster 1, wraps the input latent space samples belonging that cluster in the second latent variable space.

4. Conclusions

The coupling of the k-means clustering method with a statistical PLS model successfully identified specific rainfall features that distinguishes between the response modes of the

RWH system in evaluation. For the given case study, the consistency in the delivery of the rainfall input is key in determining its response mode. Uniform rainfall rates that contribute to a high total rainfall volume, or sporadically intense rainfall would lead the system towards a response mode with lower water availability, but with lower capacity requirements and improved integrability with downstream systems.

This framework for evaluating the performance of a given RWH system also characterised the risks of the system responding sub-optimally. Under passive operation, the three-tank RWH system examined in this work has been found to underperform in 52.7% of the simulated scenarios. With the inclusion of design and or control parameters, their associated impacts can also be evaluated to ensure that the response modes for a RWH design is adequately within the desirable performance bounds. In understanding the rainfall characteristics that drive the system from one response to another, the framework can be used to assess future risks and hence the long-term sustainability of the system using predicted changes to rainfall runoff characteristics. With this information, necessary precautions that target the root changes can be easily designed into the system.

Acknowledgements

This research is supported by the Singapore Ministry of National Development and the National Research Foundation, Prime Minister's Office under the Land and Liveability National Innovation Challenge (L2 NIC) Research Programme (L2 NIC Award No. L2NICTDF1-2017-3). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of the Singapore Ministry of National Development and National Research Foundation, Prime Minister's Office, Singapore.

References

Jing, Xueer, Shouhong Zhang, Jianjun Zhang, Yujie Wang, Yunqi Wang, and Tongjia Yue. 2018. "Analysis and Modelling of Stormwater Volume Control Performance of Rainwater Harvesting Systems in Four Climatic Zones of China." Water Resources Management 32 (8): 2649–64.

García-Muñoz, Salvador. 2019. "PyPhi." https://github.com/salvadorgarciamunoz/pyphi.

Nnaji, Chidozie Charles, Praise God Chidozie Emenike, and Imokhai Theophilus Tenebe. 2017. "An Optimization Approach for Assessing the Reliability of Rainwater Harvesting." Water Resources Management 31 (6): 2011–24.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. 2012. "Scikit-Learn: Machine Learning in Python." Environmental Health Perspectives 127 (9): 97008.

Quinn, Ruth, Charles Rougé, and Virginia Stovin. 2021. "Quantifying the Performance of Dual-Use Rainwater Harvesting Systems." Water Research X 10.

Soh, Qiao Yan, Edward O'Dwyer, Salvador Acha, and Nilay Shah. 2020. "Optimization and Control of a Rainwater Detention and Harvesting Tank." In Computer Aided Chemical Engineering, 48:547–52.

Ward, S., F. A. Memon, and D. Butler. 2012. "Performance of a Large Building Rainwater Harvesting System." Water Research 46 (16): 5127–34.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50338-9

Global assessment and optimization of renewable energy and negative emission technologies

Lanyu Li[a,b], Jiali Li[a] and Xiaonan Wang[a,c,*]

[a] *Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore, 117585*
[b] *NUS Environmental Research Institute, National University of Singapore, Singapore, 138602*
[c] *Department of Chemical Engineering, Tsinghua University, Beijing, P.R. China, 100084*
wangxiaonan@tsinghua.edu.cn

Abstract

Energy consumption can be a great environmental burden with heavy greenhouse gas emissions. Renewable energy, negative emission technologies, and waste-to-energy technologies are promising methods to assist in transitioning the energy- and carbon-intensive current energy systems towards low-carbon systems, mitigating emissions, contributing to carbon neutrality targets, and even achieving negative emissions. The availability of renewable resources is temporally uncertain and geographically different. Besides, the optimization and the conflicting economic and environmental trade-off of such systems have not yet been fully investigated in the literature. This study aims to provide a versatile framework for the assessment and sizing of renewable and negative emission technologies for global regions. The study also sheds light on the economic viability and carbon mitigation/reduction potential achievable by the combination of renewable energy and negative emission technologies regionally and globally.

Keywords: Hybrid renewable energy system, Negative emission, Design, Optimization, Analysis.

1. Introduction

Throughout history, we have gone through the energy transition from human and animal power to steam and from steam to electricity. The evolution of electricity has brought us tremendous economic prosperity and shocking changes in our daily lives, from the way we live to the way we think. However, while we enjoy an improved quality of life and convenience, we also pay a huge price for burning fossil fuels, rich reserves of carbon that can be stored well underground for thousands of years without human extraction. The world of fossil fuels has brought with it a dramatic increase in greenhouse gases in the atmosphere and the associated problem of climate change.

Figure 1 shows the emissions of the six major carbon emitters globally. Until the mid-20th century, the United States and Europe dominated the world's major carbon emissions. Emissions from the rest of the world began to rise significantly from the second half of the 20th century, with major increases in Asia, particularly China. Since 2006, China has surpassed the United States as the world's largest emitter of carbon dioxide. The main reasons can be attributed to the country's rapid industrialization, urbanization, and reliance on major fossil fuels such as coal. However, in terms of per capita greenhouse gas emissions, the United States., Russia, and Europe continue to be the highest per capita emitters. But since 2000, the per capita carbon emissions in the United States has started to fall, which is against the climb and stabilization of China's carbon emissions around the same period.

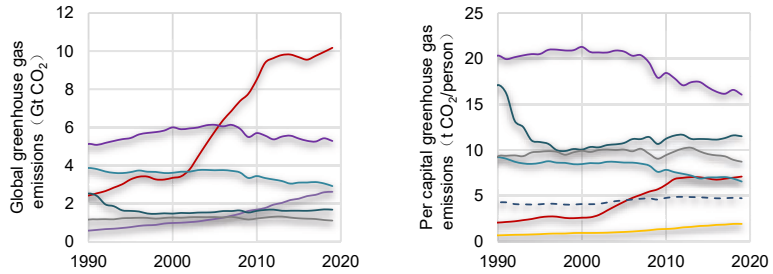


Fig. 1 Greenhouse gas emissions of the top six emitters around the world (Andrew, 2020; Hannah Ritchie; Max Roser, 2020).

Having experienced previous energy transitions, today, in the first half of the 21st century, we are about to witness another energy transition: from a fossil fuel-dependent energy system to a low-carbon, renewable, sustainable energy system. There is a growing awareness of the drawbacks and dilemmas of producing energy from fossil fuels, air pollution, greenhouse gas emissions and depletion, and the more difficult extraction that comes with it. Initiated by non-governmental organizations, the IPCC helped set the goal of carbon emissions, a goal of limiting global warming to well below 2°C, preferably 1.5°C, compared to pre-industrial levels. To achieve this goal, many countries set up their targets to peak their domestic greenhouse gas emissions in order to achieve a carbon-neutral world by the middle of the century.

Since 2015, 193 countries have submitted their climate commitments. At least 50 of these countries have set net-zero emission targets (*Members of the Carbon Neutrality Coalition*, 2021; Darby and Gerretsen, 2021; Wallach, 2021). Table 1 lists the carbon neutrality targets for several countries, including the major carbon-emitting economies such as China, the United States, the European Union, and Japan (ranked by annual carbon emissions). In addition to China's neutrality target, the United States has also proposed a statement of intent to achieve net-zero emissions by 2050 and 100% clean electricity by 2035. The governments of South Africa and Chile have also expressed their policy position of achieving net-zero emissions by 2050. The EU, Japan, South Korea, Canada, and Brazil have established explicit legislation or submitted written commitments to the United Nations to become carbon neutral in the coming decades.

Table 1 Net-zero carbon emissions goals set by different countries.

Country	Target date for carbon neutrality	Commitment
China	2060	Statement of Intent
United States	2050	Statement of Intent
EU	2050	Submitted to the United Nations
Japan	2050	Law
Korea	2050	Submitted to the United Nations
Canada	2050	Law
Brazil	2060	Submitted to the United Nations
South Africa	2050	Policy Position
Chile	2050	Policy Position

Under the common vision of humanity to reduce carbon emissions, renewable energy and negative emission technologies are promising technologies in the blueprint of low-carbon transition of energy systems (Li *et al.*, 2019). However, hybrid renewable energy systems combined with negative emission technologies, which have a synthetic effect and promising potential in building low-carbon energy systems, are still under-explored. Therefore, here we develop a versatile framework for the assessment and scaling of renewable and negative emission technologies in global regions, and to elucidate the economic feasibility and carbon reduction potential of these technologies.

2. Methodology

The study is carried out through data collection, modeling, and optimization based on the modeling method in Li *et al.* (2019), which is summarized in Figure 2. The inputs for the optimization model include meteorological, biomass, land area, economic, and social data for each of the countries. The coordinates for cities are obtained from *ArcGIS online* (2021). The models for renewable energy include solar PV, onshore wind, gasification, incineration, and pyrolysis. Biomass resources available for waste-to-energy conversion include horticultural wastes and wood wastes. The economic performance of the system is evaluated by the net present value. The environmental performance is quantified by the equivalent CO_2 emission. Either or both of them can be used as the objective function of the optimization. The constraints of technology placement include the resource limit for power generation, the requirement of demand satisfaction, and the design of operating constraints for the specific technologies. It is interesting to see what the technology planning would be like if the optimization is done to maximize each country's own benefit and the case maximizing for the global benefit as a whole. Therefore, we are carrying out the scenario analysis for three different scenarios: the case for maximized economic performance for each country, the case for maximized economic performance for the world as a whole, and the case for minimized greenhouse gas emissions globally.

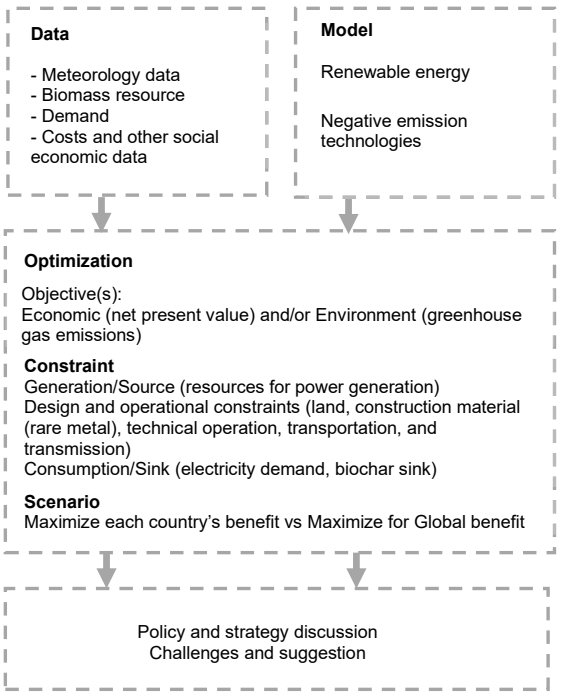


Fig. 2 Methodological framework.

3. Results and Discussion

We first calculate the maximum technology capacity achievable around the world without considering any constraints to have an overview of the resource potential by technology and by region. Table 2 summarizes the maximum potential of technology capacity sorted by power generation method.

Table 2. The maximum potential of technology capacity by power generation method (kW)

	Solar	Wind	Combustion	Gasification	Pyrolysis
Capacity	4.99×10^{13}	7.08×10^{12}	1.11×10^{7}	9.51×10^{6}	6.73×10^{6}
Percentage over total available potential	88%	12%	~ 0%	~ 0%	~ 0%

It can be found that the magnitudes of solar and wind availability are much higher than those for biomass conversion methods including combustion, gasification, and pyrolysis. Besides, the maximum total energy generation potential by the technologies by country is shown in Figure 3. It can be found that Russia, French Southern Territories, Chile, Bolivia, and the United States have the most abundant (mainly solar) energy resource for the technologies considered.

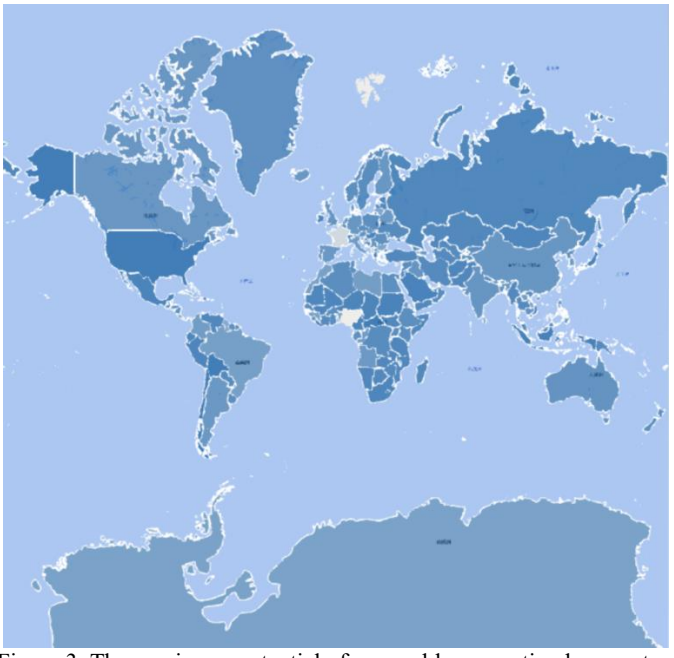


Figure 3. The maximum potential of renewable generation by country.

With this basis, we carried out the optimization by maximizing the case for maximized economic performance for each country (Scenario 1), maximizing the economic performance for the world as a whole (Scenario 2), and minimizing the greenhouse gas emissions globally (Scenario 3). The optimized results are provided in Table 3. It was found that maximizing the economic performance for countries and regions individually and optimizing globally as a whole result in almost the same economic and environmental performance. Moreover, minimizing the environmental performance as the objective function could result in an overall negative greenhouse gas emission, but it may lead to a dramatic increase in cost compared to the optimal NPV scenario. For the cases of maximizing the economic benefit, it was found that almost all regions, except 15 of them, are economically preferable for the utilization of the considered technologies no matter it is optimized for each country (Scenario 1) or globally (Scenario 2). Solar energy is the most selected technology for most regions in these cases. On the other hand, fourteen regions have no technology selected when the greenhouse gas emission is minimized as an environmental objective mainly due to the limited biomass resources.

Table 3. Result for different optimization scenarios.

	Scenario 1: Maximum economic performance for each country	Scenario 2: Maximum economic performance globally	Scenario 3: Minimum environmental performance globally
Net present value	2.15×10^{14}	2.18×10^{14}	-3.4×10^{10}
Greenhouse gas emissions	1.11×10^{12}	1.11×10^{12}	-3.7×10^{10}

4. Conclusion

This work presents the economic assessment and the potential for carbon reduction via renewable energy and negative emission technologies for countries around the world. The geographical diversification of renewable resources and system design was observed in global analysis. The result shows that except for 15 places, almost all countries and regions around the world were decided to be profitable locations for the proposed system when net present value is maximized. Negative emission was possible to be achieved globally if greenhouse gas emission was minimized, but it may lead to a dramatic increase in cost compared to the optimal NPV scenario. The methodology framework was demonstrated to be versatile and conveniently applicable to study the feasibility of the proposed renewable and negative emission technologies in multiple regions. In future studies, the impact of grid integration of renewables, a higher-precision analysis accounting for the temporal-spatial variation of various factors can be carried out to provide more insights.

Acknowledgment

This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) program.

References

Andrew, R. (2020) *Figures from the Global Carbon Budget 2020*. Available at: https://folk.universitetetioslo.no/roberan/GCB2020.shtml (Accessed: 23 August 2021).

ArcGIS online (2021). Available at: https://www.arcgis.com/index.html.

Darby, M. and Gerretsen, I. (2021) *Which countries have a net zero carbon goal?* Available at: https://www.climatechangenews.com/2019/06/14/countries-net-zero-climate-goal/ (Accessed: 1 August 2021).

Hannah Ritchie; Max Roser (2020) *CO_2 and Greenhouse Gas Emissions*, *OurWorldInData.org*.

Li, L. *et al.* (2019) 'Optimal design of negative emission hybrid renewable energy systems with biochar production', *Applied Energy*. Elsevier, 243, pp. 233–249. doi: 10.1016/J.APENERGY.2019.03.183.

Members of the Carbon Neutrality Coalition (2021). Available at: https://carbon-neutrality.global/members/ (Accessed: 1 August 2021).

Wallach, O. (2021) *Race to Net Zero: Carbon Neutral Goals by Country*. Available at: https://www.visualcapitalist.com/race-to-net-zero-carbon-neutral-goals-by-country/ (Accessed: 1 August 2021).

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50339-0

The Trade-Off between Spatial Resolution and Uncertainty in Energy System Modelling

Maria I. Yliruka[a*], Stefano Moret[b], Francisca Jalil-Vega[c], Adam D. Hawkes[a], Nilay Shah[a]

[a]*Department of Chemical Engineering, Imperial College London, United Kingdom*
[b]*Business School, Imperial College London, United Kingdom*
[c]*Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez, Chile*
m.yliruka@ic.ac.uk

Abstract

In energy system models, computational tractability is often maintained by adopting a simplified temporal and spatial representation in a deterministic model formulation i.e., neglecting uncertainty. However, such simplifications have been shown to impact the optimal result. To address the question of how to prioritize the limited computational resources, the trade-off between spatial resolution and uncertainty is assessed by applying a novel method based on global sensitivity analysis to a peer-reviewed heat decarbonization model. For all output variables apart from the total system and fuel cost, spatial resolution is ranks amongst the five most important model inputs. It is the most relevant factor for investment decisions on network capacities. For the total fuel consumption and emissions, spatial resolution turns out to be more relevant than the fuel prices themselves. Compared across all outputs, the analysis suggests the impact of spatial resolution is comparable the impact of heat demand levels and the discount rate.

Keywords: Spatial Resolution, Uncertainty, Mixed-integer Linear Program, Energy System Model, Global Sensitivity Analysis

1. Introduction

Energy system models help to explore different decarbonization pathways to reach net zero by 2050. Given the time horizon and scope, these models are often large and rely on long-term forecasts of input parameters. Computational tractability is maintained by adopting a simplified temporal and spatial representation of the system as well as a "deterministic" model formulation (i.e., neglecting uncertainty). However, the shortcomings of these approaches have been shown, notably in power systems applications: At low temporal resolutions, the dispatchable generation and storage capacities are underestimated whereas the renewable generation capacity is overestimated due to the smoothed production and demand profiles. This impact is especially pronounced for systems with high shares of intermittent and non-dispatchable renewables (Pfenninger, 2017). While the impact of temporal resolution and suitable aggregation methods have already been reviewed, the literature on the impact of spatial resolution is more limited and less conclusive. In power systems, the impact of spatial resolution on the total system cost is small compared to the changes in generation and flexibility technologies. Hörsch and Brown (2017) observe an increased investment in transmission capacity and a decrease in solar PV with increased spatial resolution for Europe, while Krishnan and Cole (2016) report the opposite effect for the US. In Jalil-Vega and Hawkes

(2018a), the averaging of heat demand densities at lower spatial resolution leads to an underestimation of district heating potential for local authorities (LA) with heterogenous demand levels in the UK. Global sensitivity analysis (GSA) has been applied to national energy system models to quantify the impact of uncertainty of model inputs (input parameters) (Pye et al., 2015; Moret et al., 2017). Economic parameters such as fuel prices and investment cost of technologies have commonly been identified as most relevant. Overall, these recent studies leave the modelling community with the question of how to prioritize computational resources: including uncertainty or increasing spatial/temporal resolution?

To our knowledge, this paper is the first to compare the impact of spatial resolution and uncertainty of input parameters. A novel methodology is applied to a case study in literature to assess this trade-off over a wide range of output variables. The fundamental novelty is that spatial resolution is considered as an uncertain input parameter in a GSA allowing to rank its importance relative to other uncertain input parameters.

2. Methodology

2.1. Novelty

GSA methods allow to quantify the impact of input parameters on one or more model outputs (Saltelli et al., 2008), where *impact* is defined as the ability of an input parameter to significantly alter a given output of interest when varied from its nominal value that corresponds to its most likely realization. Modelling choices such as spatial resolution are not commonly referred to as 'input parameters' and have therefore never been considered in GSA studies. In this work, we propose using GSA methods in a novel way that allows to assess the trade-off between spatial resolution and uncertainty in energy system models within the same methodological framework.

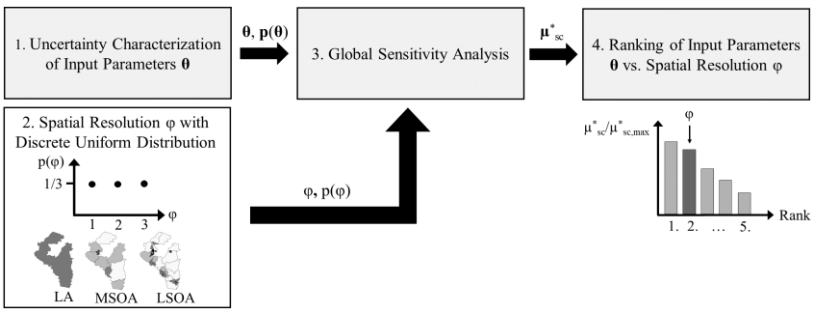


Figure 1: A summary of the three steps (1, 3, 4) involved in a conventional GSA (top) with the addition of spatial resolution (2) as an uncertain input parameter with discrete uniform distribution (bottom). In the case study, three distinct levels of spatial resolution (N=3) are modelled, having equal probability of 1/3.

Our proposed modified GSA method is illustrated in Figure 1. In conventional GSA studies, the uncertainty distributions $\mathbf{p(\theta)}$ of the input parameters $\boldsymbol{\theta}$ serve as inputs to the GSA. In our novel approach, the spatial resolution is included as an additional "uncertain" parameter (φ) characterized by a discrete uniform distribution $p(\varphi)$. The finite number of spatial resolution levels are characterized by a numerical value 1, 2, …, N with N being total number of different resolutions considered in the analysis. Each level of resolution has equal probability $1/N$. This modification allows us to consider spatial resolution alongside the usual input parameters in the GSA, hence assessing its relative impact with

respect to uncertainty in model inputs. In the following, spatial resolution and input parameters will collectively be referred to as *input factors*.

2.2. Global Sensitivity Analysis

The modified Morris Method is applied to obtain a qualitative ranking over a large set of input factors. To determine the elementary effect (EE) of each input factor on the outputs of interest, the input space is discretized into a p-level grid and systematically sampled using r trajectories. At each step of the trajectory, one of the k input factors is varied changing its value by Δ with $\Delta=p/(2(p-1))$ (Saltelli et al., 2008). To avoid the computational burden of oversampling, the enhanced Sampling for Uniformity (eSU) strategy is applied (Chitale et al., 2017). The EE of input factor θ_i on output variable Y_j in trajectory m, EE_{mij}, is subsequently calculated as the ratio between the change in the input factor and the consequent change in the output variable, as shown in Eq. (1).

$$EE_{mij} = \frac{\delta Y_j}{\Delta} \tag{1}$$

Campolongo et al. (2007) have shown that μ^*_{ij}, the mean of the distribution of the absolute values of EE_{mij}, is a good proxy to the total effect sensitivity index, that is normally determined using computationally expensive, variance-based methods. To compare the impact on multiple output variables of varying magnitudes and units, a similar approach to Sin and Gernaey (2009) is chosen by scaling μ^*_{ij} by σ_j, the standard deviation of the output variable Y_j:

$$\mu^*_{sc,ij} = \frac{1}{\sigma_j}\left(\frac{1}{r}\sum_{m=1}^{r}\left|EE_{mij}\right|\right) \tag{2}$$

2.3. Case Study

The method is applied to a spatially resolved, mixed-integer linear model for urban energy systems (Jalil-Vega and Hawkes, 2018b), which was previously used to study the impact of spatial resolution on the district heating uptake in six LAs with varying rural-urban character (Jalil-Vega and Hawkes, 2018a). The LA of Winchester is chosen and modelled at LA (1 node), middle layer super output area (MSOA, 10 nodes) and lower layer super output area (LSOA, 49 nodes) level as shown in Fig. 1. With increasing spatial resolution, the initial network capacity between and within cells, the heat and electricity demand and the distance between cells become more resolved. The domestic heat, gas and electricity supply infrastructure are explicitly modelled. The design and operation of the heat supply system minimizing the total system cost is determined from today until 2050. The same nomenclature as in Jalil-Vega and Hawkes (2018b) is used.

Apart from the lifetime of technology and network, all model parameters presented in Jalil-Vega and Hawkes (2018b) are considered as uncertain: Fuel prices ($Cost^E/Cost^G$), capital cost of technologies ($Cost^C_{tech}$), technological performance (η^{Th}_{tech} /COP_{tech} /η^{E}_{tech}), capital cost of intranodal network capacity ($Cost^{ND}_{E/G/H}$), capital cost of internodal network capacity ($Cost^{NT}_{E/G/H}$), operation and maintenance cost ($Cost^M_{tech}$), losses in heat networks ($Loss_T$), discount rate (r), electricity and heat demand (Dem^E/Dem^H). The uncertainty ranges, $R_{\%,}$ summarized in Table 1 are determined based on UK specific data where possible (Yliruka et al., in preparation). For all input parameters, a uniform distribution is assumed.

Table 1: Uncertainty ranges, $R_{\%}$, for the input parameters in Jalil-Vega and Hawkes (2018b) (Yliruka et al., in preparation).

Parameter	$R_{\%}$ [%]
$Cost^{C}_{B/Erad/HXT}$	[-40, 42]
$Cost^{C}_{ASHP}$	[-31, 49]
$Cost^{C}_{GSHP}$	[-40, 44]
$Cost^{C}_{CHP}$	[-11, 9]
$Cost^{C}_{PV}$	[-57, 76]
$\eta^{Th}_{B/Erad/HXT}$	[-0.2, 0.4]
COP_{ASHP}	[-14, 14]
COP_{GSHP}	[-11, 14]
η^{E}_{CHP}	[-10, 13]
η^{E}_{PV}	[-14, 17]

Parameter	$R_{\%}$ [%]
$Cost^{E}$	[-13, 15]
$Cost^{G}$	[-38, 0]
$Cost^{CO2}$	[-50, 50]
Dem^{E}	[-0.4, 15]
Dem^{H}	[-39, 6]
$Cost^{NT}_{E/G/H}$	[-39.3, 39.3]
$Cost^{ND}_{E/G/H}$	[-39.3, 39.3]
$Cost^{M}_{tech}$	[-49.2, 35.7]
$Loss_{T}$	[-2, 2]
r	[-81, 4]

3. Results and Discussion

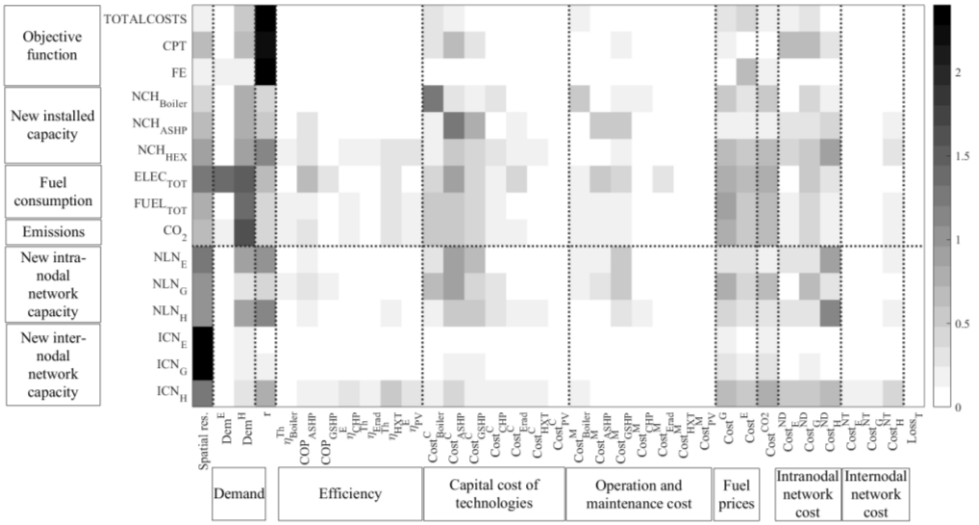


Figure 2: The impact of each input parameter on a selected subset of output variables is indicated by the shading in the heat map. The inputs and outputs are grouped by boxes and categorized by vertical lines. The horizontal line separates the network related outputs.

The results are based on r=24, p=4 for each uncertain input parameter and p=3 for the spatial resolution. The impact of the different input factors on a selection of output variables are summarized in Fig. 2. The fuel consumptions, emissions, technology and network capacities are summed over the multi-year time horizon. The scaling by σ_j in Eq. 2 also allows to compare the qualitative impact across all output variables. The darker the column, the more relevant is the parameter across multiple output variables.

The degree of spatial resolution is by far the most important factor on the installed capacity of all internodal networks ($ICN_{E/G/H}$). This ranking holds true even if the large

change in $ICN_{E/G/H}$ between LA and MSOA level is excluded. Spatial resolution also remains the most important factor for the intranodal gas and electricity network capacity ($NLN_{E/G}$). However, in the case of heat networks (NLN_H), the investment cost of the pipelines and the discount rate are more relevant. While for gas and electricity networks the initial network needs to be reinforced, the heat network has to be built from scratch. Therefore, its deployment is limited by the high upfront cost that is sensitive to the capital cost of the network and the discount rate. The general observation that the network design is the most sensitive output variable to the choice of spatial resolution agrees with Hörsch and Brown's (2017) observations in power system models.

For the other output variables presented in Fig. 2, the impact of spatial resolution is less conclusive. The total system cost (TOTALCOSTS), capital cost (CPT) and fuel cost (FE) are dominated by the discount rate. Here, the level of spatial resolution becomes irrelevant. Garcia-Gusano et al. (2016) have previously discussed the decisive role of the discount rate on the system cost in long-term energy system models. Out of all heat technologies, the installed capacity of heat exchangers (NCH_{HEX}) is most sensitive to the choice of spatial resolution. Its impact is comparable to the heat demand and heat network investment cost. However, like NLN_H, NCH_{HEX} is the most sensitive to the discount rate. For the installed capacity of boilers (NCH_B) and ASHPs (NCH_{ASHP}), the investment cost of the technology itself is most important.

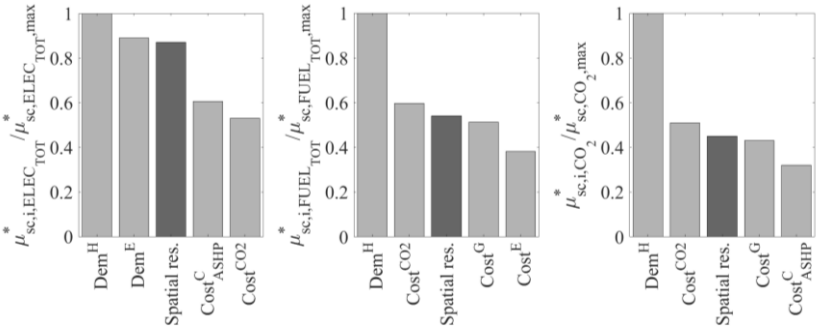


Figure 3: The spatial resolution ranks third (*dark grey*) with respect to the total emissions, electricity and natural gas consumption.

The impact of spatial resolution on the fuel consumption ($FUEL_{TOT}$, $ELEC_{TOT}$) and CO_2 emissions (CO_2) has not yet been discussed in the literature. As shown in Fig. 3, spatial resolution has a higher impact than the fuel prices. As the emissions are mostly caused by the natural gas consumption, both output variables are most sensitive to the heat demand and carbon tax. For $ELEC_{TOT}$, the demand levels for both heat and electricity are the most important. The high rank of spatial resolution indicates that the split in fuel consumption is governed by the installed network capacities that are a highly dependent on the level of spatial resolution. The choice in heat technology is secondary and subsequently determined by the relative capital cost. As the ranking can vary for different $R_{\%}$ (Moret et al., 2017), these findings should be confirmed for other $R_{\%}$ of the fuel prices. Spatial resolution ranks amongst the five most impactful input factors for all output variables apart from the total system and fuel cost. As shown previously, spatial resolution is the most relevant for the network capacities. Studying for the first time its impact on the total fuel consumption, spatial resolution turns out to be more relevant than the uncertainty in fuel prices. Overall, spatial resolution is found to be comparable to the discount rate and heat demand levels.

4. Conclusions

Neglecting uncertainties of input parameters and temporal/spatial aggregation have been shown to impact the optimal solution of energy system models, leaving the modelling community with the question of how to prioritize the limited computational resources. To guide this decision, this paper compares for the first time the impact of spatial resolution and uncertain input parameters across a range of design and operational variables of a peer-reviewed heat decarbonization model.

The choice of spatial resolution is the most relevant for the design of the networks whereas a single node system is well-suited to estimate the total system cost and capacity of heat technologies. The rankings for the total system cost and network variables agree with previous literature which serves as a verification of the method. For the first time, the impact of spatial resolution on the fuel consumption and emission levels is assessed and identified as relevant, ranking third after demand levels and carbon tax.

The novel, GSA-based framework can help prioritize the allocation of limited computational resources either on spatially detailed deterministic models or stochastic models with coarser spatial resolution in the early stages of the model development.

References

F. Campolongo, J. Cariboni, A. Saltelli, 2007, An effective screening design for sensitivity analysis of large models, Environmental Modelling & Software, 22, 10, 1509-1518

J. Chitale, Y. Khare, R. Munoz-Carpena, G. S. Dulikravich, C. Martinez, 2017, An effective parameter screening strategy for high dimensional models, IMECE2017, Tampa, USA

J. Hoersch, T. Brown, 2017, The role of spatial scale in joint optimisations of generation and transmission for European highly renewable scenarios, EEM, Dresden, Germany

D. Garcia-Gusano, K. Espegren, A. Lind and M. Kirkengen, 2016 The role of the discount rates in energy systems optimisation models, Renewable and Sustainable Energy Review, 59, 56–72

F. Jalil-Vega, A. Hawkes, 2018a, The effect of spatial resolution on outcomes from energy systems modelling of heat decarbonisation, Energy, 155, 339-350

F. Jalil-Vega, A. Hawkes, 2018b, Spatially resolved model for studying decarbonisation pathways for heat supply and infrastructure trade-offs, Applied Energy, 210, 1051-1072

V. Krishnan, W. Cole, 2016, Evaluating the Value of High Spatial Resolution in National Capacity Expansion Models using ReEDS, PESGM, Boston, USA

S. Moret, V. Codina Girones, M. Bierlaire, F. Marechal, 2017, Characterization of input uncertainties in strategic energy planning models, Applied Energy, 202, 597

G. Sin, K. V. Gernaey, 2009, Improving the Morris method for sensitivity analysis by scaling the elementary effects, Computer Aided Chemical Engineering, 26, 925

S. Pfenninger, 2017, Dealing with multiple decades of hourly wind and PV time series in energy models: A comparison of methods to reduce time resolution and the planning implications of inter-annual variability, Applied Energy, 197, 1-13

S. Pye, N. Sabio, N. Strachan, 2015, An integrated systematic analysis of uncertainties in UK energy transition pathways, Energy Policy, 87, 673-684

A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, 2008, Global Sensitivity Analysis: The Primer, John Wiley & Sons, England

M. I. Yliruka, S. Moret, N. Shah, in preparation, Uncertainty characterisation for the UK energy system

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50340-7

Designing a Resilient Biorefinery System under Uncertain Agricultural Land Allocation

Varun Punnathanam[a], Yogendra Shastri[a*]

[a]*Department of Chemical Engineering, Indian Institute of Technology Bombay, Mumbai*

yshastri@iitb.ac.in

Abstract

Agricultural residues are excellent feedstock for lignocellulosic biorefineries. However, the land allocated to various crops in a region can vary annually, thus impacting the feedstock availability for biorefineries. This work provides an optimization framework that considers uncertainty in land allocation for designing a biorefinery system that is resilient to such changes. A recently proposed decomposition-based approach is utilized to perform stochastic optimization, and the resulting design was compared to a deterministic design that considered mean land allocation. Lignocellulosic ethanol production for the state of Maharashtra, India, was taken as a case study, and the performances of both designs were evaluated on a set of 100 random land allocation instances. The resilient design had a smarter feedstock procurement strategy which resulted in a significant decrease in variation of feedstock procurement and transportation expenses. As a result, the variation in ethanol cost was 4% for the resilient design, as compared to 11% for the deterministic design.

Keywords: Stochastic optimization, large-scale optimization, uncertain feedstock availability, second generation biorefinery.

1. Introduction

The production of ethanol in India is expected to ramp up significantly to meet the target of 20% blending with gasoline by the year 2025, with significant contribution from second generation biorefineries. The feedstock for these biorefineries are to be primarily lignocellulosic biomass in the form of agricultural residues. A systems based approach is needed to design such production systems to address challenges of distributed biomass availability, multiple feedstock types, seasonality of feedstock, and the low maturity of the processing technologies (Daoutidis et al., 2013; Ng and Maravelias, 2017). Additionally, the uncertainty in demand, raw material price, and conversions pose further challenges to decision makers (Gong et al., 2016; Guo et al., 2022).

For an agricultural residue based biorefinery system, the residue availability is directly affected by the agricultural land allocation of the region, which in turn is governed by various external factors. This work presents a stochastic optimization based approach for designing a system of biorefineries under uncertain agricultural land allocation. The model is applied to a case study of Maharashtra, India, and the performance of the resilient design obtained from stochastic optimization is compared to a deterministic design.

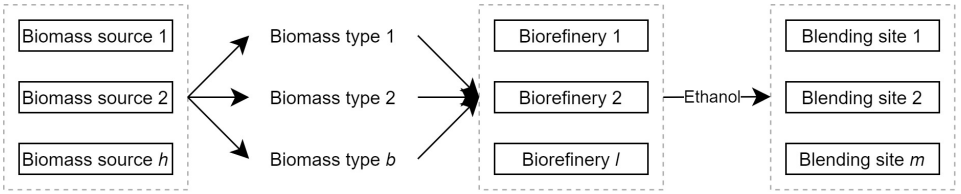


Figure 1: Schematic representation of the biorefinery system.

2. Problem formulation

The supply chain and process synthesis model for the biorefinery system from Punnathanam and Shastri (2021) is adopted in this work and schematically shown in Figure 2. This model considers multiple types of agricultural residue as feedstock, seasonal feedstock availability, biomass storage at biorefineries, multiple feedstock source locations, and multiple ethanol-gasoline blending sites. Additionally, the model considers multiple biorefinery location options, with the allowance of selection of one or more locations for setting up biorefineries. The transportation of feedstock from source to biorefineries and ethanol from biorefineries to blending sites is via trucks and tankers, respectively. The biorefinery configuration is fixed based on previous studies and is reported in Section 4.. The model is formulated as a large mixed integer linear programming problem (MILP) and is given as follows:

$$\min \quad z = \sum_{l \in L} \left(\mathbf{c}_l^{\mathrm{TC}} + \mathbf{c}_l^{\mathrm{VO}} + \mathbf{c}_l^{\mathrm{FO}} \right) \tag{1}$$

subject to Biorefinery processing constraints at location l $\qquad \forall l \in L \quad (2)$

$$\sum_{l \in L} \mathbf{f}_{h,b,l,t}^{proc} \leq \mathbf{F}_{h,b,t} \qquad \forall h \in H, b \in B, t \in T \quad (3)$$

$$\sum_{l \in L} \mathbf{e}_{l,m,t}^{\mathrm{prod}} \geq \mathbf{D}_m \qquad \forall m \in M, t \in T \quad (4)$$

The objective is to minimize to total annualized cost (TAC) of the system, which is the sum of the capital and operating expenses. The two primary sets of constraints reflect the availability constraints on the feedstock and the minimum demand for ethanol that needs to be met at each blending site by all the biorefineries. The constraints reflecting the conversion of biomass to ethanol at biorefinery l are grouped together under "Biorefinery processing constraints at location l", and are taken from Vikash and Shastri (2019). The binary variables for the model are the selection of biorefinery locations and the selection of biomass to be processed at each biorefinery at each time period. The key continuous variables are biomass procurement and ethanol distribution at each biorefinery, the capacities of the equipment within each biorefinery, and the biomass storage at each biorefinery. The mean value of the random variable can be utilized in this formulation to obtain a deterministic design for the biorefinery system. Here, design refers to the selection of biorefinery locations, capacities of equipment within each biorefinery, and the feedstock procurement plan. For the deterministic design case, the feedstock procurement plan is identical to the actual feedstock procurement in the solution to the optimization problem.

3. Stochastic optimization

While the deterministic design is generated considering the mean value of the random variable, the resilient design is generated by considering a set of random instances. The goal of the resulting stochastic optimization problem is to obtain the optimal biorefinery design that minimizes the expected TAC considering all the random instances. This optimization is formulated as follows:

$$\min \quad \bar{z} = E(z) = \frac{\sum_{s \in S} z'_s}{\mathrm{N}_S} \tag{5}$$

subject to Biorefinery processing constraints at location l for instance s $\qquad \forall l \in L, s \in S \qquad (6)$

$$z'_s = \sum_{l \in L} \left(\mathbf{c}'^{\mathrm{TC}}_{l,s} + \mathbf{c}'^{\mathrm{VO}}_{l,s} + \mathbf{c}'^{\mathrm{FO}}_{l,s} \right) \qquad \forall s \in S \tag{7}$$

$$\sum_{l \in L} \mathbf{f}'^{\mathrm{proc}}_{h,b,l,t,s} \leq \mathbf{F}_{h,b,t} \qquad \forall h \in H, b \in B, t \in T, s \in S \tag{8}$$

$$\sum_{l \in L} \mathbf{e}'^{\mathrm{prod}}_{l,m,t,s} \geq \mathbf{D}_m \qquad \forall m \in M, t \in T, s \in S \tag{9}$$

The objective function of TAC from the model presented in Section 2. has been replaced with the expected TAC, which is the average TAC across all instances assuming all instances have identical probabilities. All constraints have an additional index $s \in S$ to reflect the set of random instances. Similarly, all variables except for the design variables have an additional index $s \in S$. The design variables, being common for all instances, do not have this additional index. Each instance considered within the stochastic optimization framework has a corresponding feedstock procurement variable. The feedstock procurement plan, however, is a design variable and common for all instances. The positive difference between the instance specific feedstock procurement and the feedstock procurement plan is the quantity of feedstock that would be procured on short notice. Feedstock procured in this manner is termed as short-term procurement, and is more expensive. This is taken into account as part of the stochastic optimization formulation.

Depending on the number of random instances considered, the stochastic problem can potentially be very big. For the case study in consideration, the stochastic problem considering 10 random instances has 140,297 constraints, 420,354 variables, and 19,899 binary variables, and was not solvable in a reasonable time on an INTEL® i7-4770 3.40GHz CPU with 4GB RAM using the CPLEX® 12 MILP solver. Hence, the decomposition based approach presented in Punnathanam and Shastri (2020) was employed in this work. This method employs Dantzig-Wolfe decomposition (DWD) to simplify the original MILP such that it can be solved using the CPLEX® solver. The solution method involves the following steps: first, the deterministic optimization problem was decomposed to be solved within a DWD framework, where each sub-problem represents the constraints corresponding to a single biorefinery location. Next, the mean value of the random variable was used to solve the problem up to a specified termination criteria using DWD. Here, sub-problems and the master problem are iteratively solved; the sub-problem solutions are integer feasible solutions for each biorefinery and the mater problem assigns weights to the solutions obtained

Table 1: Types of feedstock in the form of agricultural residues and their prices ($/ton).

Cotton stalk	Sugarcane bagasse	Rice straw	Wheat stalk	Sorghum Stalk
6.7	53.3	10.4	9.5	7.2

from the sub-problems. DWD stage is terminated when the change in objective value of the master problem over iterations is below a specified threshold. On termination, based on the weights assigned to the sub-problem solutions by the master problem at the final iteration, certain binary selections corresponding to biorefinery location and biomass to be processed are rejected. This heuristic drastically simplifies the original MILP. The resulting simplified problem can be solved in the stochastic optimization framework to obtain the resilient design for the biorefinery system. As a result of the DWD-assisted simplification step, the final simplified problem had half the original number of constraints and variables, and 92% fewer binary variables, and could be solved in 12 hours. However, due to the heuristic which was employed for simplification, the optimality of the final solution is not guaranteed.

4. Case study details

The presented model is applied for the case study of ethanol production for Maharashtra, India. Maharashtra is an agriculturally intensive state with a large variety of crops under cultivation. The feedstock for biorefineries are assumed to be residues from five prominent crops; cotton, wheat, sugarcane, sorghum, and rice. The residues and their costs are provided in Table 1. These feedstocks are available only in specific months each year, and only 10% of the residue generated in farms are available for biorefineries as the rest are consumed within the farms for various purposes. Feedstock procured in short term is assumed to be twice as expensive as its nominal price. The major towns of 33 districts in the state act as collection sites from which feedstock can be distributed to biorefineries. These 33 locations also act as potential biorefinery locations. The processing configuration within each biorefinery is as follows: hammer milling for size reduction followed by dilute acid pretreatment, washing and detoxification, bioconversion via SSCF, and purification by conventional and extractive distillation. This configuration was found to be optimal for this case study Punnathanam and Shastri (2021). Ethanol blending sites are assumed to be located in 35 districts in the state. The demand for ethanol is calculated as the quantity of ethanol required to meet a 10% blending target at each district. The resulting total ethanol demand for the state was 42.47 million litres per month. Additional details on the case study is available in Punnathanam and Shastri (2021).

The land allocated towards the cultivation of a specific crop directly impacts the quantity of the corresponding crop residue available in that region. The allocation of agricultural land to various crops in the different districts of Maharashtra changes every year based on numerous external factors and hence is considered as an uncertain parameter in this work. The mean land allocated towards each crop in each district between 2010 to 2018 was utilized for the deterministic optimization problem presented in Section 2. and the DWD stage in the solution method presented in Section 3.. Random land allocation instances are generated by identifying the lower and upper bounds on land allocated towards each crop

Table 2: Comparison of the feedstock procurement plan recommended by the deterministic and resilient designs ('00 million kg).

Design	Cotton stalk	Sugarcane bagasse	Rice straw	Wheat stalk	Sorghum stalk	Total
Deterministic	634	210	190	446	507	**1,988**
Resilient	772	147	251	585	578	**2,334**

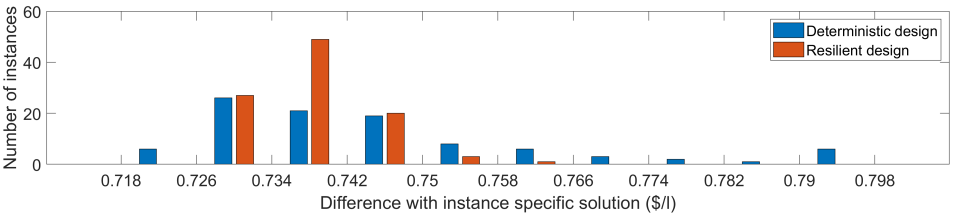


Figure 2: Frequency chart representing the performance of the deterministic and resilient designs over 100 random land allocation instances.

in each district between 2010 and 2018, and sampling from uniform distributions between these bounds. Note that each district and crop has a unique lower and upper bound. Additionally, a lower and upper bound on the total agricultural land allocated at each district was determined from historical data as well, and instances where the total land allocated was outside these bounds were scaled appropriately. For stochastic optimization, increasing the number of random instances can improve the design but significantly increase the computational time required for optimization. Based on this computational restriction, 10 random instances were considered for stochastic optimization. Similarly, a set of 100 random instances were generated for evaluating the performance of the deterministic and resilient biorefinery system designs. For a fair comparison, the total expense allocated for the feedstock procurement plan is enforced to be identical for both designs.

5. Results and discussion

The deterministic and resilient designs recommended similar locations for setting up biorefineries; with the deterministic design recommending 11 and the resilient design recommending 10 biorefineries. The feedstock procurement plans for both designs are presented in Table 2. The resilient design planned for the procurement of significantly higher quantities of cotton, sorghum, rice, and wheat based biomass, while the deterministic design planned for a higher quantity of sugarcane bagasse. Note that sugarcane bagasse is a much more expensive feedstock as compared to the others and that both designs had identical feedstock procurement plan expenses. Hence, the total quantity of feedstock planned by the resilient design was 17.4% higher than the deterministic design. As a result, the resilient design was better equipped to handle the uncertainty in feedstock availability.

Figure 2 presents the performance of both designs on the set of 100 random land allocation instances in the form of a frequency chart. At the lower end, the deterministic design was

observed to perform marginally better; the lowest ethanol costs obtained by both designs varied by 0.5%. However, at the higher end, the resilient design performed significantly better than the mean design; the highest ethanol cost obtained by the resilient design was 6% lower than the deterministic design. Moreover, the variation in cost over 100 instances was 4.2% for the resilient design as compared to 10.7% for the deterministic design. The difference in costs between the designs was primarily due to the differences in operating expenses, in particular, the feedstock procurement and transportation expenses.

6. Conclusions

This work presents a stochastic optimization framework to generate a biorefinery system design that is resilient to changes in feedstock availability caused by annual changes in agricultural land allocation. The model was applied to a case study of ethanol production in the state of Maharashtra, India. The resilient design obtained from stochastic optimization was compared to a deterministic design obtained by considering mean land allocation. The performances of both designs were evaluated on a set of 100 random land allocation instances. The feedstock procurement plan for the resilient design took into consideration the uncertainty in feedstock availability. Hence, the variation in ethanol cost for the resilient design was 61% lesser as compared to the deterministic design. This work can be extended to consider uncertainties in other parameters, such as the composition of feedstock and the yield at various processing stages.

References

Daoutidis, P., Kelloway, A., Marvin, W. A., Rangarajan, S., and Torres, A. I. (2013). Process systems engineering for biorefineries: new research vistas. *Current Opinion in Chemical Engineering*, 2(4):442–447.

Gong, J., Garcia, D. J., and You, F. (2016). Unraveling optimal biomass processing routes from bioconversion product and process networks under uncertainty: an adaptive robust optimization approach. *ACS Sustainable Chemistry & Engineering*, 4(6):3160–3173.

Guo, C., Hu, H., Wang, S., Rodriguez, L. F., Ting, K., and Lin, T. (2022). Multiperiod stochastic programming for biomass supply chain design under spatiotemporal variability of feedstock supply. *Renewable Energy*.

Ng, R. T. and Maravelias, C. T. (2017). Design of biofuel supply chains with variable regional depot and biorefinery locations. *Renewable Energy*, 100:90–102.

Punnathanam, V. and Shastri, Y. (2020). Efficient optimization of a large-scale biorefinery system using a novel decomposition based approach. *Chemical Engineering Research and Design*, 160:175–189.

Punnathanam, V. and Shastri, Y. (2021). Optimization based design of biomass-based energy systems: A case study of the state of maharashtra, india. *Clean Technologies and Environmental Policy*.

Vikash, P. V. and Shastri, Y. (2019). Conceptual design of a lignocellulosic biorefinery and its supply chain for ethanol production in India. *Computers & Chemical Engineering*, 121:696–721.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50341-9

LCA modelling as a decision-tool for experimental design: the case of extraction of astaxanthin from crab waste

Carina L. Gargalo[a*], Liliana A. Rodrigues[b,c], Alexandre Paiva[c], Krist V. Gernaey[a], Ana Carvalho[c]

[a] *Process and Systems Engineering Centre (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark*
[b]*iBET - Instituto de Biologia Experimental e Tecnológica, Oeiras, Portugal*
[c]*NOVA School of Science and Technology, Universidade NOVA de Lisboa, Caparica, Portugal*
[c]*CEG-IST, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais, 1, 1049-001 Lisboa, Portugal*
carlour@kt.dtu.dk

Abstract

The worldwide consumption of crustaceans, mainly crabs and shrimp, has increased significantly over the last decades. Noteworthy is that this waste inherently contains high-value-added compounds, such as astaxanthin. Hence, it is economic- and environmentally beneficial to extract astaxanthin effectively and sustainably from seafood production wastes and thrive towards a circular economy.
Therefore, the objective of this work is three-fold: (i) to propose a novel integrated process for the extraction of astaxanthin-oleoresin (AXT-oleoresin) from crab shell wastes; (ii) to assess the environmental performance of this new process at an industrial scale in Portugal; and, (iii) to identify latent production and environmental bottlenecks as well as provide suggestions for process and/or model re-design.
A process environmental and human health impact must become an essential consideration when designing new processes. Henceforth, aiming at a more sustainable process development and production, by applying LCA at the early stage of process development, we aim to identify critical issues early on, then to re-design, and reassess. Hence, this study concludes that the side-by-side application of LCA modelling and experimental process development is the best proactive approach for designing new processes and strategies as early as at the experimental level.
Keywords: LCA, astaxanthin, oleoresin, circular economy

1. Introduction

More than 6 million tons of crustacean shell waste is produced worldwide per year (Rodrigues et al., 2020). Astaxanthin (AXT), a lipid-soluble carotenoid from the xanthophyll family, is among other high-value compounds present in this waste. It has the leading antioxidant activity compared to other antioxidants (e.g., lycopene, vitamins E and A). Other natural sources of AXT are *Haematococcus pluvialis* (highest yield) and the yeast *Xanthophyllomyces dendrorhous* (Rodrigues et al., 2020; Sajna et al., 2015). Although there are many challenges when competing with synthetic AXT production,

natural AXT has been approved as food coloring agent and to be used in the formulation of cosmetics and nutraceuticals (Rodrigues et al., 2020; Sajna et al., 2015).
Therefore, to transition to a circular economy emphasizing the conversion of waste into value, in this work, we propose a novel process for the extraction and production of astaxanthin-oleoresin (AXT-oleoresin) from crab shell wastes. Astaxanthin (AXT) is isolated from the shells using a menthol:myristic acid-based deep eutectic solvent as an alternative to conventional approaches. This leads to a bioactive extract, AXT-oleoresin, with combined solute and solvent properties that can find potential applications as a functional ingredient and as a natural preservative in the pharmaceutical, nutraceutical, or cosmetic/personal care industries (Rodrigues et al., 2020). Furthermore, to ensure sustainable progress, LCA is applied at the early stage of process development to identify critical issues early on, then to re-design, and reassess.

2. Methodology

The objective of this work is three-fold: (i) to propose a novel integrated process for the extraction and production of astaxanthin-oleoresin (AXT-oleoresin) from crab shell wastes; (ii) to assess the environmental performance of this new process at an industrial scale in Portugal; and, (iii) to identify latent production and environmental bottlenecks as well as provide recommendations for process and/or model re-design. To this end, the methodology illustrated in Figure 1 is implemented, which implies applying the LCA framework in a stepwise manner.

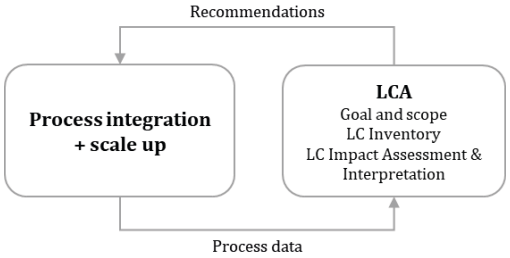


Figure 1: Methodology outline.

The LCA was performed following the ISO 14040 and 14044 guidelines (*ISO* 2006a,b). SimaPro vs.9.1 (Pré Consultants 2020) was the software used. Consequential modelling was used in the inventory analysis as described in (Weidema et al., 2009).

2.1. Goal and scope

As previously stated, the goal of the LCA study is two-fold; (i) to assess the environmental performance of the novel AXT-oleoresin production process in Portugal at industrial scale; and, (ii) to identify latent production/environmental bottlenecks and provide suggestions for process re-design. Figure 2 shows the product system for the overall production of AXT-oleoresin. LCA is performed with cradle-to-gate boundaries. The functional unit requires an additional 1 kg of AXT-oleoresin at the production facility's gate. It is important to note that the cut-off criteria are applied to the crab shell wastes according to the ILCD handbook (JRC-IES, 2010): the environmental impacts related to the production of crab shells are not included in the model. It is assumed that the crab fishery industry is responsible and accounts for all impacts related to the production and transport of crab shell wastes.
Furthermore, as presented in Figure 2, two critical aspects of consequential modelling have been employed: (a) the marginal use of crab shell wastes is identified as being

composting (Muñoz et al., 2018) (identification of marginal suppliers), and (b) the process by-product rich in minerals is assumed to substitute mineral fertilizer in the market (product substitution).
The life cycle impact assessment method chosen is the one identified by the International Life Cycle Data (ILCD) handbook, at the midpoint level.

2.1.1. Production process: production of astaxanthin-oleoresin (AXT-oleoresin)

The first step involves the drying and milling of the raw materials since the efficiency of the subsequent extraction process may be influenced by the moisture content and/or particle size. The dried and milled residue is then directed to the extraction step, in which AXT is extracted from the shells using a menthol:myristic acid (ME:MA,8:1) deep eutectic solvent (DES). The AXT-rich extract is separated from the solid residue by filtration, which leads to the final product and a by-product stream. The final product is the AXT-oleoresin, which is composed of DES and AXT-rich extract. The by-product stream is rich in minerals and proteins.
Due to the novelty of the proposed production process, it has only been developed and tested at the lab scale. Therefore, in this work, the experimental results have been appropriately scaled up (Piccinno et al., 2016) in order to reflect a credible industrial scale (see Section 3).

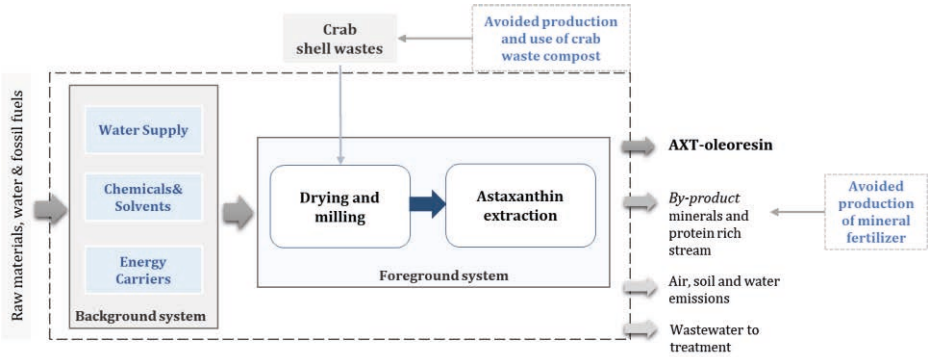


Figure 2: Product system for the production of AXT-oleoresin. System boundaries: cradle-to-gate.

2.2. Life Cycle Inventory

We believe a credible scenario to be the production of AXT-oleoresin that fulfills 50% of the forecasted amount of natural AXT to be consumed in 2021 in Europe (corresponding to approx. 6.9 tons) ("Global Natural Astaxanthin Market insights"). We assume that the natural AXT is produced/consumed in the form of AXT-oleoresin. Therefore, the hypothetical plant's production capacity is approximately 6.9 tons of AXT-oleoresin. The foreground data was obtained by applying appropriate scale-up factors (Piccinno et al., 2016) to the experimental results of the process described in Section 2.1. As previously stated, the tasks in the background system were modelled with the consequential version of the ecoinvent database v.3.1 (ecoinvent Centre 2016). The life cycle inventory and details concerning primary data sources and assumptions are presented in Table 1.

Table 1: Life cycle inventory of the industrial production of AXT-oleoresin from crab shell waste.

Exchanges	Unit	Amount	LCI data
Output of products/services:			
AXT-oleoresin	ton	6.9	Reference flow. Yearly production of AXT-oleoresin (AXT + ME:MA)
Avoided products:			
Mineral Fertilizer	ton	1.85	Displaced production of mineral fertilizer: disrupted production of mineral fertilizer through traditional routes. Composition approx. Minerals (65%), Protein (21%), Chitin (11%). Ecoinvent dataset: Lime fertilizer, from sugar production, at plant/ES Mass Assumption: it is assumed that the by-product stream is directly used and replaces mineral fertilizer.
Brown crab waste composting and soil application	ton	3.73	Avoided composting of crab shell wastes and use of compost. Amount of crab shell waste used. Model developed in this work based on brown crab shell composition and the modelling strategy proposed in (Muñoz et al., 2018) and references within.
Input of products/services:			
Crab shell waste	ton	3.73	Inflow of waste. Applied cut-off criterion, no environmental impacts associated.
Menthol (ME)	ton	6.25	Ecoinvent dataset: Cyclohexanol {RER}| market for cyclohexanol | Conseq, U Assumption: production of cyclohexanol used as proxy for the production of Menthol. (Kendall et al., 2011a)
Myristic acid (MA)	ton	1.14	Ecoinvent dataset: Crude palm kernel oil (incl. LUC incl. peat emissions), at producer/GLO Assumption: production of myristic acid from palm kernel oil. 1.14 ton of MA result in 8.90 ton of palm kernel oil. Based on the content of fatty acids in the oil (82%) and the content of myristic acid within the fatty acids (15.7%). (Tambun et al., 2019)
Heating energy	MJ	6.76E+03	Ecoinvent dataset: Heat, from steam, in chemical industry {RER}| market for heat, from steam in chemical industry | Conseq, U
Electricity	MWh	3.49E-02	Ecoinvent dataset: Electricity, medium voltage {PT}| market for | Conseq, U Assumption: production plant to be located in Portugal.
Emissions to air: direct emissions from composting and use of the compost.			
Wastewater treatment:	ton	1.88	Ecoinvent dataset: Wastewater, average (waste treatment) {RoW}| treatment of, capacity 5E9l/year | Conseq, U

3. Results & Discussion

Figure 3 shows the life cycle impact assessment results (LCIA) and the relative contribution of the different activities in producing AXT-oleoresin in Portugal, from cradle-to-gate. To simplify the interpretation of results, only nine categories are selected out of sixteen reported in ILCD. Due to space constraints, the numerical values are omitted here. The LCIA shows credits (savings, negatives values in Figure 3) for several impact categories. This is especially noticeable in the categories: Acidification (AC), Terrestrial eutrophication (EUTT), Marine eutrophication (EUTM), and water use/depletion (WD). The water savings are associated with the avoided traditional mineral fertilizer production (water-intensive) since mineral fertilizer (minerals-rich side stream) is a by-product of AXT-oleoresin production. The savings in AC, EUTT, and EUTM are

mostly related to the fact that, as previously mentioned, the crab shell wastes are being diverted from the composting activity and subsequent use of compost. The use of compost usually leaves behind excessive amounts of nutrients in the soil (e.g., nitrogen and phosphorus), which can be leached into lakes, streams, and coastal waters and thus lead to a gradual increase in nutrient concentration (terrestrial and marine eutrophication). Additionally, commonly the same nutrients also cause soil acidification by changing the soil pH levels. Moreover, there are also some minor credits in the Climate Change (CC), Particulate Matter (PM), and Photochemical ozone formation (POF) categories of impact due to avoided air emissions of small amounts of, for example, nitrous oxide gas.

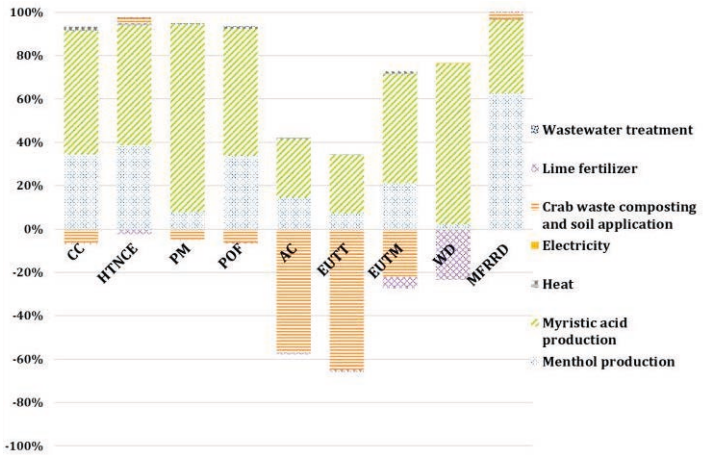


Figure 3: Impact assessment results for extraction and production of AXT-oleoresin from crab shell wastes in Portugal by activity (characterization values, per kg of AXT-oleoresin). Climate change (CC), Human toxicity, non-cancer effects (HTNCE), Particulate matter (PM), Photochemical ozone formation (POF), Acidification (AC), Terrestrial eutrophication (EUTT), Marine eutrophication (EUTM), Water resource depletion (WD) and Mineral, Fossil & Ren resource depletion (MFRRD).

Figure 3 also illustrates the effect of other activities in the LCIA results. The production and use of menthol (ME) and myristic acid (MA) play a crucial role; they are the most environmentally damaging activities in all impact categories. Thus, they are critical process hotspots. In this study, MA is produced from palm kernel oil, which affects/disrupts the palm oil kernel market, thus leading to increased extraction and associated environmental impacts. This is further aggravated by the low MA to oil yield. The production of cyclohexanol was taken as a proxy for ME production (Kendall et al., 2011b) due to the lack of information. This also implies a lack of accuracy and the propagation of uncertainty to the results. The production of cyclohexanol is based on the hydrogenation of benzene (resource intensive). Henceforth, this is in line with, for example, the results obtained for the Mineral, Fossil & Renewable resource depletion (MFRRD) category. However, it is important to note that the AXT-oleoresin production process's energy needs are somewhat irrelevant among the other contributors. Besides, the process does not consume freshwater. Therefore, this is a positive indication that if improving the critical process points (ME and MA models), the process has a good probability of improving its environmental performance. For example, the production of MA from nutmeg butter and coconut oil will be investigated (model improvement). Although both ME and MA have been identified as the process' critical points, replacing these solvents is not in question at this point. Both solvents were chosen due to their properties for producing a bioactive extract for the potential formulation of cosmetics and

nutraceuticals. Henceforth, to overcome the described process limitations and hotspots, the overall recommendations are to (i) test if smaller quantities of solvents will have the same properties for the formulation of the bioactive extract, (ii) have a dialogue with the solvent producer in order to find a greener source, and/or (iii) attempt to synthesize the solvents in-house and reassess the process's environmental performance.

4. Conclusions

To the best of our knowledge, we have presented the first model of the cradle-to-gate LCA for the novel production of natural AXT-oleoresin from crab shell wastes using DES. Primary data was obtained by scaling up experimental results. Furthermore, this study addresses process evaluation using consequential modelling principles and thus performs a prospective and proactive assessment rather than retroactive. The production of the DES components, ME and MA, was identified as process hotspots, which in fact translates into targets for model improvement. Noteworthy is that the energy and water consumption of the AXT-oleoresin production process seems to be immaterial when comparing to the remaining contributors. This is a reasonable indication that the process has a good environmental performance. As previously stated, replacing ME and MA is not an option at this point since both solvents were chosen due to their properties for producing a bioactive extract for the potential formulation of cosmetics/nutraceuticals. Therefore, the overall recommendation is to re-evaluate the solvent needs and its raw materials, potentially synthesize them in-house, and finally reassess the process's environmental performance. This study concludes that the side-by-side application of LCA modelling and experimental process development is the best proactive approach for designing new processes and strategies as early as at the experimental level.

References

Global Astaxanthin Market insights, forecast to 2025, 2019.

ISO 14040—environmental management—life cycle assessment—principles and framework, 2006. . Genève.

ISO 14040: Environmental management – Life cycle assessment – principles and framework, 2006. . Gnève.

JRC-IES, 2010. ILCD handbook—general guideline for life cycle assessment—detailed guidance. First edition. Ispra.

Kendall, A., Yuan, J., Brodt, S., Jan Kramer, K., 2011a. Carbon Footprint of U.S. Honey Production and Packing Report to the National Honey Board.

Kendall, A., Yuan, J., Brodt, S., Jan Kramer, K., 2011b. Carbon Footprint of U.S. Honey Production and Packing Report to the National Honey Board.

Muñoz, I., Rodríguez, C., Gillet, D., M. Moerschbacher, B., 2018. Life cycle assessment of chitosan production in India and Europe. Int. J. Life Cycle Assess. 23, 1151–1160.

Piccinno, F., Hischier, R., Seeger, S., Som, C., 2016. From laboratory to industrial scale: a scale-up framework for chemical processes in life cycle assessment studies. J. Clean. Prod. 135, 1085–1097.

Rodrigues, L.A., Pereira, C. V., Leonardo, I.C., Fernández, N., Gaspar, F.B., Silva, J.M., Reis, R.L., Duarte, A.R.C., Paiva, A., Matias, A.A., 2020. Terpene-Based Natural Deep Eutectic Systems as Efficient Solvents To Recover Astaxanthin from Brown Crab Shell Residues. ACS Sustain. Chem. Eng. 8, 2246–2259.

Sajna, K.V., Gottumukkala, L.D., Sukumaran, R.K., Pandey, A., 2015. White Biotechnology in Cosmetics. Ind. Biorefineries White Biotechnol. 607–652.

Tambun, R., Ferani, D.G., Afrina, A., Tambun, J.A.A., Tarigan, I.A.A., 2019. Fatty Acid Direct Production from Palm Kernel Oil. IOP Conf. Ser. Mater. Sci. Eng. 505.

Weidema, B.P., Ekvall, T., Heijungs, R., 2009. Guidelines for applications of deepened and broadened LCA. Deliverable D18 of work package 5 of the CALCAS project.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50342-0

Decomposition of Organic Compounds in Water from Oil Refineries

Shoma Kato[a], Yasuki Kansha[a*]

[a]*Organization for Programs on Environmental Sciences, Graduate School of Arts and Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902 JAPAN*

kansha@global.c.u-tokyo.ac.jp

Abstract

In oil refineries, water is used for many different purposes and a substantial amount of wastewater is generated. The traditional wastewater treatment process in oil refineries is comprised of primary and secondary treatment, and the treated water is discharged to water bodies. By recycling the treated water, water withdrawal from the environment can be reduced. Additionally, the organic sulfur, organic nitrogen, and other organic compounds may not be fully removed by biological treatment. For the treatment of these organics and for the recycling of wastewater from oil refineries, a wastewater treatment system for oil refineries has been proposed. The proposed process consists of 1) primary and secondary treatment 2) the removal of remaining organic sulfur and nitrogen using hydrogen, and 3) the removal of remaining organics by photocatalysis. To realize the proposed system, the removal of organic sulfur using hydrogen by simulation and photochemical reactions by experiments were studied. The removal of organic sulfur using hydrogen was analyzed using oil refinery plant data. The results showed that organic sulfur and nitrogen were removed in the process, and the conversion ratio highly depended on the chemical structure of the components. For the experiment, as a representative of the remaining organic compound in the wastewater from the oil refinery, phenol was chosen. In the experiment, the decomposition of phenol using UV and TiO_2 was investigated. The results showed that phenol was successfully decomposed. From the results of the investigations, this process shows promise to improve the treatment of industrial wastewater and contribute to the conservation of water resources.

Keywords: Water Treatment, Photocatalyst, Environmental Systems, Process Design

1. Introduction

Globally, there is a rise in the demand for freshwater resulting from the growing world population and industrialization. In fact, the global demand for water has been increasing by 1 % per year (UNESCO, 2020). For sustainable development and the well-being of the global ecosystem, it is necessary that water usage is reduced, that water is properly treated, and that water is recycled. Wastewater is generated from the agricultural sector, the industrial sector, and the domestic sector.

The contaminants present in industrial wastewater vary from industry to industry. In the case of the petroleum industry, most processes in oil refineries use water, so large amounts of wastewater are generated. For example, 0.60 – 0.71 L of water is used to produce 1 L of gasoline (Sun et al., 2018). To reduce the environmental impact of the discharge of wastewater and to reduce the water withdrawal of the petroleum industry, the treated wastewater can be recycled to use in the oil refinery or other industries. For

example, the Kalundborg Industrial Symbiosis in Denmark was developed to recycle the by-products in an industrial community, and the initial reason for the development was to reduce the usage of groundwater in oil refineries (WWAP, 2017).

The traditional treatment of wastewater from oil refineries includes primary treatment to remove the oils and sediments physically and secondary treatment to remove the organics in water biologically. Then, the water is discharged to water bodies. For the water to be recycled, additional treatment of water is necessary to remove the remaining contaminants after primary and secondary treatment of wastewater as the contaminants left in the treated water may affect the industrial processes such as fouling and degradation of catalysts. Additionally, the treatment process should be non-toxic and safe for the environment.

Advanced Oxidation Processes (AOPs) that generate strong oxidants such as OH radicals are effective to treat the refractory pollutants remaining in the water (Miklos et al., 2018). However, the chemical oxidation using AOPs is expensive, so they may be used in combination with biological treatment to reduce operation costs and for the effective treatment of contaminants in the wastewater (Oller et al., 2011).

In this research, a process was proposed to enhance the recycling of the water generated from oil refineries. In the proposed process, the remaining organic sulfur and nitrogen are removed after the biological treatment of water. Then, the remaining organic is removed by photocatalysis which is an AOP with the generation of OH radicals. The validity of the process was verified through simulation and experiments.

2. Proposed process for the treatment of water from oil refineries

Figure 1 shows the proposed process to recycle the water generated from the oil refineries. The proposed process aims to recycle the water in the oil refinery to reduce water withdrawal from the environment.

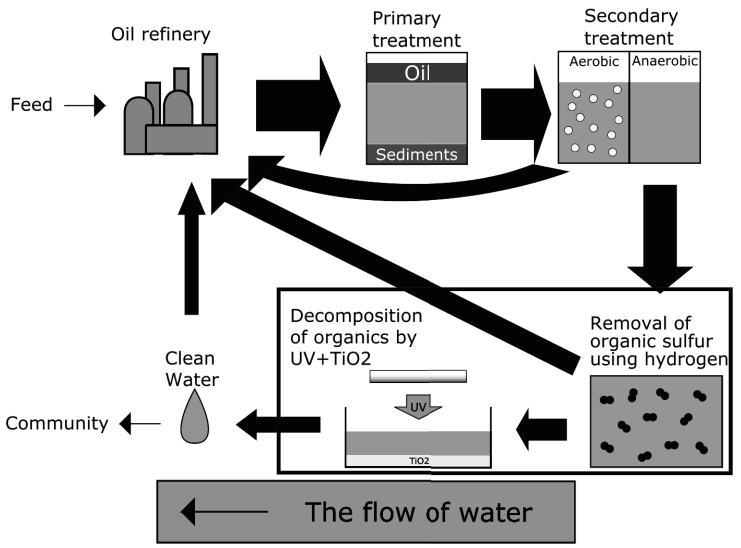


Figure 1: The proposed wastewater treatment system for the recycling of water from oil refineries is shown.

After primary treatment to remove the oils and sediments and secondary treatment to remove some of the organic compounds, organic sulfur, organic nitrogen, and other organic compounds may remain in the water. This is because wastewater from oil refineries contains a high concentration of refractory organics which are difficult to remove by biological treatment, and it may show higher COD values even after primary and secondary treatment (Diya'uddeen et al., 2011).

To treat the remaining contaminants in the water for recycling, two additional processes were added. The first process is to treat the remaining organic sulfur and organic nitrogen using hydrogen. In an oil refinery, desulfurization processes such as residue desulfurization (RDS) use hydrogen to upgrade residual fuel by removing contaminants, by converting them into lighter products, and by promoting hydrogenation (Marafi et al., 2006). Like the RDS process, the remaining organic sulfur and organic nitrogen may be removed using hydrogen. After the removal of organic sulfur and organic nitrogen, the remaining contaminants are further removed through photocatalysis with TiO_2 as the catalyst. TiO_2 was chosen as the catalyst for its ability to oxidize organic pollutants, chemical stability, and nontoxicity (Nakata and Fujishima, 2012). Depending on the necessary quality of the water needed for the industrial processes, the water can be recycled back to the oil refinery instead of going through the full cycle. Additionally, the process is safe for the environment as it does not require the use of flocculants or other chemical agents. The energy usage of the process can be further reduced by using sunlight as the light source for the photocatalyst process.

3. Methods

The removal of organic sulfur, organic nitrogen, and other organic compounds was explored through simulation. The treatment of the remaining organic contaminants was evaluated through experiments using phenol as a representative and TiO_2 as the photocatalyst.

3.1. The removal of organic sulfur, organic nitrogen, and other organic compounds using hydrogen

The data of the components of the feed and product data from the RDS process was used to calculate the overall conversion ratio of the organic sulfur and nitrogen. The conversion ratio was calculated as shown in Eq.(1).

$$Conversion\ ratio = \frac{Product\ concentration}{Feed\ concentration} \quad (1)$$

By understanding the reaction in the process using hydrogen, the effectiveness of the process can be evaluated, and the potential products from the reaction can be known.

3.2. The treatment of phenol through photocatalysis

Phenol (purity 99 %, FUJIFILM Wako Pure Chemical Corporation) was used as the representative for the organic compound. It was dissolved in Milli-Q water so that the concentration was 50 mg/L. Aeroxide TiO_2 P 25 (specific surface area 53 m^2/g, NIPPON AEROSIL CO., LTD.) was used as the photocatalyst, and it was added to the phenol solution so that the TiO_2 dosage was 1 g/L. A 300 mL Pyrex beaker (diameter 77 mm and height 110 mm) was used as the vessel. The vessel was filled with 200 mL of the mixture of TiO_2 and phenol solution. After filling the vessel, it was covered with a quartz lid (thickness 3 mm). A 4 W UV lamp with the main wavelength at 365 nm (model LUV-4,

AS ONE Corporation) was placed directly on top of the quartz lid. The experiments were conducted at room temperature.

The samples were taken from the vessel using syringes, and they were filtered using PES syringe filters of pore size 0.1 μm to remove TiO_2 particles. The absorbance of the samples was measured with a UV-vis spectrophotometer (model ASUV-1100, AS ONE Corporation) at 270 nm which is the absorbance peak of phenol. The phenol removal was calculated by the following equation:

$$Phenol\ removal\ (\%) = \frac{C_0 - C_t}{C_0} \times 100 \quad (2)$$

where C_0 and C_t are the concentrations of phenol at time 0 and at time t respectively.

4. Results and Discussion

4.1. The overall conversion ratio of the organic sulfur and organic nitrogen

Table 1 shows the conversion ratio by using hydrogen for organic sulfur and organic nitrogen present in the feed and product data. Depending on the structure of the component, the conversion ratio varies. Oil B was a heavier oil compared to Oil A, and the conversion ratio for Oil B was less than that of Oil A. In both oils, there is remaining organic sulfur and organic nitrogen which needs further treatment. Additionally, the process consumes a large amount of hydrogen. Further investigation is needed to increase the efficiency of this process to reduce hydrogen consumption.

Table 1: The conversion ratio by hydrogen treatment is shown.

Name	Conversion ratio of molecule with at least one sulfur atom	Conversion ratio of molecule with at least one nitrogen atom
Oil A	0.655	0.141
Oil B	0.575	0.084

4.2. The removal of phenol by photocatalysis

Figure 2 shows the comparison of the removal of using UV only versus using UV and TiO_2 as the photocatalyst. For UV photolysis alone, only 0.14 % of phenol was removed after 180 min. The result that phenol is difficult to degrade through direct photolysis is consistent with that of Lin et al. (2011). For UV+TiO_2 photocatalysis, 8.4 % of the phenol was removed after 180 min. With the presence of the TiO_2 photocatalyst, phenol which is the representative organic can be removed. However, the rate of removal is slow under the experimental conditions. The removal rate may be improved by increasing the light intensity, by optimizing TiO_2 dosage.

For the photocatalytic process, it is necessary to consider the recovery of the TiO_2 catalyst. One way is to use a TiO_2 photocatalytic membrane reactor which combines photocatalytic oxidation and membrane filtration (Leong et al., 2014). In this way, the TiO_2 can be recovered. Additionally, sunlight may be used as the light source to reduce energy consumption.

Figure 3 shows the possible mechanism of the removal of the organic compounds using the TiO_2 membrane. By irradiating the TiO_2 membrane with UV light, water may be converted into H^+ ions and OH radicals (Sobczyński et al., 2004). Additionally, through

a series of reactions, the organic compounds are completely mineralized into water and CO_2.

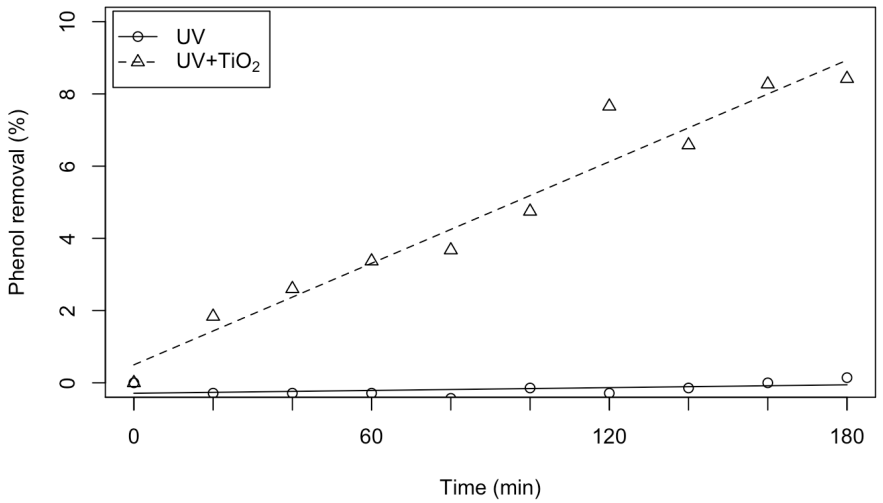


Figure 2: Comparison of the phenol removal using UV photolysis only versus UV+TiO_2 photocatalysis.

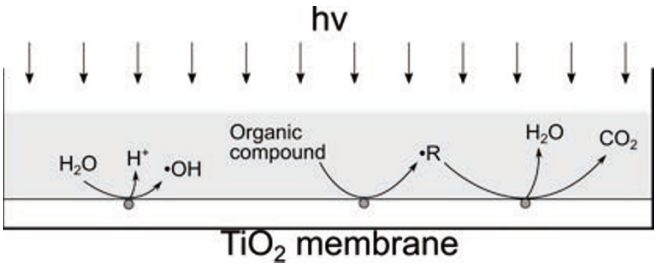


Figure 3: The schematic of photocatalytic treatment of organic contaminants in wastewater.

5. Conclusion

A system was proposed to recycle the wastewater from an oil refinery. The proposed process consists of 1) primary treatment, 2) secondary treatment, 3) treatment of remaining organic nitrogen, organic sulfur, and other organics by hydrogen, and 4) treatment of the remaining organics by photocatalysis with TiO_2 as the catalyst. To evaluate the treatment using hydrogen, simulation was conducted. The results showed that although the process can treat some of the contaminants, further treatment is necessary to remove the remaining contaminants. To evaluate the treatment using photocatalysis, experiments were conducted. The results showed that phenol, a representative of a refractory organic can be removed. Depending on the demanded quality of the water, the water may be recycled back to the oil refinery and the industrial community to reduce the water withdrawal from the environment. The proposed process is promising to reduce the amount of water withdrawal, and it is a green process that does not require the use of toxic chemicals.

References

Diya'uddeen B.H., Daud W.M.A.W., Abdul Aziz A.R., 2011, Treatment technologies for petroleum refinery effluents: A review, Process Safety and Environmental Protection, 89 (2), 95–105.

Leong S., Razmjou A., Wang K., Hapgood K., Zhang X., Wang H., 2014, TiO_2 based photocatalytic membranes: A review, Journal of Membrane Science, 472, 167–184.

Lin S.H., Chiou C.H., Chang C.K., Juang R.S., 2011, Photocatalytic degradation of phenol on different phases of TiO_2 particles in aqueous suspensions under UV irradiation, Journal of Environmental Management, 92 (12), 3098–3104.

Marafi A., Hauser A., Stanislaus A., 2006, Atmospheric Residue Desulfurization Process for Residual Oil Upgrading: An Investigation of the Effect of Catalyst Type and Operating Severity on Product Oil Quality, Energy & Fuels, 20 (3), 1145–1149.

Miklos D.B., Remy C., Jekel M., Linden K.G., Drewes J.E., Hübner U., 2018, Evaluation of advanced oxidation processes for water and wastewater treatment – A critical review, Water Research, 139, 118–131.

Nakata K., Fujishima A., 2012, TiO_2 photocatalysis: Design and applications, Journal of Photochemistry and Photobiology C: Photochemistry Reviews, 13 (3), 169–189.

Oller I., Malato S., Sánchez-Pérez J.A., 2011, Combination of Advanced Oxidation Processes and biological treatments for wastewater decontamination—A review, Science of The Total Environment, 409 (20), 4141–4166.

Sobczyński A., Duczmal Ł., Zmudziński W., 2004, Phenol destruction by photocatalysis on TiO_2: an attempt to solve the reaction mechanism, Journal of Molecular Catalysis A: Chemical, 213 (2), 225–230.

Sun P., Elgowainy A., Wang M., Han J., Henderson R.J., 2018, Estimation of U.S. refinery water consumption and allocation to refinery products, Fuel, 221, 542–557.

UNESCO, UN-Water, 2020, *United Nations World Water Development Report 2020: Water and Climate Change*, UNESCO, Paris, France.

WWAP (United Nations World Water Assessment Programme), 2017, *The United Nations World Water Development Report 2017. Wastewater: The Untapped Resource*, UNESCO, Paris France.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50343-2

Energy Harvesting Wireless Sensors Using Magnetic Phase Transition

Yasuki Kansha[a], Masanori Ishizuka[b]

[a]*Oraganization for Programs on Environmental Sciences, Graduate School of Arts and Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro-ku Tokyo 153-8902, Japan*
[b]*Collaborative Research Center for Energy Engineering, Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan*
kansha@global.c.u-tokyo.ac.jp

Abstract

In this research we proposed energy harvesting data acquisition sensors using the magnetic phase transition resulting from changes in temperature, and electromagnetic induction resulting from changes in magnetic flux. The proposed system can provide wireless temperature or velocity sensors that directly measure electromotive forces generated by a solenoid following Faraday's law without any additional energy input. Our proposed energy harvesting sensors have the potential to contribute significantly to the development of CPS in the near future.

Keywords: Wireless sensors; Energy harvesting; Cyber-physical systems

1. Introduction

The Japanese government has proposed ‘Society 5.0’, under which, not only industry, but society itself will be changed by information and communication technology (ICT) and the internet of things (IoT) to allow sustainable development. In the energy related field, dig data such as climate and environmental information and energy usage in communities will be acquired and analysed by Artificial Intelligence (AI). Furthermore, energy will supply following the analysed data for providing a stable energy supply etc. (Cabinet Office in Japan).

To deploy cyber-physical systems (CPS) and the IoT in society, it is necessary to develop overall security systems (Alguliyev et al., 2018), efficient information and communication technologies (ICT) including data transfer systems in the network, and acquisition systems, such as sensors and actuators (Patil and Fiems, 2018). It is also important to find new energy sources to connect and transfer digital data to the cloud and to develop intelligent decision/control systems (Shen et al., 2019) and learning algorithms (Zhang et al., 2018) for rational operation of the overall networks, incorporating the use of artificial intelligence.

In meeting the requirements of energy for CPS, the term ‘energy harvesting’ is commonly used. Energy harvesting involves electric power generation for online operation of sensor and electric devices from currently unused low level energies such as vibrations, radio frequencies, light and low temperature heat, into electricity. Piezoelectric elements and antennae are common devices used for converting vibrations or radio frequencies into electricity (Wang et al., 2018). Photovoltaics is also a familiar way of converting light into electricity to supply CPS (Gunduz and Jayaweera, 2018). Thermoelectric elements based on the Seebeck effect are also commonly used to convert heat into electricity (Ando

Junior et al., 2018). However, the energy efficiency of many of these energy harvesting technologies is still low, and much research is focused on increasing their efficiency. In fact, the thermos-electric device generate the electric power by p- and n-type semiconductors. However, it is well-known that the efficiency of this device is so low due to \small entropy change by electron or hole transfer (Chen and William, 1996)

On the contrary, much recent research has focused on the possibility of combining energy harvesting and sensors, including wearable sensors for human information (Myers et al., 2017) and wireless sensors for environmental information (Babayo et al., 2017). These sensors called energy harvesting sensors directly sense the measured target without any additional energy conversion, leading to increase overall energy efficiency of CPS. Thus, the energy harvesting sensors are expected as a key technology for propagation of a CPS (Kausar et al., 2014).

2. Energy harvesting using magnetic phase transition and sensors

Kansha et al. (2018) proposed an energy harvesting system from sub-ambient heat sources, such as exhaust heat from refrigerators or coolers, that uses magnetic phase transition integrated with electromagnetic induction. In this system, a magnetic material such as gadolinium is cooled by a sub-ambient temperature heat to below the Curie temperature (292 K for gadolinium). The material becomes ferromagnetic. It is placed near a solenoid as shown in Figure 1 a) and heated by a high temperature heat to its Curie temperature. The material is demagnetized, becoming paramagnetic, and the magnetic flux from the material that passes through the solenoid decreases as the temperature increases. An electromotive force is generated in the solenoid following Faraday's law of induction;

$$\varepsilon = -N\frac{d\Phi}{dt} \tag{1}$$

where ε is the electromotive force, N is the number of turns on the solenoid, Φ represents the magnetic flux, and t is time.

Figure 1 b) shows a schematic image of the thermodynamic cycle in temperature-entropy diagram of the series of action. The theoretical thermodynamic cycle of this system transitions to a closed trilateral cycle suitable for sensible heat recovery to generate electric power.

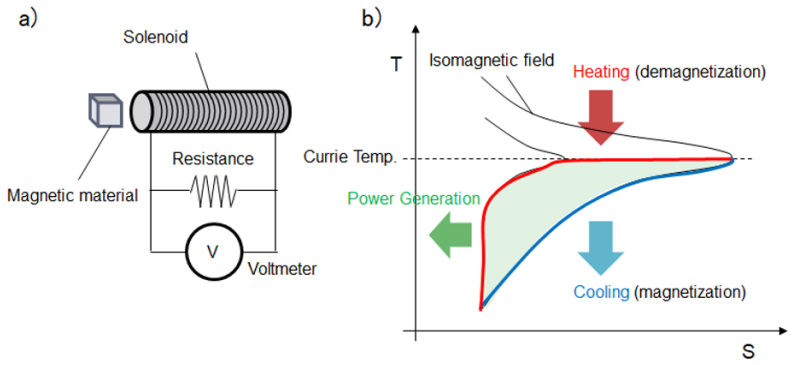


Figure 1: Thermodynamic cycle of the proposed power generation system.

A study of the adiabatic temperature change due to the magnetic phase transition during magnetization of gadolinium from 0 T to 1 T indicates that the change is almost linear from 270 K to the Curie temperature (292 K) and from the Curie temperature to 320 K at a peak of 292 K as shown in Figure 2. Furthermore, the adiabatic efficiency of magnetization/demagnetization changes of gadolinium was examined about 0.92 (Kotani et al., 2013). From these aspects, the change in entropy at the magnetic phase transition around the Curie temperature may have a linear relationship with the temperature.

Thus, observing the flux changes for magnetic materials at different temperatures could allow it to be used as a temperature sensor integrated with energy harvesting around the Curie temperature, without requiring any additional energy instead of thermocouples.

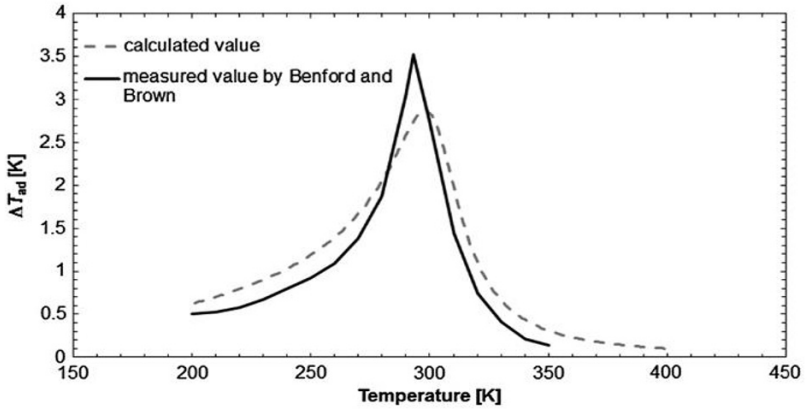


Figure 2: Adiabatic temperature change when gadolinium is magnetized from 0 T to 1 T (Kotani et al., 2013).

3. Experimental procedure and set-up

To examine the possibility of energy harvesting sensors, the relationship between the temperature of a magnetic material and the change in its magnetic flux was investigated by measuring the electromotive force. Gadolinium was selected as the magnetic material because of its Curie temperature.

Before measuring the electromotive forces, a gadolinium nugget (2.73 g) was left at each temperature for more than 30 mins until the nugget temperature became constant. The maximum magnetic flux, which the gadolinium nugget creates, was about 25 mT measured by a gauss meter (HMMT-6J04-VF, Lake Shore Cryotronics Inc.).

Using the following experimental set-up shown in Figure 3, the electromotive forces generated by Faraday's law of induction were monitored at four different temperatures (256, 280, 292, 296 K) by an oscilloscope (InfiniiVision DSO-X 2002A, Agilent Technologies Inc.). A permanent magnet (275 mT) was positioned at the edge of a solenoid with 500 coils of iron wire.

Using the gauss meter, the magnetic flux density at the other end of the solenoid, the nearest point to the gadolinium, was 19 mT. The gadolinium nugget, which had a magnetic flux density of 0 mT without the magnetic field, was fixed to the end of a 40-mm arm, which was rotated horizontally by a motor (rotation speed: 545 degree/s). The

gadolinium was passed over the solenoid. The minimum distance from the gadolinium to the solenoid was 3 mm. Thus, the magnetic flux through the solenoid changed with the position of the gadolinium. It is noted that the magnetic flux was changed by heat transfer to the magnetic materials to use the proposed system as an ambient temperature sensor. However, this change might be too sensitive to examine in the experiments. Therefore, the gadolinium was forced to move by motor for changing the distance to the solenoid at the constant temperature in order to sense the targeted temperature in this experiment. The electromotive force produced by electromagnetic induction and the current were measured by the oscilloscope with a shunt resistance of 4.7 Ω.

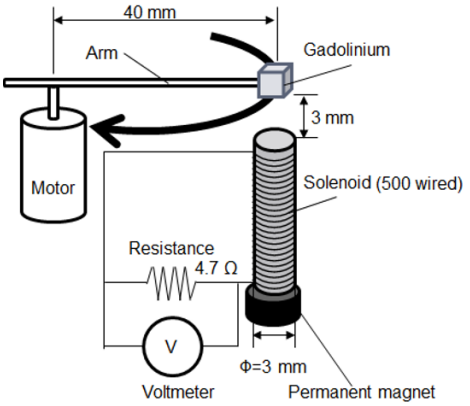


Figure 3: Adiabatic temperature change when gadolinium is magnetized from 0 T to 1 T.

4. Experimental results

The large electromotive forces generated are shown in Figure 4 at the Curie temperature (292 K). The grey trace shows the raw data and the black line is the 10-ms moving average. The figures show positive and negative peaks. A positive peak was created when the gadolinium came close to the solenoid and the negative peak was created when the gadolinium passed. To determine the relationship between temperature and the generated electromotive force, the amplitude of the electromotive force was measured. Table 1 lists the amplitude of the electromotive force generated as a moving average at each temperature. From this table, it can be understood that the maximum power output is distributed following the temperature. Thus, the proposed system has a possibility to use it as a temperature sensor with calibration.

Table 1. Relationship between temperature and amplitude of electromotive forces.

Temperature [K]	256	280	292	296
minimum value of electromotive force [mV]	-0.182	-0.466	-0.505	-0.228
maximum value of electromotive force [mV]	0.023	0.239	0.267	-0.018
amplitude of electromotive force[1)] [mV]	0.102	0.352	0.386	0.105
maximum power ouput[2)] [nW]	1.8	28.2	34.7	2.4

[1)] amptitude = (maximum - minimum)/2

[2)] power output = V^2/R

Figure 5 shows a comparison of the 10-ms moving averages of the electromotive forces generated for several motor rotation speeds (182, 363, and 545 degree/s) at 292 K; the

amplitudes were 0.188, 0.309, and 0.386 and the intervals between the peaks (53.5, 40.5 and 31.5 ms) changed following the rotational speed of the gadolinium. The amplitudes are linearly increase with rotation speeds. Therefore, it can be worked as a velocity sensor.

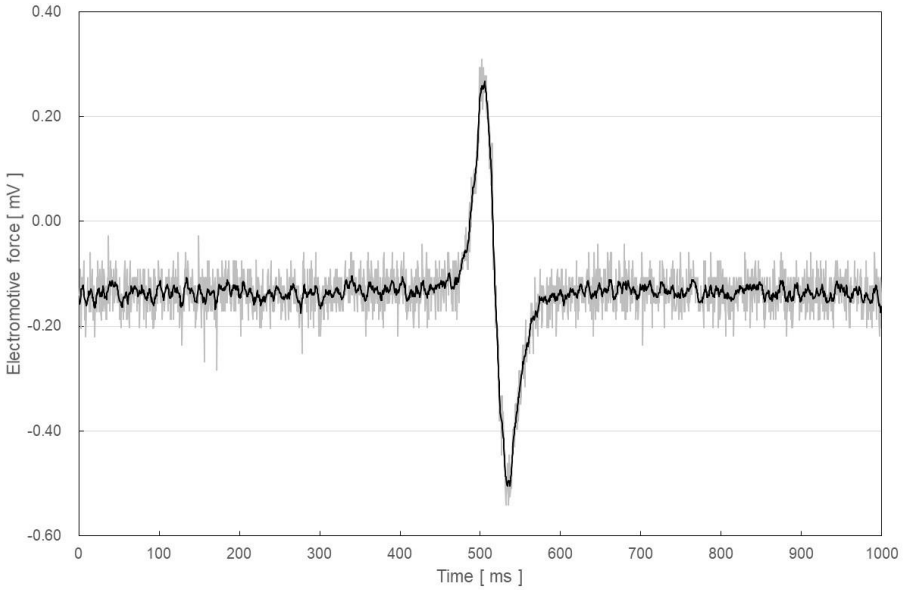


Figure 4: Generated electromotive force at 292 K during one cycle.

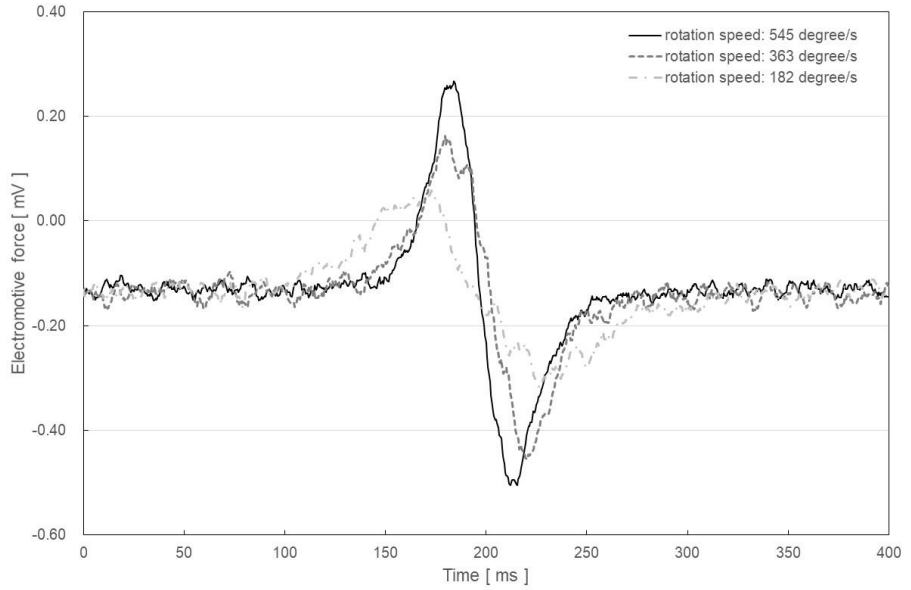


Figure 5: Moving average of generated electromotive force at 292 K by different rotation speeds.

5. Conclusion

This paper proposes designs for energy harvesting temperature and velocity sensors. By integrating magnetic phase transition with electromagnetic induction, these sensors can sense without needing any additional energy, leading to an increase in the energy efficiency of CPS. Furthermore, as the sensor itself is not attached to a receiver, it functions as a wireless sensor. Our proposed energy harvesting sensors have the potential to contribute significantly to the development of CPS in the near future.

Acknowledgement

This work was supported by the Japan Society for the Promotion of Science, (KAKENHI Grant Number 16K14544 and 21H01868) and the Tonen General Sekiyu Research & Development Encouragement & Assistance Foundation.

References

R. Alguliyev, Y. Imamverdiyev, L. Sukhostat, 2018, Cyber-physical systems and their security issues, Comput. Ind., 100, 212–223.

O.H. Ando Junior, A.L.O. Maran, N.C. Henao, 2018, A review of the development and application of thermoelectric microgenerators for energy harvesting, Renew. Sust. Energ. Rev., 91, 376–393.

A.A. Babayo, M.H. Anisi, I. Sli, 2017, A review on energy management schemes in energy harvesting wireless sensor networks, Renew. Sust. Energ. Rev., 76, 1176–1184.

Cabnet office in Japan https://www8.cao.go.jp/cstp/english/society5_0/index.html (access 10/Nov./2021)

K. Chen, S.B. William, 1996, An analysis of the heat transfer rate and efficiency of TE (thermoelectric) cooling systems. Int. J. Energy Res., 20, 399–417.

H. Gunduz, D. Jayaweera, 2018, Reliability assessment of a power system with cyber-physical interactive operation of photovoltaic systems, Int. Electr. Power Energy Syst., 101, 371–384.

Y. Kansha, M. Ishizuka, 2018, Power generation from low temperature waste heat using magnetic phase transition, Chemical Engineering Transactions, 70, 43–48.

A.S.M.Z. Kausar, A.W. Rezan, M.U. Saleh, H. Ramiah, 2014, Energizing wireless sensor networks by energy harvesting systems: Scopes, challenges and approaches, Renew. Sust. Energ. Rev., 38, 973–989.

Y. Kotani, Y. Kansha, A. Tsutsumi, 2013, Conceptual design of an active magnetic regenerative heat circulator based on self-heat recuperation technology, Energy, 55, 127–133.

A. Myers, R. Hodges, J.S. Jur, 2017, Human and environmental analysis of wearable thermal energy harvesting, Energy Conv. Manag., 143, 218–226.

K. Patil, D. Fiems, 2018, The value of information in energy harvesting sensor networks, Oper. Res. Lett., 46, 362–366.

B. Shen, Z. Wang, D. Wang, J. Luo, H. Pu, Y. Peng, 2019, Finite-horizon filtering for a class of nonlinear time-delayed systems with an energy harvesting sensor, Automatica, 100, 144–152.

D.W. Wang J.I. Mo, X.F. Wang, H. Ouyang, Z.R. Zhou, 2018, Experimental and numerical investigations of the piezoelectric energy harvesting via friction-induced vibration, Energy Conv. Manag., 171, 1134–1149.

X. Zhang, T. Yu, Z. Xu, Z. Fan, 2018, A cyber-physical system with parallel learning for distributed energy management of a microgrid, Energy, 165, 205–221.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50344-4

Competitive Adsorption of Copper, Nickel, and Chromium Ions onto Amine Functionalized SBA-15

Bawornpong Pornchuti[a*], Yuttana Phoochahan[a], Prarana Padma[a], Suchada Ruengrit[a], Pravit Singtothong[b]

[a]*Chemical Engineering Department, Mahanakorn University of Technology, Bangkok 10530, Thailand*
[b]*Chemistry Department, Mahanakorn University of Technology, Bangkok 10530, Thailand*
bawornpong@yahoo.com

Abstract

Wastewater from electroplating industry is composed of heavy metals. Adsorption is an effective method treating this type of hazardous wastewater. Many works studied adsorption of single metal ion however industrial wastewater is always multicomponent system. As a consequence, this work aimed to study the competitive adsorption of copper, nickel, and chromium ions onto amine functionalized SBA-15. It was found that adsorption capacity increased with an increase of pH. Moreover, removal efficiency of metal ion decreased when initial metal concentration increased. To represent the adsorption data, Langmuir, Freundlich, extended Langmuir, and modified competitive Langmuir models were selected. Their SSEs (Sum of squared errors) were 98.58, 118.40, 165.84, and 156.18, respectively.

Keywords: competitive adsorption, heavy metals, multicomponent isotherms, SBA-15.

1. Introduction

Electroplating is one of finishing steps used to improve mechanical and chemical properties of products. It also gives an attractive surface. However, wastewater from electroplating process is always comprised of heavy metals such as copper, nickel, and chromium. These heavy metals are harmful to human and can accumulate in ecosystems. Adsorption is one of the promising techniques treating wastewater containing heavy metals. As a consequence, many researches aimed to develop novel adsorbents having high adsorption capacities.

SBA-15 was chosen in this work because it has large surface area and easy to modify its surface. If any functional groups containing N, O, S, or P atom is incorporated onto adsorbent surface, the removal efficiency of metal ions is improved (Maleki, 2016). Therefore adsorption of copper, nickel, and chromium ions onto amine functionalized SBA-15 was studied in this work.

Most works in literature reported the adsorption of single heavy metal ion while industrial wastewater is generally composed of several heavy metal ions. The adsorption of a component may be affected by the other component (Girish, 2017). Therefore, the study of multicomponent adsorption is necessary. Isotherm models are used to predict the adsorption behaviour. Langmuir, Freundlich, extended Langmuir, and modified competitive Langmuir isotherm models were chosen in this study.

2. Materials and methods

2.1. Materials

TEOS (Tetraethyl orthosilicate), Pluronic P123, toluene, ethanol, APTES (Aminopropyltriethoxysilane), and hydrochloric acid were purchased from Sigma Aldrich or Merck. They were used as received.

2.2. Synthesis of SBA-15

SBA-15 was synthesized by using the method of Naik et al. (2011). Briefly, 6.64 g of Pluronic P123, 13.5 mL of HCl solution, and 202 mL of deionized water was mixed together. Next, 13.86 g of TEOS was dropped into the mixture. After aging at 90 °C for 24 h, the solid obtained was filtered and washed with deionized water. Then, the solid was dried and calcined at 550 °C for 3 h.

2.3. Preparation of amine functionalized SBA-15

The method of Parida and Rath (2009) was used to graft amino functional group onto the surface of SBA-15. In brief, 2.0 g of SBA-15 was mixed with 60 mL of toluene. Next, 1.2 mL of APTES was dropped into the mixture. After reflux for 8 h, the solid was washed with ethanol and deionized water, respectively. Then it was dried at room temperature for 12 h.

2.4. Batch adsorption studies

In general, 100 mg of adsorbent was added into 50 mL of heavy metal solution. The mixture was shaken at 105 rpm for 48 h. After vacuum filtration, the metal concentration was determined by atomic absorption spectroscopy (AAS). The adsorption capacity was calculated by Eq. (1).

$$q_e = \frac{(C_i\text{-}C_e)V}{m} \tag{1}$$

q_e is the adsorption capacity at equilibrium; C_i is the initial metal concentration; C_e is the metal concentration at equilibrium; V is the volume of metal solution; m is the mass of adsorbent. For the calculation of removal efficiency, Eq. (2) was used.

$$\text{Removal efficiency} = \frac{(C_i\text{-}C_e)}{C_i} \times 100\ \% \tag{2}$$

3. Results and discussion

3.1. Effect of pH

To study the effect of pH, the experiments were conducted at pH 2 – 5. It was found that the amount of nickel ion adsorbed by amine functionalized SBA-15 was very low as shown in Figure 1. The nickel adsorption was suppressed by copper and chromium ions, suggesting antagonistic interaction. This effect was also found in the adsorption of nickel in the presence of copper when using olive stones as adsorbent (Girish, 2018). Moreover, adsorption of copper and chromium ions increased with an increase of pH. This may be caused by competition with hydrogen ion at low pH (Sertsing et al., 2018).

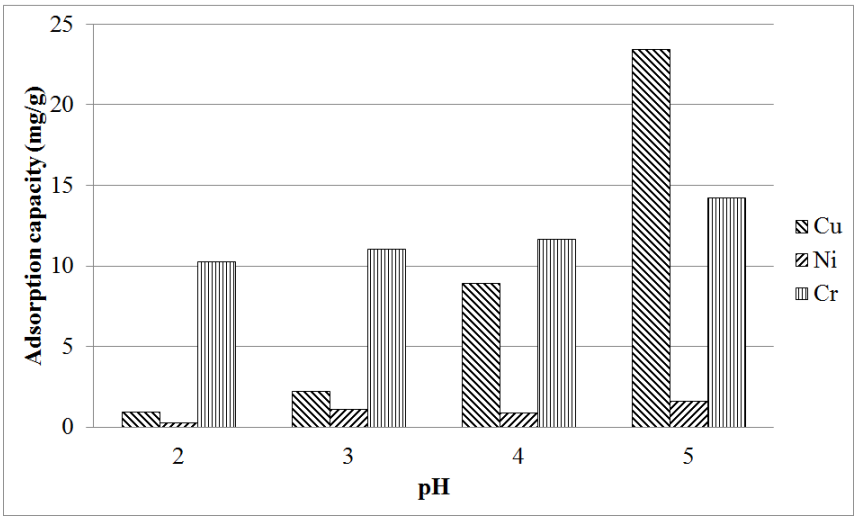


Figure 1. Effect of pH on adsorption capacity of copper, nickel, and chromium ions in multicomponent adsorption.

3.2. Effect of initial metal concentration

The effect of initial metal concentration was studied at pH 5. The results were illustrated as Figure 2. When initial metal concentration increased, the removal efficiency of metal ion decreased. This phenomenon was also found in single metal adsorption (Pornchuti et al., 2020). By fixing the amount of adsorbent, the number of adsorption site was limited. That was why the reduction of the removal efficiency occurred.

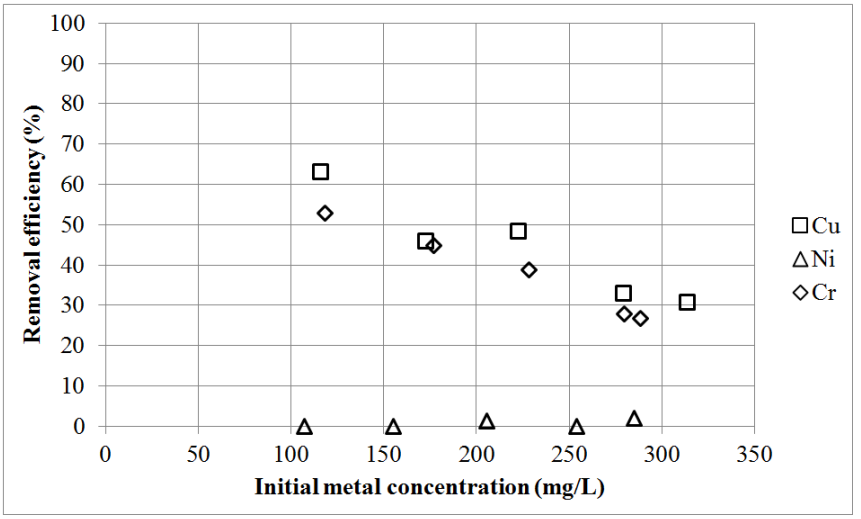


Figure 2. Effect of initial metal concentration on removal efficiency of metal ion in multicomponent adsorption.

To treat wastewater containing copper, nickel, and chromium ions by amine functionalized SBA-15, at least two adsorption columns are required. The first one is used to adsorb copper and chromium ions. The other one is used to remove nickel ion.

When copper and chromium ions are removed, nickel ion can be adsorbed in the second adsorption column.

3.3. Mathematical modeling

Langmuir and Freundlich isotherms were used widespread for adsorption of single component. There were some works using these models representing experimental data of multicomponent adsorption such as the work of Qi and Pichler (2017). Due to their simplicity, both models were chosen in this work. Langmuir isotherm is expressed as Eq. (3).

$$q_{e,i}=\frac{q_{m,i}b_{L,i}C_{e,i}}{1+b_{L,i}C_{e,i}} \tag{3}$$

$q_{e,i}$ is the adsorption capacity at equilibrium for component i; $q_{m,i}$ is the maximum adsorption capacity for component i; $b_{L,i}$ is the Langmuir constant for component i; $C_{e,i}$ is the metal concentration at equilibrium for component i. Freundlich isotherm is expressed as Eq. (4).

$$q_{e,i}=K_{F,i}C_{e,i}^{1/n_i} \tag{4}$$

$K_{F,i}$ is the Freundlich constant for component i; n_i is the adsorption intensity for component i. Isotherm parameters of both models can be determined from the linearized forms of each model and listed in Table 1. Since the amount of adsorbed nickel was very low, the experimental data of nickel was excluded from the models.

Table 1. Adsorption isotherm parameters.

Adsorption isotherm parameters	Cu	Cr
Langmuir isotherm		
$q_{m,i}$	50.2513	43.8597
$b_{L,i}$	0.0511	0.0407
Freundlich isotherm		
$K_{F,i}$	18.4204	16.7687
n_i	5.7438	6.2972
Extended Langmuir isotherm		
$q_{m,i}$	111.7629	76.2243
$b_{L,i}$	0.0249	0.0284
Modified competitive Langmuir isotherm		
$q_{m,i}$	78.1250	163.9344
$b_{L,i}$	0.0230	0.0268
$\eta_{L,i}$	1.0174	3.2165

The extended Langmuir isotherm is developed for multicomponent adsorption. This model can be represented as Eq. (5).

$$q_{e,i}=\frac{q_{m,i}b_{L,i}C_{e,i}}{1+\sum_{j=1}^{N}\left(b_{L,j}C_{e,j}\right)} \tag{5}$$

N is the total number of ions in the solution. The evaluation of parameters can be achieved by minimization of the error in non-linear regression analysis. The parameters of this model were listed in Table 1. In case of Modified competitive Langmuir isotherm, the interaction factor ($\eta_{L,i}$) is included to explain the competitive effect between the individual components. This model can be expressed as Eq. (6).

$$q_{e,i} = \frac{q_{m,i} b_{L,i} \left(C_{e,i} / \eta_{L,i} \right)}{1 + \sum_{j=1}^{N} \left(b_{L,j} \left(C_{e,j} / \eta_{L,j} \right) \right)} \tag{6}$$

All parameters in this model except the interaction factor, was determined from experimental data of single component adsorption. The interaction factor was evaluated from experimental data of multicomponent system by minimization of the error in non-linear regression analysis. The parameters were summarized in Table 1.

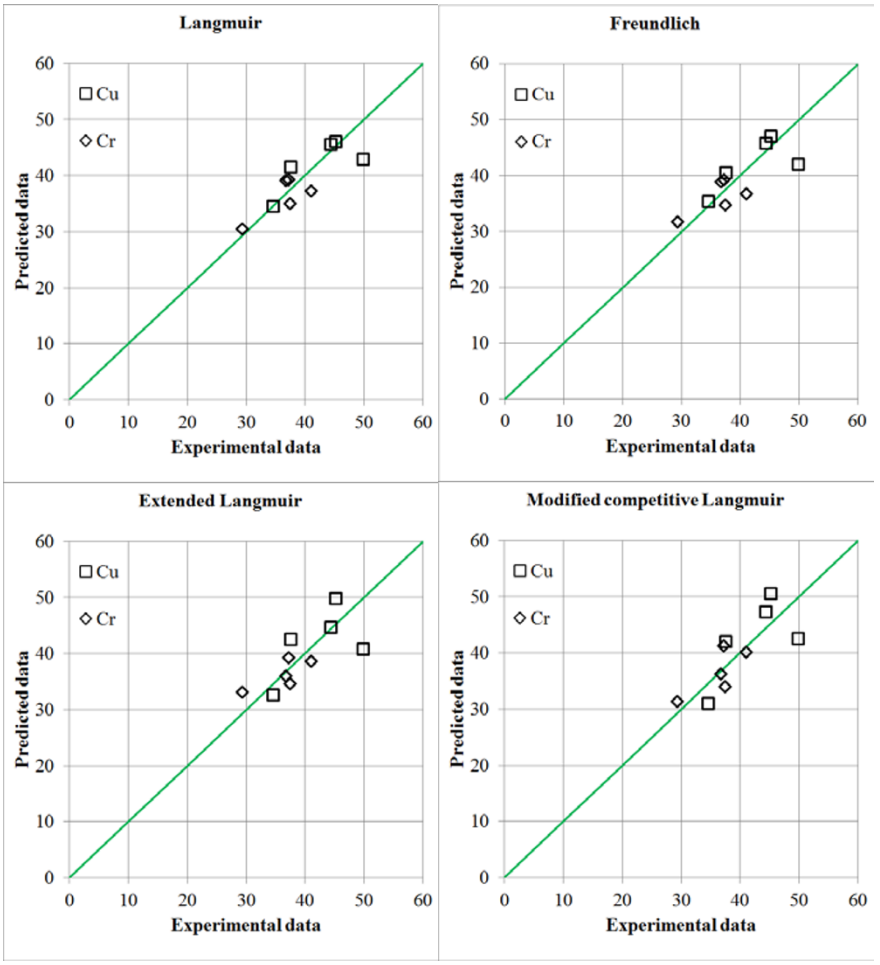


Figure 3. Comparison of experimental and predicted adsorption capacities.

The predicted data of adsorption capacity from each model were compared with the experimental data in Figure 3. The SSEs of Langmuir, Freundlich, extended Langmuir, and modified competitive Langmuir models were 98.58, 118.40, 165.84, and 156.18, respectively. Langmuir model was better than Freundlich model. Anyway, application of both models is limit. In addition, they do not include the effect of other components. When other metal concentrations were changed, the adsorption capacity was not altered as shown in Eq. (3) and (4). As a result, the adsorption behavior could not be described.

Since wastewater is always composed of various metal ions, the multicomponent models are useful for the design of adsorption column. The effect of other components is included in multicomponent adsorption models. Extended Langmuir isotherm and modified competitive Langmuir isotherm were developed for multicomponent adsorption. Modified competitive Langmuir model gave better result because it includes interaction factor which showed the competitive effect of metal ions (Girish, 2017).

4. Conclusions

Our study showed that there were interactions between metal ions during adsorption process. The adsorption affinities of metal ions depend on pH. The efficiency of metal removal decreased when initial metal concentration increased. Although SSEs of Langmuir and Freundlich models were better than those of multicomponent adsorption models, the application of single component model was limit.

References

C. R. Girish, 2017, Various Isotherm Models for Multicomponent Adsorption: A Review, Journal of Civil Engineering and Technology, 8, 80-86

C. R. Girish, 2018, Multicomponent Adsorption and the Interaction between the Adsorbent and the Adsorbate: A Review, International Journal of Mechanical Engineering and Technology, 9, 177-188

H. Maleki, 2016, Recent Advances in Aerogels for Environmental Remediation Applications: A Review, Chemical Engineering Journal, 300, 98-118

B. Naik, V. Desai, M. Kowshik, V. S. Prasad, G. F. Fernando, N. N. Ghosh, 2011, Synthesis of Ag/AgCl-Mesoporous Silica Nanocomposites using a Simple Aqueous Solution-Based Chemical Method and a Study of Their Antibacterial Activity on E. coli, Particuology, 9, 243-247

K. M. Parida, D. Rath, 2009, Amine Functionalized MCM-41: An Active and Reusable Catalyst for Knovenagel Condensation Reaction, Journal of Molecular Catalysis A: Chemical, 310, 93-100

B. Pornchuti, B. Pongpattananurak, D. Sutthiard, P. Singtothong, 2020, Adsorption of Copper, Nickel, and Chromium Ions using Silica Aerogel Synthesized by Ambient-Pressure Drying and Modified with EDTA, IOP Conference Series: Materials Science and Engineering, 778, 012133

P. Qi, T. Pichler, 2017, Competitive Adsorption of As(III), As(V), Sb(III), and Sb(V) onto Ferrihydrite in Multi-Component Systems: Implications for Mobility and Distribution, Journal of Hazardous Materials, 330, 142-148

S. Sertsing, T. Chukeaw, S. Pengpanich, B. Pornchuti, 2018, Adsorption of Nickel and Chromium Ions by Amine-Functionalized Silica Aerogel, MATEC Web of Conferences, 156, 03014

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50345-6

Use of Environmental Assessment and Techno Economic Analysis (TEA) to Evaluate the Impact and Feasibility of Coatings for Manufacturing Processes

Antoine Merlo[a*], Grégoire Léonard[a]

[a]*Department of Chemical Engineering, University of Liège, Quartier Agora B6a Sart-Tilman, 4000, Belgium*
**antoine.merlo@uliege.be*

Abstract

Coatings have become ubiquitous in modern manufacturing processes as the mechanical improvements or the new properties that they offer bring a lot of added value in the processes in which they are used. Coatings offer interesting challenges in terms of environmental and economic assessment as the deposition process and the use phase of the coating are often decoupled, although both need to be considered for a proper evaluation.

In this paper, the impact of the choices made during the deposition process will be demonstrated through a case study about TiAlN coatings for machining applications. Two deposition techniques are evaluated: Magnetron Sputtering in Direct Current (DC-MS) and in Hi-Powered Impulse (HiPIMS) regime. While coatings deposited with the HiPIMS technology are costlier and have a higher carbon impact, their increased coating life compensates that higher cost. Inclusion of the impact and costs of other aspects such as the steel substrate production would further increase the benefits of using HiPIMS.

Keywords: Environmental Assessment, Economic Assessment, Machining, TiAlN, HiPIMS

1. Introduction

Machining is an important part of many modern supply chains, for the automotive or electronical industries for example. Global market for machining is estimated to have reached a worth of $341.91 bn in 2019 in which China is the biggest shareholder [1]. In just the machining sector, the global market for cutting tools reached a size of $34.42 bn [2]. Due to the impact of that sector, finding ways to increase the performance of the machining process would be of great importance. One of the ways the machining process can be improved is, for example, by using coatings to improve the mechanical properties and the durability of the cutting tools. In the present paper, a method for economic and environmental joint assessment for different coating technologies will be presented to evaluate feasibility of the HiPIMS technology for TiAlN coatings.

2. Goal and scope

The goal of the present work is is to compare Ti0.5Al0.5N coatings deposited by HiPIMS (Hi-Powered Impulse Magnetron Sputtering) and by DC-MS (Direct Current Magnetron Sputtering) on cutting tools. The motivations are that previous works have shown that using HiPIMS to deposit TiAlN has the capacity to greatly extend the tool life time under

cutting conditions (75 vs 50 minutes) [3]. That result combined with the fact that HiPIMS, while having interesting properties, tends to be costlier due to lower deposition rates and higher investment costs [4], make up an interesting trade-off between cost and performance to study in the use phase. The target public for this work is mostly the scientific community interested in either joint economic and ecological assessment, HiPIMS and cutting tools technology. The functional unit (FU) is an 8 mm X 8 cm high-speed steel (HSS) tool coated by 4µm Ti0.5Al0.5N coatings sputtered by DC-MS or HiPIMS.

This study will take into account the extraction phase for all materials and the coating phase. Instead of a full Life Cycle Analysis, the study will focus on CO_2 emissions only because they are the main environmental impact linked to cutting and to the present coating technologies as their main difference is electricity consumption. This is also a way to streamline data acquisition.

3. Data collection and inventory

Several subprocesses are necessary to assess the cost and impacts of the coating process. Those are: production of the TiAlN target, gas production and the deposition process itself.

Target production itself is comprised of several different phases: Metal extraction, melting, powderization, annealing, compaction and sintering. To assess the energy consumption of those phases, the calculation of Kruzhanov is used [5]. The generic data provided in [5] is used for all phases except for metal extraction where emission factors (EF) from [6] are used for the CO_2 emitted for aluminium and titanium production. The energy requirements are shown in Table 1.

Table 1: Energy requirements of TiAlN target production

	Powder production	Compaction	Sintering	Total
Energy requirement (kWh per kg)	1.32	1	4	6.32

Because the target's production country is unspecified, the EF of the energy production is taken as the global average: 475 gCO_2/kWh [7]. Cost of target is assumed to be of 400 €/kg based on prices of partners.

The deposition process itself must also be evaluated, which takes place in a vacuum chamber. A representation of the chamber is given in Figure 1. The chamber is assumed to be octagonal and to be able to accommodate 5 rows of 5 tools on each of its sides. The substrate holder is able to rotate and to heat the tools to 450°C. A thickness of 4 µm and a deposition rate of 1.8 µm/h for HiPIMS are assumed [4]. The only difference for the two technologies is that a deposition rate of twice the one for HiPIMS, i.e. 3.6 µm/h, is assumed for DC-MS [3]. No material losses are assumed as a first approximation as the impact of TiAlN is negligible in the results.

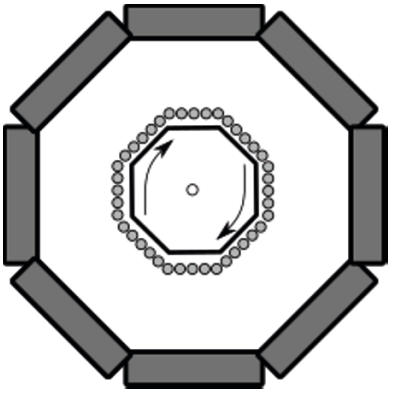


Figure 1: Representation of the deposition chamber for DC-MS and HiPIMS

The details for chamber operation are given in Table 2 for calculation of energy and gas consumption through deposition time.

Table 2: Energy consumption and time requirements of the deposition phase subprocesses of HiPIMS and DC-MS

	Power (W)	Time (Min)	Source
Heating	2000	30	[3]
Etching	120000	30	[3]
Deposition	120000	HiPIMS: 133.3 DC-MS: 66.7	[3]
Cooling	2000	20	[8]
Grinding	1000	10	[8]
Loading/Unloading	-	20	[8]
Pumping	8000	continuous	[8]

Coating is assumed to take place in Belgium, the EF for the energy production is then of 174 gCO_2/kWh [9]. Based on invoices, cost of nitrogen is assumed to be 0.332 \$/kg. Cost of argon is assumed to be 0.976 \$/kg. A dollar-euro conversion rate of 1.18 \$/€ is used for all of the present work. The EF for argon is considered to be 0.385 $kgCO_2/m^3$ [10]. As for nitrogen, the energy needed for air separation is 243 kWh/t of nitrogen [11] and global average EF for the energy used in the separation process is considered. A consumption of 400 sccm of Ar is assumed during the heating, etching and deposition phases as well as a consumption 400 sccm of N_2 during the deposition phase.

4. Cost and environmental assessment

The first goal is to differentiate the costs and impacts of DC-MS and HiPIMS for the different steps of coatings production.

4.1 Gas production

Following the assumptions made in section 3.2, consumption, cost and emissions of the gases used in the deposition process for a batch of 200 tools are summarized in Table 3.

Table 3: Gas consumption, costs and related CO_2 emissions per batch

	HiPIMS		DC-MS	
	Ar	N	Ar	N
Gas consumption (g)	131.5	64.2	86.1	32.1
Total Gas cost (€)	0.11	<0.02	0.08	<0.01
Gas CO_2 emissions (g CO_2)	29.8	7.7	19.5	3.8

4.2 Target production

Using a TiAlN density of 4.8 g/cm³, a coating of 38.6 mg is deposited on every tool, from which 28.1 mg comes from the sputtered target and 10.5 mg comes from nitrogen. Following assumptions from section 3, 62.4 gCO_2 are emitted for target production per batch for both DC-MS and HiPIMS. Every batch induces a cost of 2.24 € of target consumption.

4.3 Electricity consumption

The electricity consumption for the coating process is summarized in Table 4.

Table 4: Electricity consumption of the deposition subprocesses

Subprocess	HIPIMS	MS
Heating (kWh/**batch**)	5.00	5.00
Etching (kWh/**batch**)	65.00	65.00
Deposition (kWh/**batch**)	297.78	148.89
Cooling (kWh/**batch**)	0.67	0.67
Grinding (kWh/**batch**)	0.17	0.17
Total (kWh/batch)	368.61	219.72
Total (kWh/tool)	1.84	1.10

Most of the energy is used in the deposition and the etching phase due to the area of targets in those types of MS installations. Using the Belgian EF, 64 138.3 gCO_2 and 38 231.7 gCO_2 are emitted per batch for HiPIMS and DC-MS respectively. Using a cost of 0.0807 €/kWh, electricity costs per batch are 29.7 € and 17.7 € for HiPIMS and DC-MS respectively.

4.4 Labor, maintenance and annuities

The final metrics to evaluate the costs of each technology are labor, maintenance and annuities costs. In order to assess those aspects per coating produced, it is necessary to establish working parameters.

For maintenance, the annual costs are assumed to be 2% of the equipment costs each year. For annuities, a time of return on investment of 5 years is assumed, while inflation is neglected as a first approximation. For labor, each installation is assumed to work 300 days a year with 2 shifts of 8 hours every day. A single operator per shift with a hourly

salary cost of 40 € is assumed to be assigned to the installation. Due to the longer time per batch for the HiPIMS process, the annuity cost per FU will be higher.

4.5 Summary

Using the previously stated costs and assumptions, costs and CO_2 emissions per FU can be worked out. The total costs per FU are shown in Figure 2. They amount to 2.36 € per HiPIMS coating and 1.67 € per DC-MS coating.

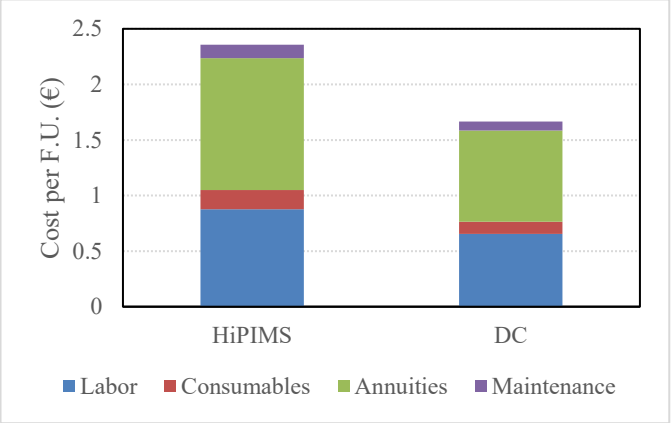


Figure 2: Cost breakdown of the coatings for DC-MS and HiPIMS

Most of the costs are related to labor and annuities. This is due to the relatively high investment cost of the installations. For HiPIMS, that investment cost is higher, and the cost per coating is exacerbated by HiPIMS' lower productivity.

CO_2 emissions for each technology are presented in Figure 3. Electricity production is overwhelmingly responsible for CO_2 emissions for both technologies. This is mainly due to the large amount of energy required for coating compared to the amount of material deposited. Due to the lower deposition rate of the HiPIMS technology, the amount of CO_2 emitted by this technology compared to DC-MS is noticeably higher (320.7 gCO_2 vs 191.2 gCO_2).

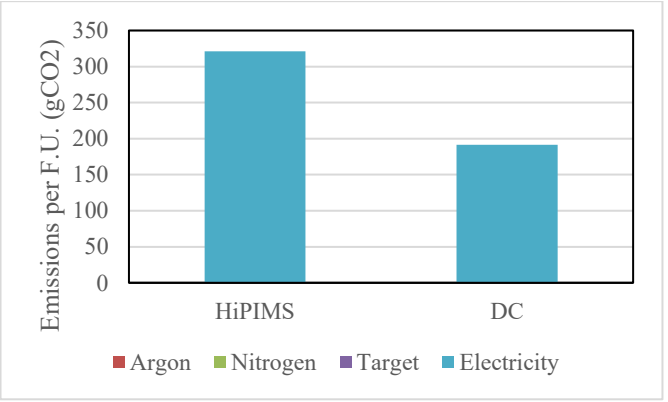


Figure 3: Total CO2 emissions for DC-MS and HiPIMS technologies

5. Conclusions

In summary, HiPIMS coatings have a 41% increased cost and a 68% increase in CO2 emissions compared to their DC-MS counterpart. However, with tool lifetimes of 75 minutes and 50 minutes for HiPIMS and DC respectively, one can reasonably justify the choice of HiPIMS coatings, at least in terms of costs. Indeed, the tool lifetime is increased by 50% for HiPIMS, while its costs only increases by 41%, making this choice a priori advantageous especially as a longer tool lifetime will also reduce the downtimes of the machining process and thus improve its cost efficiency. Further work will include the use phase in the analysis as well as technical data in order to verify these assumptions. Finally, further perspectives will also consider the cost of the HSS tool substrate and the CO_2 emissions linked to its production. Despite a lower productivity of HiPIMS, these elements will also presumably benefit the HiPIMS technology due to the longer tool lifetime it allows.

6. References

[1] Beroe Inc., 2019, Machining Market to Reach $414.17 Billion by 2022, Says Beroe Inc[Online]. Last accessed: 19/10/2021 [Available at : https://www.prnewswire.com/news-releases/machining-market-to-reach-414-17-billion-by-2022--says-beroe-inc-300939464.html]

[2] IndustryArc, 2021, Cutting Tools Market - Forecast(2021 - 2026)

[3] Weichart, J., Lechthaler, M., 2012, Titanium aluminum nitride sputtered by HIPIMS, M. IOP Conf. Series: Mater. Sci. Eng., 39, 012001.

[4] Anders, A.,2010, Deposition Rates of High Power Impulse Magnetron Sputtering: Physics and Economics. Lawrence Berkeley National Laboratory. Retrieved from https://escholarship.org/uc/item/7cz4g23b

[5] Kruzhanov, V., Arnhold, V., 2012, Energy consumption in powder metallurgical manufacturing. Powder Metallurgy, 55(1). DOI 10.1179/174329012X13318077875722

[6] Nuss, P. and Eckelman, M. J., 2014, Life Cycle Assessment of Metals: A Scientific Synthesis. PLoS ONE 9(7): e101298. https://doi.org/10.1371/journal.pone.0101298

[7] IEA, 2019, Global Energy & CO2 Status Report 2019

[8] Centre de Recherche Métallurgique (CRM)'s installations

[9] European Environment Agency, 2021, Indicator assessment: Greenhouse gas emission intensity of electricity generation in Europe. https://www.eea.europa.eu/data-and-maps/indicators/overview-of-the-electricity-production-3/assessment-1

[10] Nakhla, H., Shen, J.Y. and Bethea, M., 2012, Environmental Impacts of Using Welding Gas. The Journal of Technology, Management, and Applied Engineering, 28(3).

[11] European Industrial Gases Association, 2010, Position Paper, Indirect CO2 emissions compensation: Benchmark proposal for Air Separation Plants

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50346-8

Forecasting Operational Conditions: A case-study from dewatering of biomass at an industrial wastewater treatment plant

Sebastian Olivier Nymann Topalian[a], Pedram Ramin[a], Kasper Kjellberg[b], Murat Kulahci[c], Xavier Flores Alsina[a], Damien J. Batstone[d], Krist V. Gernaey[a]

[a]*Technical University of Denmark, Department of Chemical and Biochemical Engineering, Søltofts Plads 228A, Kgs. Lyngby, 2800, Denmark*
[b]*Novozymes A/S, Hallas Alle 1, Kalundborg, 4400, Denmark*
[c]*Technical University of Denmark, Department of Applied Mathematics and Computer Science, Richard Petersens Plads 324, Kgs. Lyngby, 2800, Denmark*
[d]*Advanced Water Management Centre, The University of Queensland, Brisbane, QLD 4067, Australia*

Corresponding Author's E-mail: kvg@kt.dtu.dk

Abstract

In this paper, we present a data-driven approach to predicting polymer dosages for industrial decanters based on upstream production data. First, a data extraction algorithm using on-line sensors is developed to identify when the operational mode is changed with a 99 % accuracy. Next, an investigation of process delays in the collected data is carried out by analysing partial autocorrelation matrix eigenvalues upon which is it concluded to transform the data by summarising the data by batch and including lagged summaries to account for a time delay of 2 hours. Finally, a random forest forecasting model is trained capable of learning structured information from the lagged summaries producing decent predictions for both low and high polymer dosages (RMSE 14.89). The proposed approach could potentially save operators 3-6 hours a day.

Keywords: Control; Operation; Forecasting; Environmental Systems;

1. Introduction

Decanters are widely used in the biotech industry to carry out solid-liquid separations. Achieving adequate separation in the decanters reduces the amount of energy, and therefore money, required to treat the reject water in the waterline, however this is at the cost of adding chemicals to the decanters. At the current point in time, it is deemed unfeasible to control the settings of the decanters based on a cost-benefit analysis due to the complexity of accounting for savings in the waterline, so the problem is reduced to achieving decent separation with the necessary operational conditions as determined by the operators. Running decanters with a highly variable feed composition in a satisfactory manner is a task for operators who manually carry out flocculation tests to estimate the proper chemical dosage at a given point in time. This procedure is labour intensive and cumbersome. It also has a strong time component i.e delay between the result of the flocculation essay and the need of deciding for a dosage strategy. In order to circumvent this limitation, we developed a mathematical model based on on-line sensor data to estimate dosing of polymer based on upstream process information. The proposed methodology is tested at case study largest industrial wastewater treatment plant in Northern Europe.

2. Methods

2.1. Plant description

The plant has a hall with decanters that treat approximately 200 m^3/h of biomass and waste activated sludge with operation around the clock. The stream that goes into the

decanter hall comes from one of two storage tanks that are operated in a semi-continuous manner, and the biomass in the storage tanks consists of inactivated biomass from the upstream biotechnological production as well as waste activated sludge from the waterline at the plant. Apart from storing material, the storage tanks also buffer the flow to the decanter hall allowing for smoother operation, and a regular emptying pattern, whereas the filling pattern is dictated by upstream production. There are approximately 4-6 batches going through the decanter hall every day. The decanters are controlled by operators who adjust the polymer dosage, among other settings, depending on the biomass and sludge that is being treated for each batch. The operators judge the quality of the separation by looking at the reject water from the decanters with a visual test, and then decide to change the polymer dosage, among other settings, accordingly. A conceptual illustration of the plant is provided in Figure 1.

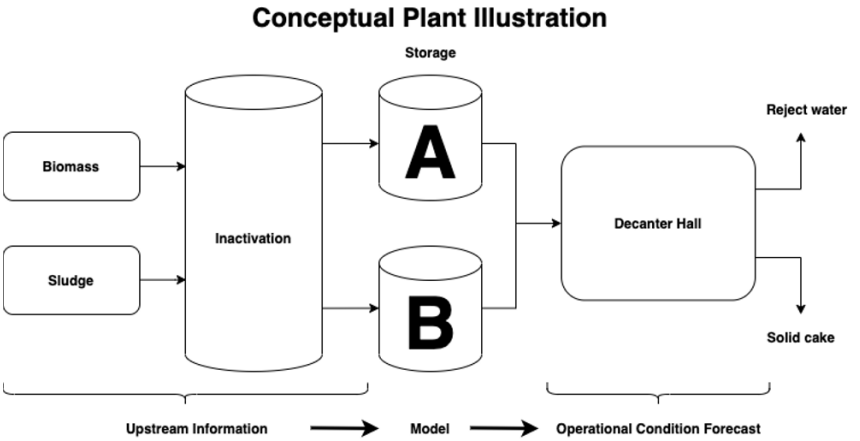


Figure 1 – Conceptual illustration of the plant. Biomass and waste activated sludge are inactivated and stored in two storage tanks that semi-continously feed into a hall with centrifugal decanters. The upstream information is used to produce operational condition forecasts to assist operators in controlling the decanters.

2.2. Data availability

The plant is equipped with online sensors of pH, temperature, tank volumes, flowrate among others, and historic data from the sensors is available from an online database. Historic time-series data from the 1st of July to 31st of December 2020 were extracted from the online database with a resolution of 15-minute averages. Categorical information such as production codes were 15-minute medians. In total 302 variables from the plant were selected for analysis. The data retrieved from the online database is ordered by time of observation, however the plant contains several unit operations with residence times larger than the 15-minute averages for which data is obtained, and hence comparing upstream variables with downstream variables at the same time should exhibit poor correlation unless the variables being compared are severely autocorrelated.

2.4. Batch data-extraction algorithm

A custom algorithm is developed to automatically identify when a batch begins when the other tank starts to empty (see Figure 2). In this way, a batch change is identified by checking that the tank volume is above approx. 50 %, and that the gradient of that volume is negative, and that the gradient is larger than 0.5 % as to prevent identifying a batch change time when operation is stopped due to sensor noise (approx. 0.1 %).

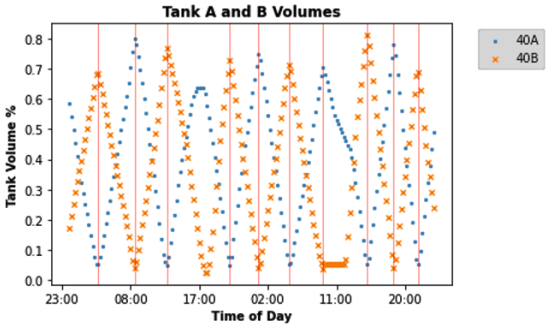


Figure 2 – Algorithm for detection of batch changes for the semi-continuous tank system. Identified batch changes are marked with a vertical red line.

2.3. Partial autocorrelation matrix (PACM)

The partial autocorrelation matrix (PACM) is used to study process delays due to unit operations. Due to the semi continuous operation of the tanks, a large delay is expected that correlates the upstream data with the downstream data at a lag corresponding to the residence time in the storage tanks. Summarising The PACM is constructed by calculating the partial correlation matrix for lagged versions of the dataset, and the eigenvalues for each matrix. Next, the obtained PACM is compared to the eigenvalues of the matrix at prior and future lags where large and small eigenvalues indicate overall good and correlation between the variables and their lagged states respectively (Vanhatalo et al., 2017).

2.5. Predictive method

Random forests are widely regarded as a good off-the-shelf regression model for structured data, such as the data presented herein (Hastie et al., 2017). The random forest (RF) algorithm is used to construct a forecast model that predicts the operator decided polymer dosage for a batch based on the upstream information obtained while the batch is in storage. The first 80 % of the batch observations are used as training data to decide which hyper-parameters to utilize, and the final 20 % of the batch observations are used to evaluate the model performance by recursively refitting the model and predicting the next batch polymer dosage.

3. Results

3.1. Analysis of batch data

The proposed algorithm identified batch changes with a 99% accuracy. In total 869 batches were analyzed for the 6-month period. The target objective of the forecast model is the polymer dosage after the operators have evaluated and changed the operational conditions, implying that the mean for a batch averages out the conditions before and after the operators have done their job, so to have a more accurate estimate of the polymer dosage a zero-excluded median is calculated for the decanters in the hall, since this removes idle machines or machines that exhibit irregular behaviour. The zero-excluded median is calculated after rounding the dosages to nearest base 5 as this is the typical magnitude of change that the operators will apply and the final value for a batch is considered as the suggested dosage from the side of the operators. In table 1 the initial polymer dosage, final polymer dosage and percentage of batches with new conditions is shown for all the identified batches. A batch is deemed to have new conditions if the calculated starting median differs from the final one. The statistics are shown for all three shifts in Table 1. The afternoon shift appears to utilize less polymer and change their operational conditions more frequently than the morning and night shift.

Table 1- Initial polymer dosage, final polymer dosage and % of batches optimized.

	Morning Shift	Afternoon Shift	Night Shift
Initial polymer dosage	49.6	45.8	46.3
Final polymer dosage	47.4	45.5	50.2
Batches w. new conditions	79.5 %	92.4 %	81.2 %

4.2. Assessment of process delays

In Figure 3 the largest absolute eigenvalue for the partial autocorrelation matrix (PACM) as a function of lag is shown for the untreated data on the left, and for batch summarised data on the right for the first two months of data (Vanhatalo et al., 2017). The PACM eigenvalues for the untreated data display a high correlation between the variables at delay 20, which could correspond to the process delay caused by the storage tanks that feed the decanter hall.

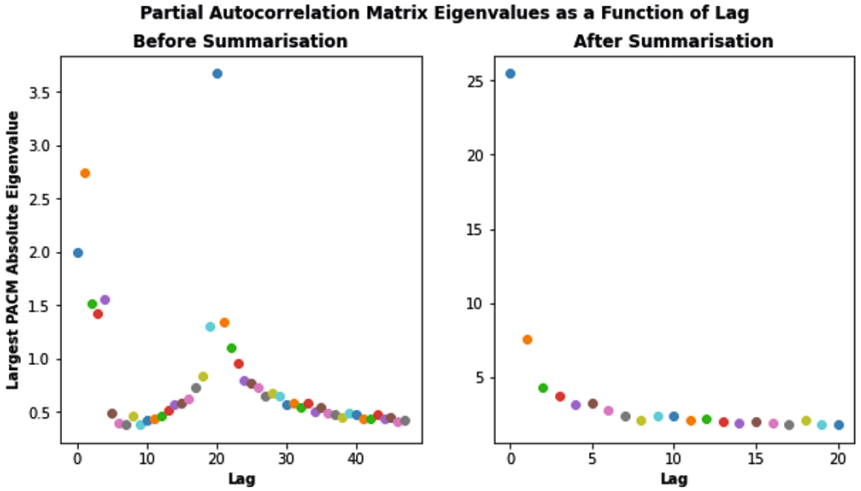


Figure 3 – Largest absolute eigenvalues for the PACM at different lags before and after batch summarization left and right respectively.

The summarised data is calculated by averaging all values for a given batch time, and the lag corresponds to moving the starting and end point of the batch by one timestep, to check if different variables should be summarised by different time indices to account for delays occurring before the semicontinuous storage tanks. An illustration is provided in Figure 4. The eigenvalues past lag 8 appear to stabilize, so for the forecasting model these lagged summaries will be included when developing the forecasting model, effectively accounting for the past 2 hours.

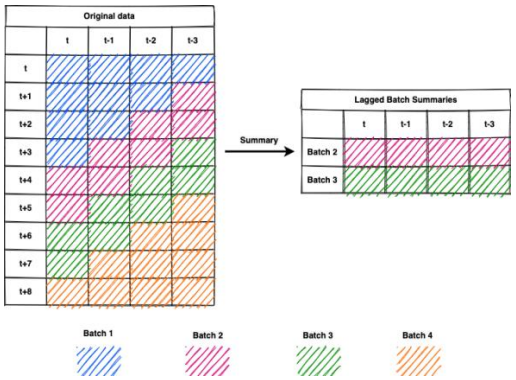


Figure 4 – Batch summarization illustration. For each identified batch the mean is calculated for each variable, and for each lagged version of the variables.

4.3. Random forest

4.3.1. Model training

The sci-kit learn library is used to carry out the hyper-parameter optimization and train the models (Pedregosa et al., 2011). The hyper-parameter optimization is carried out by building 500 random forests with 1024 trees where the complexity parameter, maximum number of features and maximum depth are drawn from the uniform distributions U(0.2, 0.8), U{1, 2700} and U{1,10} for each random forest respectively. The random forest with the best performance has complexity parameter, maximum number of features and maximum depth values of 0.23, 105 and 9 respectively.

4.3.2. Model predictions

For each observation in the test set the model is refit on all prior data to produce the forecasts. The root mean square error (RMSE) on the test set is 14.89 whereas using a naïve guess such as the mean of polymer dosage results in an RMSE of 18.44 and the RF algorithm thus successfully learns structured information, however considering that the average polymer dosage is between 45-50 the RMSE is still considered as high. In Figure 5 two residual plots are shown. The left one shows the residuals of the model on the test set as a function of time, and the right one shows the residuals as a function of the model prediction. Each residual is shown with a dot, if the batch conditions are new compared to the previous conditions, and a cross, if the batch conditions are old corresponding to unchanged conditions from the last batch.

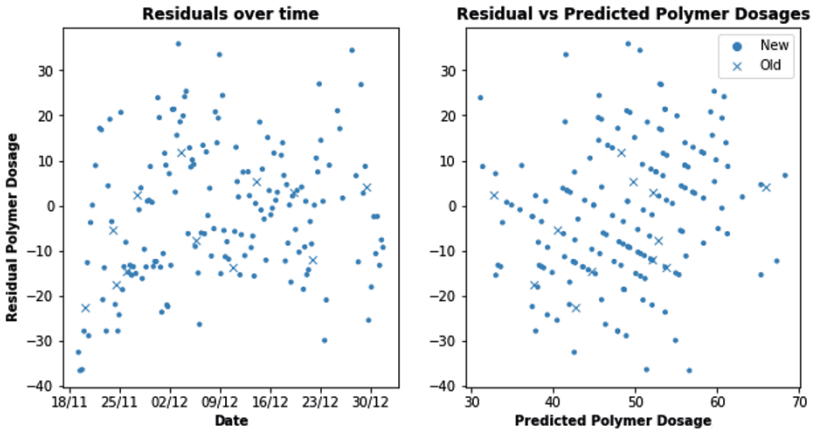


Figure 5 – Model residuals a function of date on the left, and as a function of predicted polymer dosage on the right. A dot represents a batch with new operational conditions, and a cross old operational conditions compared to that of the previous batch.

4.3.3. Discussion

From the left plot in Figure 5 there does not appear to be an increasing trend in the model residuals indicating that there is no systematic error in the model, and that the model can predict both low and high polymer dosages. The batches that utilize old conditions have a lower residual on average, which could correspond to batches where the operators deem the performance of the decanters sufficient, and therefor keep the old operational conditions. The forecasting model could therefore be applied in situations where the old conditions are insufficient to generate a starting point for the operators. For each batch where the model provides an adequate estimate 1-2 hours of time could potentially be saved for the operators allowing them to perform other crucial tasks. Upon investigating the random forests feature importance it becomes evident that the tree utilizes primarily flow rates, and information about time to make predictions, and that information related to upstream product information, pH and temperature are not utilized as much. Models that inherently utilize time as information can be impractical since they often do not convey implications of physical phenomena, however for systems with temporal patterns they can provide a significant boost in forecasting accuracy. As a means of comparison, a naïve forecast where the prediction for each value is the prior value yields a RMSE of 17.43 which is 17.06 % larger than that of the RF model. Random forests are considered as a good off-the-shelf data mining procedure, and here they also achieve moderate success encouraging further data collection and investigation of more sophisticated and time-consuming forecasting methods such as neural network approaches. One inherent drawback of predicting the actions of operators is that the operation could change depending on the operators working a shift and including the operator schedule could lead to a better forecasting model and provide a method for evaluating operator performance through data-engineered key performance indicators. The framework presented herein can also be utilized for predicting other operational conditions, and transferred to other plants with similar plant layouts, or for predicting decanter failure.

4. Conclusion

This study demonstrates performance analysis of an industrial-scale process using plant-wide operational data. With proper data treatment, data analysis, feature selection and data-driven modelling it was possible to make an automated algorithm to handle complex datasets for prediction purposes and later optimization and control. We propose that it could save operators between 3-6 hours of work every day, leaving room to carry out other important tasks at the plant, however further model development and following verification is required to increase the accuracy of the forecasts.

5. Acknowledgement

The authors would like to acknowledge the financial and academic support of the Technical University of Denmark and Novozymes for their contributions to the project.

6. References

T. Hastie, R. Tibshirani, J. Friedman, 2017, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd Edition, pp. 350-352.

Pedregosa et al., 2011, *Scikit-learn: Machine Learning in Python*, JMLR, 12, pp. 2825-2830.

E.Vanhatalo, M. Kulahci, B. Bergquist, 2017, *On the structure of dynamic principal component analysis used in statistical process monitoring*, Chemometrics and Intelligent Laboratory Systems, 167, pp. 1-11.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50347-X

Plant wide modelling of a full-scale industrial water treatment system

Vicente T. Monje[a], Helena Junicke[a], Kasper Kjellberg[b], Krist V. Gernaey[a], Xavier Flores Alsina[a]

[a]*Technical University of Denmark, Department of Chemical and Biochemical Engineering, Søltofts Plads, Building 228A, 2800 Kgs. Lyngby, Denmark*
[b]*Novozymes A/S, Hallas Alle 1, 4400 Kalundborg, Denmark*
Corresponding Author's E-mail: xfa@kt.dtu.dk

Abstract

In this study a set of mathematical tools are developed and assembled together to assess and predict mass and volumetric flows in industrial water treatment systems (iWTS). The proposed approach is constructed upon a set of data reconciliation methods, influent fractionation routines and process simulations models (and model interfaces) to balance, analyse, reproduce and forecast the behaviour of different compounds within treatment facilities. The proposed approach is tested on full-scale data collected after a five week measuring campaign at the largest iWTS in Northern Europe. Results show that the proposed approach is capable to predict the occurrence, transformation and fate of COD, N, P, S and multiple metals (Na, K, Ca, Mg and Al).

Keywords: Data reconciliation, Mass balancing, Model simulation, Process systems engineering, Scenario analysis, Wastewater

1. Introduction

Industrial wastewaters have very diverse dynamics (compared to urban wastewater), which is a result of different production schemes/schedules within the factory. Variable pH, influent biodegradability, non-standard COD/N and COD/P ratios might challenge traditional biological processes. In some cases, high loads decrease methane/biogas production (and potential energy recovery). This reduction is attributed to two factors: 1) loss of electron equivalents due to the presence of sulfate reducing bacteria; and, 2) decrease of acetoclastic and hydrogenotrophic methanogenesis due to sulfide inhibition. Metals and some inorganic/organic compounds can inhibit microbial growth and/or have severe toxicity effects. The high content of cations and anions promotes the formation of precipitates at different locations in the reactor (granules, pipes), which can have detrimental (decrease of methanogenic activity) or catastrophic (cementation) effects on reactor performance (Feldman et al., 2017). Hence, mathematical models describing iWTS should include all these (hostile) phenomena in order to produce reliable predictions.

The objective of this study is to present a set of mathematical tools to assess and predict mass and volumetric flows in iWTS. The study presents the following novelties: 1) The results of a 5 week sampling campaign at the largest iWTS Norther Europe; 2) A reconciled mass balance analysis showing the occurrence, transformation and fate of traditional (COD, N, P & S) but also non-traditional (Na, K, Ca, Mg, Al) compounds; and, 3) A customized/calibrated model library describing multiple technologies treating different types of waste streams (liquid, solid).

This work goes beyond state of the art by presenting a modelling approach: 1) dealing with extremely concentrated streams (2,5 M PE in 10,000 m^3/day); 2) modifying the existing mathematical model structures to adapt to the harsh industrial conditions; and, 3) extending the quantity of monitored compounds up to 10 (Q, traditional and non-traditional compounds) + pH + VSS/TSS ratio; 4) presenting for a first time an integrated plant-wide model dealing with a large industrial iWTS at this level of detail.

2. Methods

2.1 Plant description and measuring campaign

Figure 1 shows a schematics + detailed description of the case study. Influent flow may be treated anaerobically (PAT & AGSR) or aerobically (ASR, SEC, FLOT) or both. Biogas goes through a cleaning process (SCRUBB, REAC & SET) before being introduced to a gas motor for energy (electricity / heat) recovery purposes. Reject water from biomass dewatering can be sent to either the PAT & AGSR or the ASR, SEC, FLOT section (both is also an option). The output of PAT & AGSR is sent to the ASR.

A five week measuring campaign was conducted. Samples were taken from 11 locations within the plant: 3 influent streams (PWW_2, $_3$ & $_4$), after PRIM ($PRIM_{over}$), after PA (PAT_{liq}), after AGSR ($AGSR_{liq}$), after ASR (ASR_{liq}), after SEC (SEC_{eff}, SEC_{RASS} & SEC_{WAS}) and after dewatering (DEW_{under}, DEW_{over}). Additional one day samples were taken to characterize: primary underflow ($PRIM_{under}$) and the output of the inactivation tank (IT_{liq1}, IT_{liq2}). Measurements involved the determination of: TSS, VSS, COD, TN, TP and TS and multiple metals (Al, Ca, Mg, Na and K), in both unfiltered and filtered samples. The analysis also includes the quantification of nitrates (NO_x).

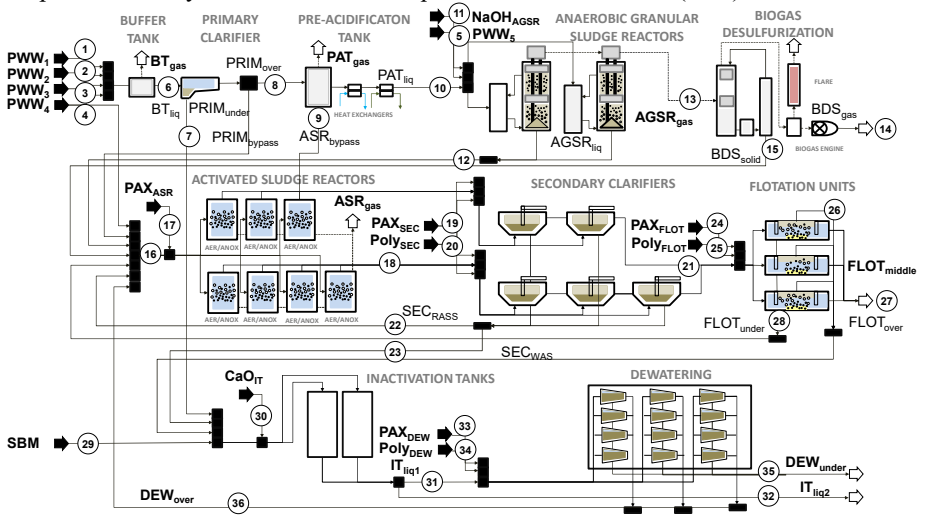


Figure 1. Flow diagram, measuring points, defined sub-systems and mass balances of the iWTS under study: **1-5** process waste water ($PWW_{1\text{-}5}$), **6** - effluent buffer tank (BT_{liq}), **7-8** –primary clarifier (PRIM) overflow and underflow ($PRIM_{under}$, $PRIM_{over}$) **9**- by pass activated sludge reactor (ASR) to pre-acidification tank (PAT) (ASR_{bypass}), **10** – PAT effluent (PAT_{liq}), **11** – NaOH addition to the anaerobic granular sludge reactor (AGSR)($NaOH_{AGSR}$), 12 – AGSR effluent ($AGSR_{liq}$), **13** – AGSR biogas (untreated) ($AGSR_{gas}$), **14** – AGSR biogas (treated) (BDS_{gas}), **15** – S recovered (BDS_{Solid}), **16** – input AER, **17** – Poly-aluminum chloride (PAX) addition to ASR (PAX_{ASR}), **18** – effluent ASR (ASR_{liq}), **19-20** – PAX and Polymer addition to secondary settler (SEC) (PAX_{SEC}, $Poly_{SEC}$), **21-23** – SEC effluent, recirculation and waste flow ($SEC_{eff,}$ SEC_{RASS}, SEC_{WAS}), **24-26** – PAX and polymer addition to flotation (FLOT) (PAX_{FLOT}, $Poly_{FLOT}$), **26-28** – flotation streams ($FLOT_{over}$, $iWTS_{eff}$, $FLOT_{under}$), **29** – spent biomass stream (SBM), **30** – lime addition to inactivation tanks (IT) (CaO_{IT}), **31-32** – IT outputs (IT_{liq1}, IT_{liq2}), **33-34** – PAX and Polymer addition to dewatering (DEW) (PAX_{DEW}, $PolyD_{EW}$), **35-36 –** dewatering under (cake) and overflow (reject water) (DEW_{under}, DEW_{over})

2.3 Mass balancing and data reconciliation

A five step methodology is used to reconcile the data obtained during the measuring campaign (Puig et al., 2008, Behami et al., 2019): 1) definition of the identity matrix; 2) curation, processing, cleansing and data analysis; 3) estimation of the missing fluxes; 4) calculation of optimal flows using Lagrange multipliers; and, 5) new data set quality verification.

2.4. Influent fractionation

Influent fractionation is based on earlier work (Feldman et al., 2017; Monje et al., 2021). Essentially, ADM states were estimated by assuming: 1) degree of COD biodegradability (D_{BIO}) in both COD_{sol} and COD_{part}; 2) degree of acidification (D_{ACID}) in CODs; 3) fraction of ethanol (D_{ETOH}). Hence, it is possible to determine S_I, X_I, S_{ac} and S_{ETOH}. Once this is established, X_{li}, X_{prot} and S_{aa} are quantified using P and N content of the aforementioned compounds. Since the influent wastewater originates from the fermentation industry, sugars ($S_{su} = 0$) and carbohydrates ($X_{ch} = 0$) should not be present (consumed upstream). The quantity of P not associated with organics is assumed to be inorganic and precipitated (mainly calcium phosphate) ($X_{Ca\text{-}P}$). Any remaining particulate calcium and magnesium will be assumed to be calcium carbonate (X_{CaCO3}). The influent inorganics (X_{ISS0}) are determined using the TSS/VSS ratio minus the quantity of precipitates.

2.5. Main mathematical models

The main model is based on: (1) a biological model; (2) a physico-chemical model; and, (3) model interfaces. The biological models (BM) comprise an anaerobic digestion model (ADM) and an activated sludge model (ASM). The ADM is used to describe influent conditions, BT, PRIM, PA, AGSR, IT and DW, while the ASM describes the ASR, SEC, and FLOT units. The physico-chemical model (PCM) includes an aqueous phase + precipitation model and a gas transfer model. Finally, the model interfaces include an ADM/ASM/ADM interface and PCM/ADM/ASM interface. The ADM/ASM is incorporated before the AS unit and the ASM/ADM after the SEC. The outputs of the ASM/ADM at each integration step are used as inputs for the PCM module to estimate pH, ion speciation/pairing, precipitation potential and stripping. A comprehensive description of these models can be found elsewhere (Feldman et al., 2017; Flores-Alsina et al., 2019).

3. Results

3.1. Reconciled mass balances

Result of reconciled mass balances revealed that the effluent of the iWTS under study ($iWTS_{eff}$) contains the largest Q fraction (95 %). The dewatered cake (DEW_{under}) and IT liquor (IT_{liq2}) only account for a marginal contribution (<5 %). About 32 % of the incoming COD is captured in the anaerobic granular sludge reactor ($AGSR_{gas}$) and then potentially converted into electricity and heat. The remaining COD is lost in intermediate operations (BT_{gas}, PAT_{gas}) (12 %), burned aerobically/anoxically in the activated sludge section (ASR_{gas}) (23 %), part of the effluent ($iWTS_{eff}$) (1 %) or trucked to an external biogas facility as sludge cake (32 %) (IT_{liq2}, DEW_{under}). Regarding N, 65 % is removed in the activated sludge section via nitrification/denitrification (ASR_{gas}). A significant part of N ends up being part of the bio-solids to be disposed (IT_{liq2}, DEW_{under}) (33 %) while a small fraction leaves with the effluent ($iWTS_{eff}$) (2 %). The analysis of P, Ca, Mg and Al reveals that these compounds are basically accumulated in the bio-solids (IT_{liq2}, DEW_{under}) (> 70 %) as precipitates. Contrary to that, Na, K and S remain soluble and leave via the liquid stream ($iWTS_{eff}$) (> 70 %). It is important to mention that 11 % of S is converted to H_2S in the anaerobic granular sludge reactor ($AGSR_{gas}$,) and then captured in a desulfurization tower. Next, it is re-oxidized again to S_0 and introduced (BSD) in the biological reactor.

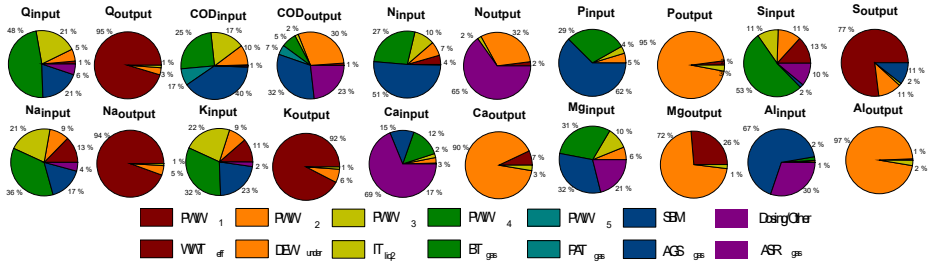


Figure 2. Contribution to influent and effluent loads of the different streams. For each reconciled magnitude (Q, COD, N, P, S, Na, K, Ca, Mg, Al), pie charts represent influent and effluent compositions, respectively.

3.2. Influent fractionation

The results of the influent fractionation are illustrated in **Figure 3.** Four examples are presented: 1) process wastewater 2, 3 & 4 (PWW $_{2, 3, 4}$); and, 2) the spent biomass stream (SBM). Indeed, **Figure 3** depicts the proportions of the different model states (both soluble and particulate) in form of a stack bar adding up to 100 % of the particulate or soluble fraction of each reconciled major component (COD, N, P & Ca). In all PWW streams, COD_{sol} and COD_{part} is mainly composed of fatty acids (S_{fa}) and lipids (X_{li}), whereas SBM mainly contains amino acids (S_{aa}) and proteins (X_{pro}). The remaining fraction is allocated to non-biodegradable organics (S_i, X_i) and VFAs (S_{ac}). With respect to N (soluble), PWW_3 and PWW_4 have a large contribution of nitrate, which comes from the use of nitric acid for cleaning equipment. The remaining N is linked to ammonia (S_{NHx}) and amino acids (S_{aa}). Particulate N is (almost) entirely allocated to X_{prot}. All the soluble P is assumed to be phosphate (S_{PO4}). The particulate fraction is allocated to lipids (X_{li}) and calcium precipitates ($X_{Ca\text{-}P}$). PWW$_{2, 4}$ and SBM have an important contribution of precipitates, which is also confirmed by their lower VSS/TSS ratios. In contrast, P in PWW_3 is mainly linked to organics (X_{li}). In both, N and P cases, a small fraction is associated with organic inert material (S_I and X_I). Ca has a soluble (S_{Ca+2}) and a particulate fraction in the form of carbonates (X_{CaCO3}) and phosphates ($X_{Ca\text{-}P}$). In the vast majority of cases, calcium carbonate is assumed to be the major component, but a small fraction is linked to phosphates depending on the availability of inorganic (particulate) P (see PWW$_{2,4}$ and SBM). A special case is PWW_3 where the low content of inorganic material is in the form of phosphates.

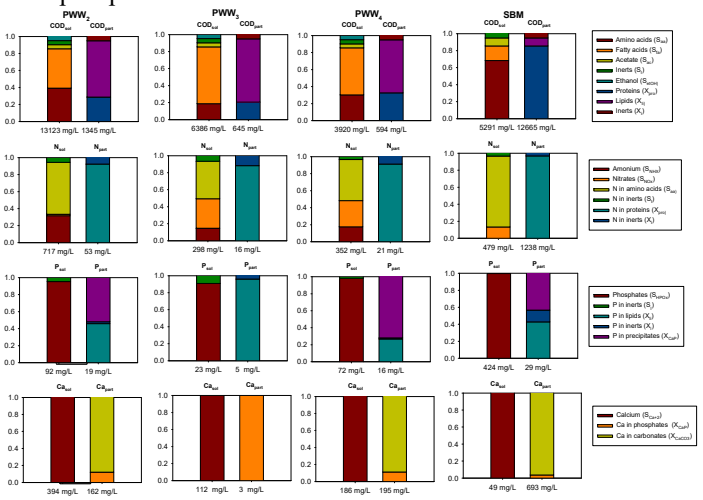


Figure 3. Fractionation of PWW$_{2, 3, 4}$ and SBM streams: COD, N, P and Ca.

A similar approach is used for Mg i.e. the soluble part is S_{Mg+2} and the particulate form is magnesium carbonate (X_{MgCO3}). For the sake of simplicity, the soluble fraction of S is associated to sulfate (S_{SOx}), while the particulate fraction is assumed to be sulfur mineral (X_{S0}). With respect to the remaining compounds (see **Fig 4**): 1) K and N are assumed to be only soluble and present in ionic form (S_{Na+} and S_{K+}); 2) Al is assumed to be only in particulate form (X_{AlOH}). The pH values were adjusted by modifying the Cl concentration. A pool of undefined particulates is used to calibrate the VSS/TSS ratio.

3.3. Computer simulations

3.3.1. Buffer tank (BT) + primary clarifier (PRIM)

COD_{sol}/COD_{part} values are the result of hydrolysis, acidogenesis and H_2 stripping. N and P transformations include hydrolysis and uptake/release during biomass growth/decay. Specifically, for N the model considers dissimilatory nitrate reduction to ammonium (see TN, NH_X and NO_x). Precipitation of P is assumed to happen (justified with the high alkalinity of PWW_2). The other studied compounds are assumed to be non-reactive (S, Mg, K, Al). In PRIM, a fraction of all compounds with a particulate fraction (COD, N, P, Ca, Mg, Al) is removed. The model also reflects well the hydraulic balance (see Q values), weak acid chemistry (see pH values) and the differences in settling velocities between organics and inorganics (see VSS/TSS values) (see **Figure 4** & **5**)

3.3.2. Pre-acidification tank (PAT)

The model predicts a reduction of COD_{sol} and an increase of the COD_{part} as result of the acidogenic activity. The first is due to hydrogen formation (and subsequently stripping). The second is the result of biomass production/growth. The models also suggest a change in the composition of COD_{sol} (~60 % is acidified) which would explain the lower pH experimental values (well predicted by the model too). Both, plant data and simulation values, show total denitrification NO_x (~ 0 g/m^3) (see **Figure 5**). Since TN values do not seem to change, DNRA was again the main mechanism. No further reactions were assumed for the other compounds (see **Figure 4** & **5**).

3.3.3. Anaerobic granular sludge reactor (AGSR)

The proposed approach is capable to reproduce COD values (soluble/particulate) resulting from the methanogenic microorganisms (see **Figure 4**). In addition, it also predicts the S removal by sulfate reducing bacteria (see **Figure 4**). No further reactions e.g. precipitation were assumed to occur. The pH and VSS/TSS predictions match well with plant data (see **Figure 4** & **5**)

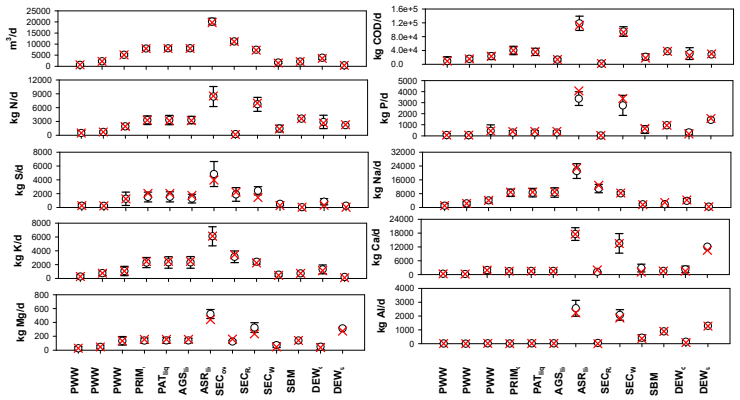


Figure 4. Steady state model predictions (red crosses) and reconciled measurements mean and standard deviation (black circles and whiskers) for several plant locations (see labels in X-axis). A: volumetric flow rate; B: TCOD; C: TN; D: TP; E: TS; F: TNa; G: TK; H: TCa; I: TMg; J: TAl. Deviation between measurements and simulations = 10 %

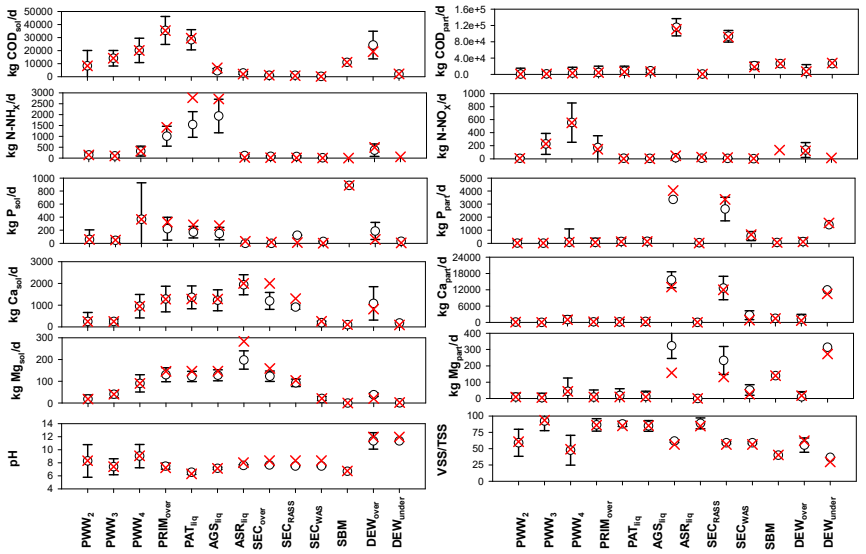


Figure 5. Steady state model predictions (red crosses) and reconciled measurements mean and standard deviation (black circles and whiskers) for several plant locations (see labels in X-axis). A, CODpart; B, CODsol; C, NH4+; D, NO3-; E, TPpart; F, TPsol; G, TCapart; H, TCasol; I, TMpart; J, TMgsol/TSS ratio. K, pH; L, VSS/TSS ratio. Deviation between measurements and simulations = 12 %

3.3.4. Activated sludge reactor (ASR) + secondary clarifier (SEC) + flotation (FLOT)

Simulations results show that both biological reactor + secondary settler model describe COD, N and P removal (the latter due to precipitation with Al). Another important aspect predicted by the model is the VSS/TSS ratio, including the effect of the particulate Ca compound arriving via reject water on the VSS/TSS ratio. The latter leads to an increase of the biomass operational mixed liquor suspended solids (MLSS) concentration in the reactor to achieve enough nitrification capacity (see **Figure 4** & **5**)

3.3.5. Inactivation tank (IT) + dewatering unit (DEW)

The model predicts three types of behaviour: 1) heavily hydrolysed; 2) precipitated; and, 3) unaltered. The dewatering module can reproduce the measuring data, i.e. soluble compounds were concentrated in the reject water stream, while particulate compounds exited the system via the IT liquor or the cake (see **Figure 4** & **5**).

4. Conclusion

This study demonstrates that the proposed approach is capable for the main streams in the iWTS to reproduce neutralization, volatile fatty acid production, particulate removal and nitrate denitrification in the first units of the flow diagram (buffer tank, primary clarifier, pre-acidification tank). It also correctly predicts biogas composition and COD recovery in form of electricity and heat in the anaerobic granular sludge reactor. Lastly, biological and chemical N and P removal processes in the activated sludge and the quality of bio-solids after inactivation/dewatering (reject water /cake) are predicted by the model.

5. References

Behami et al., 2019. Journal of Cleaner Production.218.616-618
Flores-Alsina et al., 2019. Water Research.156. 264-276
Feldman et al., 2017. Water Research. 126,488-500
Puig et al., 2008. Water Research.42,18,4645-4655
Monje et al., 2021. Journal of Environmental Managament. 293, 112806

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50348-1

A Systematic Framework for the Integration of Carbon Capture, Renewables and Energy Storage Systems for Sustainable Energy

Manali S. Zantye, Akhilesh Gandhi, Mengdi Li, Akhil Arora, M. M. Faruque Hasan*

Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX 77843-3122, USA
**Email: hasan@tamu.edu*

Abstract

In this work, we address the challenges associated with decarbonizing electricity grids through a decentralized integration scheme of individual fossil power plants with energy storage and flexible CO_2 capture. To this end, we develop a technology design and downselection framework and demonstrate a prototype tool called THESEUS (TecHno-Economic framework for Systematic Energy storage Utilization and downSelection) which enables an extensive techno-economic analysis of integrating fossil power plants with several candidate energy storage technologies which include cryogenic, molten-salt, compressed air, and batteries. The core of THESEUS is a large-scale mixed integer nonlinear programming (MINLP)-based optimization that determines an optimal selection and combination of energy storage technologies for minimizing the cost of meeting the grid electricity demand. We demonstrate THESEUS through a case study for reducing the fossil power plant cycling and minimizing carbon emissions while satisfying sharp spikes in energy demands.

Keywords: Energy Storage, Simultaneous Design and Operation, Downselection.

1. Introduction

As the global energy demand increases, there is a push to adopt sustainable renewable energy sources. However, the seamless integration of clean renewable energy with electricity grids requires measures to address the challenges arising from its inherent intermittency. Conventional measures include the cycling of the fossil-based generating units and the installation of large-scale energy storage. While power plant cycling reduces its efficiency and leads to thermal and mechanical stresses in critical plant components, grid-scale energy storage is cost-intensive with only a limited number of suitable technologies. These integration challenges make it difficult for renewables to completely replace the dispatchable fossil generators from electricity grids. CO_2 capture presents a promising solution to decarbonize fossil-based power generation, but its large-scale deployment is limited by its high energy requirement and cost (Hasan et al., 2012).

To address the challenges associated with the decarbonization technologies, we propose that the operational synergies between them can be leveraged through their localized integration with individual fossil power plants to achieve low-cost as well as reliable clean energy systems. To this end, we study a decentralized power generation system comprising of a natural gas combined cycle (NGCC) power plant integrated with co-located CO_2 capture and energy storage facilities. The benefits include the ability to meet

demand spikes without increased cycling of the power plant. Furthermore, along with reducing plant emissions by up to 90%, CO_2 capture also acts as a form of 'indirect storage' to counter renewable intermittency (Zantye et al., 2021).

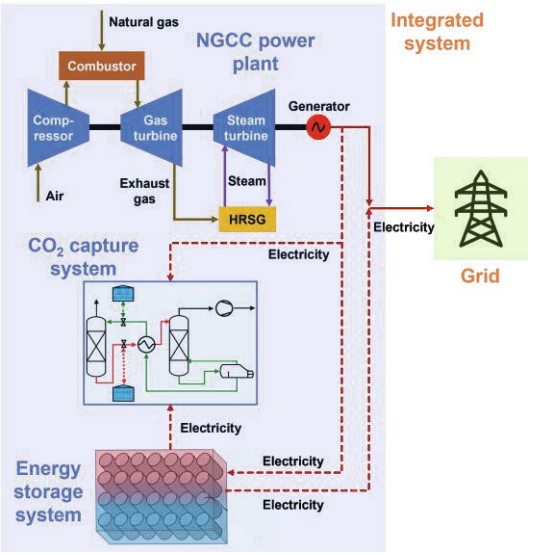


Figure 1: Integrated system schematic.

The integrated system is depicted in Figure 1 where it is connected to the electricity grid and is required to satisfy the time-varying electricity demand. The net demand is considered to incorporate the variability of renewable energy and is given by the total grid demand less the renewable generation. Four candidate storage technologies are considered for the storage block: mechanical energy storage through compressed air energy storage (CAES), thermal energy storage using cryogenic and molten salt-based high temperature storage (CES and HTTS, respectively), and electrochemical energy storage in the form of sodium sulfur (NaS) battery systems. For the CO_2 capture system, storage of the solvent enables time-varying operation of the energy-intensive solvent regeneration step and enables the flexible operation of the capture system. The power generated by the NGCC plant is used to meet the grid demand. A portion of this can also be used for CO_2 capture or stored in the energy storage for cases when the net demand is low from excess renewable availability. On the other hand, if the power plant output is insufficient to meet the grid demand, the energy stored in the storage system can be discharged to provide the required electricity.

The economic viability of this integration is influenced by the high investment cost of the energy storage and CO_2 capture systems. The candidate storage technologies also exhibit trade-offs between the various factors such as lifetime, efficiency and cost. It is crucial to determine the overall dynamic operation of the different systems to cost-effectively ensure that the time-varying grid demand is met while accommodating variable renewable energy. To consider these trade-offs and determine if it is profitable to invest in the integrated system under the spatio-temporal variability of electricity markets and renewable availability, we develop a mathematical programming-based simultaneous design and scheduling framework. This framework forms the back-end of our user-friendly software program: THESEUS (TecHno-Economic framework for Systematic Energy storage Utilization and downSelection). THESEUS enables the user to evaluate and compare the different energy storage alternatives for various demand profiles, region specific factors, and power plant types and conditions. Section 2 in this article presents the THESEUS program and the back-end optimization formulation. Section 3 depicts the framework demonstration to determine system design and operation for a case incorporating a sharp demand spike.

2. THESEUS Framework

The overall framework is depicted in Figure 2. From the user interface, the user can input the power plant parameters such as the cost parameters, ramping limits, nominal capacity and minimum load factor. The user can also input several region-specific parameters including the cost of electricity, ambient conditions, as well as specify a time-varying electricity demand profile to be met by the system. These inputs are combined with the power plant models and the technology models for the various direct and indirect storage

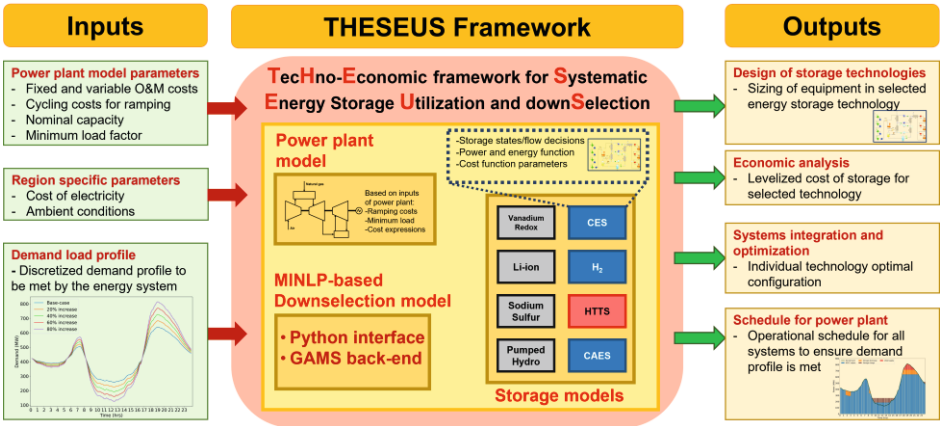


Figure 2: THESEUS software framework for energy storage and CO_2 capture (indirect storage) technology downselection.

systems in the back-end. Our mathematical programming-based optimization formulation which accounts for the trade-offs between system costs and flexibility comprises the unifying element connecting the different modules in an overall decision framework. Here, the problem statement is as follows: given a time-varying net electricity demand profile, determine the optimal storage technology, size and operation to integrate decentralized energy storage with existing fossil power plants for minimizing the overall system cost of meeting the demand. The general formulation of the optimization framework is given below:

$$\min TC = \sum_{i=1}^{NI}\left(C_i^{S,iv} + C_i^{S,of} + \sum_{t=1}^{NT}\left(C_{i,t}^{S,ov} + C_{i,t}^{FP,ov}\right) + C_t^{FP,rc} + C^{os} + C^{us} + C^{co2}\right) \tag{1}$$

$$P_t^{dem} = \sum_i P_{i,t}^S + P_t^{FP} - P_t^{os} + P_t^{us} - P_t^{co2} \tag{2}$$

$$|P_{t+1}^{FP} - P_t^{FP}| \le ro^{FP} P^{FP,nom} \Delta t \qquad \forall t \in T \setminus \{NT+1\} \tag{3}$$

$$P^{FP,nom} lf^{FP,min} \le P_t^{FP} \le P^{FP,nom} \qquad \forall t \in T \tag{4}$$

$$C_t^{ov,FP} = c4\, P_t^{FP} \eta^{FP,nom} / \eta_t^{FP} \qquad \forall t \in T \tag{5}$$

$$C_t^{FP,rc} = c5 |P_{t+1}^{FP} - P_t^{FP}| \qquad \forall t \in T \setminus \{NT+1\} \tag{6}$$

$$P_{i,t}^S = f1_i(s_{i,t},\, l_{i,t},\, z_{i,t}^b,\, x_i) \qquad \forall t \in T, \forall i \in I \tag{7}$$

$$E_{i,t} = f2_i(s_{i,t}) \qquad \forall t \in T, \forall i \in I \qquad (8)$$

$$-z^C_{i,t} P^S_{i,t} \leq y_i P_i^{max,C} \qquad \forall t \in T, \forall i \in I \qquad (9)$$

$$z^D_{i,t} P^S_{i,t} \leq y_i P_i^{max,D} \qquad \forall t \in T, \forall i \in I \qquad (10)$$

$$E_{i,t+1} = E_{i,t} - (\eta^S_i z^C_{i,t} + z^D_{i,t}) P^S_{i,t} \Delta t \qquad \forall t \in T \setminus \{NT+1\}, \forall i \in I \qquad (11)$$

$$E_{i,t=NT+1} = E_{i,t=1} \qquad \forall i \in I \qquad (12)$$

$$z^{idle}_{i,t} + z^C_{i,t} + z^D_{i,t} = 1 \qquad \forall t \in T, \forall i \in I \qquad (13)$$

$$0 \leq x_i \leq E_i^{max} y_i \qquad \forall i \in I \qquad (14)$$

$$0 \leq E_{i,t} \leq x_i \qquad \forall t \in T, \forall i \in I \qquad (15)$$

$$C_i^{S,iv} = c1_i(x_i) CRF_i \, T/8760 \qquad \forall i \in I \qquad (16)$$

$$C_i^{S,of} = c2_i \, (P_i^{max,D}) \, T/8760 \qquad \forall i \in I \qquad (17)$$

$$C_{i,t}^{S,ov} = c3_i(P^S_{i,t}) \qquad \forall t \in T, \forall i \in I \qquad (18)$$

The set $t \in T = \{1, 2, \dots, NT, NT+1\}$ denotes the set of time periods in the scheduling horizon, while the candidate storage technology set is given by: $I = \{ces, htts, caes, nas\}$. The time resolution is denoted by Δt, with T representing the time horizon length (in hrs). The optimization design decisions comprise of the selection y_i, energy capacity x_i, design discharging and charging power $P_i^{max,D}$ and $P_i^{max,C}$ respectively of storage technology i. The operational decisions for technology i includes the manipulating/flow variable $l_{i,t}$. In addition, the state of operation $z^{idle}_{i,t}$, $z^C_{i,t}$, $z^D_{i,t}$ i.e. if the technology is in the idle, charging or discharging state comprises the operational storage decisions. The power output from the power plant at time t, P_t^{FP}, is the plant-level operational decision. Here, the cost minimization objective given by Eq.(1) represents the sum of the storage investment cost $C_i^{S,iv}$, fixed operating cost $C_i^{S,of}$, variable operating cost $C_{i,t}^{S,ov}$, the fossil power plant operating cost $C_{i,t}^{FP,ov}$, plant cycling cost $C_t^{FP,rc}$, the electricity oversupply penalty C^{os}, the undersupply penalty C^{us}, and the cost of CO_2 capture C^{co2}. The optimization constraints consist of the following 3 categories: the grid-level constraints, the power plant model and energy storage model.

Eq.(2) represents the grid-level constraints and denotes the overall energy balance. Here, P_t^{dem}, $P^S_{i,t}$, P_t^{os} and P_t^{us} denote the grid power demand, power output from the storage, electricity oversupply and undersupply at time t, respectively. $P^S_{i,t}$ is positive when the storage is discharging and negative when charging. Eqs.(3)-(6) represent the power plant model. Here, $ro^{FP}, P^{FP,nom}$, $lf^{FP,min}, \eta^{FP,nom}, c4$ and $c5$ are constants representing the unit ramping rate of the power plant (%/MW.hr), nominal plant capacity (MW), minimum load factor (%), base-load efficiency (%), unit operational cost ($/MW.hr) and the unit cycling cost ($/MW.hr) respectively. Eqs.(7)-(8) represent the technology-specific models for the storage power $P^S_{i,t}$ and the storage energy capacity $E_{i,t}$ in terms of the storage state variable $s_{i,t}$, storage flow variable $l_{i,t}$, storage operational state $z^b_{i,t}$, and storage design x_i. Eqs.(9)-(15) represent the general operational model of each storage technology, while Eqs.(16)-(18) denote the general cost models. For these equations, η_i^S and E_i^{max} denote the storage efficiency and maximum possible energy capacity, while $c1_i, c2_i$ and $c3_i$ are constants denoting the unit storage investment, fixed operating and

variable operating costs respectively. The operational and cost models of the CO_2 capture system are adapted from Zantye et al., 2021.

3. Results and Discussion

We demonstrate THESEUS for an NGCC power plant of 641 MW nominal capacity to study if it is beneficial for the plant to invest in a co-located energy storage facility and/or a CO_2 capture system. We consider a scheduling time discretization of 5 minutes over a time horizon of one day and a net demand profile with a sharp spike in the evening hours. Furthermore, we consider a futuristic carbon pricing scenario with a CO_2 tax of \$80/ton and a selling price of \$35/ton. This scenario is found to be one of the economically viable cases to provide enough incentive for investment in carbon reduction technologies (Zantye et al., 2021).

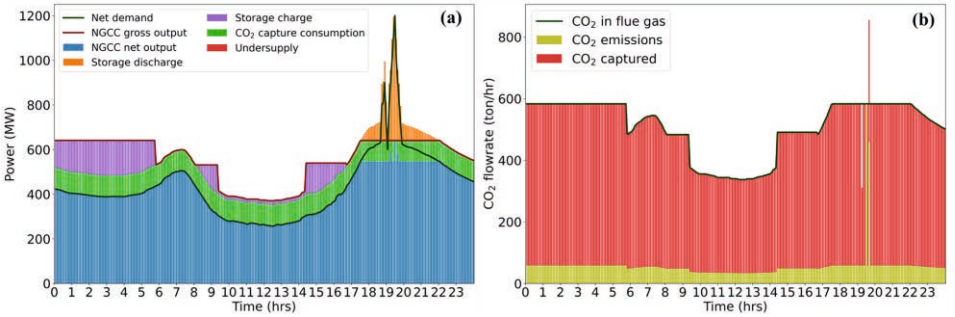


Figure 3: Optimal operational profiles for (a) the integrated system including power plant, CO_2 capture and energy storage, (b) CO_2 emission and capture.

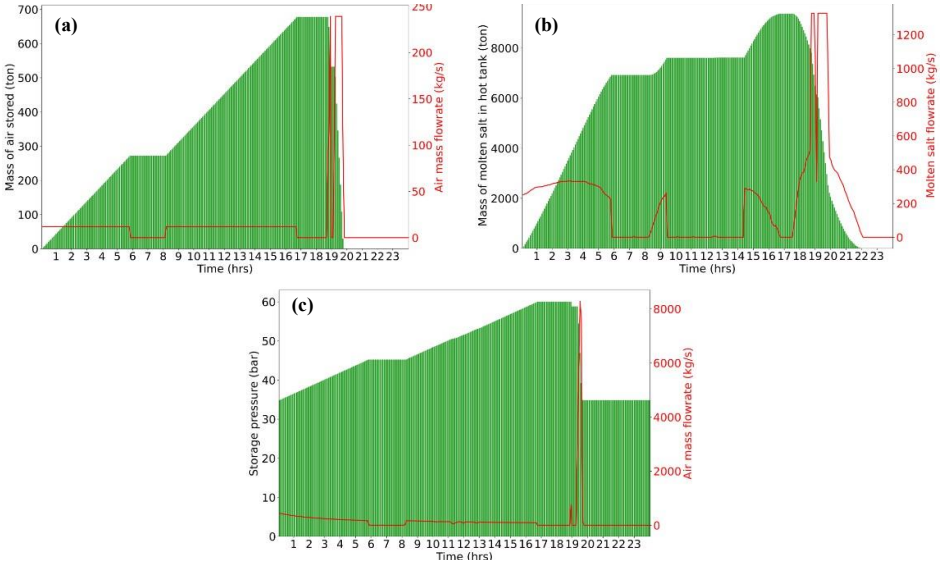


Figure 4: Optimal storage operational profiles for (a) CES technology, (b) HTTS technology, (c) CAES storage system.

THESEUS reports the optimal solution for the NGCC plant to invest in energy storage comprised of CES, HTTS and CAES technologies and a CO_2 capture system to minimize

the overall cost. The optimal storage integration size for CES is 100 MW/71 MWh, HTTS is 217 MW/386 MWh, and CAES is 242 MW/66 MWh. The optimal operational profile of the integrated system is demonstrated in Figure 3a. The operation of the CO_2 capture system is shown in Figure 3b. We observe that during periods of troughs in the net demand curve, the power plant produces additional power than the grid demand to charge energy in storage systems and in CO_2 capture. The stored energy is then discharged by the storage system to meet the demand spike during hours 17-22. From Figure 3b, we can see that the high CO_2 price ensures that the capture system is operating almost throughout the day, with the cumulative capture of 89%. The capture system also enables the decreased cycling and increased base-load operation of the power plant and reduces the overall cycling by 24% compared to a case without CO_2 capture. The optimal operational profiles for the selected CES, HTTS and CAES are shown in Figure 4. We find that the storage technologies slowly charge to their maximum capacities when excess energy is available during the day and are discharged almost instantaneously in the evening when the demand spike occurs. Among the three technologies, HTTS has the highest installed energy capacity with the lowest levelized cost of storage (LCOS) of $149/MWh. The LCOS for CAES is $302/MWh and CES is $345/MWh.

4. Conclusions

A decentralized integration schematic of energy storage and CO_2 capture systems with individual fossil power plants is proposed to address the challenges associated with power plant cycling and large-scale energy storage while accommodating variable renewable energy. The THESEUS framework can systematically determine both the integration and dynamic operational decisions. By extensively modelling the interactions between the system components in an overall optimization-based decision framework, THESEUS enables the user to evaluate the different integration alternatives, compare the costs and visualize the system operation. Demonstration of THESEUS for an NGCC plant under a futuristic carbon pricing scenario shows that the integration of both CO_2 capture and energy storage is optimal to reduce the power plant cycling by 24%, reduce emissions by nearly 90%, and meet a sharp demand spike. CO_2 capture is shown to act as an effective indirect energy storage system and enables increased base-load operation of the power plant under renewable integration.

5. Acknowledgements

The authors gratefully acknowledge support from U.S. Department of Energy (Grant number DE-FE0031771).

References

M. M. F. Hasan, R. C. Baliban, J. A. Elia, C. A. Floudas, 2012. Modeling, simulation, and optimization of postcombustion CO_2 capture for variable feed concentration and flow rate. 1. Chemical absorption and membrane processes. Industrial & Engineering Chemistry Research 51 (48), 15642–15664.

M. S. Zantye, A. Arora, M. M. F. Hasan, 2021. Optimal integration of renewables, flexible carbon capture, and energy storage for reducing CO_2 emissions from fossil power plants. In: Computer Aided Chemical Engineering. Vol. 50. Elsevier, pp. 1535-1540.

M. S. Zantye, A. Arora, M. M. F. Hasan, 2021. Renewable-integrated flexible carbon capture: A synergistic path forward to clean energy future", Energy & Environmental Science, 14, 3986 - 4008.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50349-3

Integration of experimental study and computer-aided design: A case study in thermal energy storage

Shoma Fujii[a*], Yuichiro Kanematsu[b], Yasunori Kikuchi[abc]

[a]*Institute for Future Initiatives, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku Tokyo, 113-8654, Japan*
[b]*Presidential Endowed Chair for "Platinum Society", The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku Tokyo, 113-8656, Japan*
[c]*Department of Chemical System Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku Tokyo, 113-8656, Japan*
shoma.fujii@ifi.u-tokyo.ac.jp

Abstract

The integration of experimental studies and computer-aided design strengths the coupling of the interfaces between different research scales. Using the mobile thermal energy storage system as a case study, we demonstrated the connection from the small scale of the material-level to the system-level researches. A full-scale conceptual design was carried out by numerical analysis including material properties that was validated by experiments, and the relationship between design parameters and performance indicators was summarized. By utilizing sensitivity and regression analysis using design parameters, inventory data that can be easily used in process flow models and system evaluations was generated. In addition, system hot spots were extracted from the system evaluation, and the summarized relationship between the design parameters and performance indicators provided feedback on the requirements for proceeding to the real site demonstration.

Keywords: Interdisciplinary approach, Life cycle assessment, Techno-economic analysis.

1. Introduction

To achieve a sustainable society, the time to adoption of emerging technologies must be minimized. To facilitate the adoption of technologies, the connectivity of interfaces between different research areas, and between scales from the material to the system level needs to be further enhanced. For example, integrated simulations of agricultural and industrial processes have made it possible to study the generation of by-products from different cultivars and the introduction of technologies to exploit them (Ohara et al., 2019; Ouchida et al., 2017). As for differences in scale, for example, in thermal energy storage technology, there were cases where bench-scale experiments (Nonnen et al., 2016) and numerical analysis (Mette et al., 2014a) were conducted based on material-level analysis (Mette et al., 2014b). However, now that lifecycle thinking is strongly required, it is necessary to consider new technological developments, combinations of mature technologies, and changes in socioeconomic and environmental conditions in design. This requires not only modeling of material properties, design of equipment incorporating material-level research, and process flow of the plant, but also

seamless integration of future system-level analysis such as life cycle assessment and techno-economic analysis in computer-aided design.

In this study, the integration of experimental studies and computer-aided design is demonstrated through a case study of a thermal energy storage system (Fujii et al., 2019) that can charge unused heat from the industrial waste heat and renewable resources, shift the heat in time or space, and release the stored heat according to the heat demand.

2. Material and Methodology

A mobile thermal energy storage (m-TES) system based on the water vapor ad/desorption cycle of zeolites can eliminate the spatial and temporal mismatch between unused heat generation and heat demand. Zeolites generate heat (heat of adsorption) when they capture water vapor (= heat discharging). The adsorbed water can be released by adding dry and heated air at the location of unused heat generation (=heat charging). The heat-charged zeolite can be packed in containers and transported by truck to surrounding heat demand areas. We studied the implementation of a m-TES system between a sugar mill and a food processing factory in Japan. Sugar mills usually burn sugarcane bagasse as fuel to generate steam for their demand. However, sugarcane bagasse is generated more than what is needed, therefore, more bagasse than necessary is burned in the bagasse boiler, emitting unused heat with a temperature of around 200 °C. On the other hand, the surrounding food processing factories consume fossil fuels for process steam generation. Industrial symbiosis of heat through m-TES systems has the potential for effective utilization of unused heat derived from biomass.

Scenario analysis, which considers various socioeconomic and environmental conditions and their combination with other technologies, plays an important role in the design of system. However, when conducting system-level analysis, inventory data extracted from lab-scale experimental results have large errors with real data. For example, in the case of a cylindrical packed bed reactor, the proportion of heat leakage from outside the walls to the total heat balance of the reactor is larger in a laboratory system with a smaller diameter than in a full-scale reactor. In this example, the inventory data extracted from the experimental results will underestimate the heat recovery efficiency. Therefore, it is necessary to predict the full-scale performance utilizing an experimentally validated numerical model and use the inventory data calculated from that simulation. However, it is difficult for researchers in system-level analysis and planners to build numerical simulation models and calculate inventories because it requires specialized skills in the target technology. In addition, when estimating the effects of horizontal deployment, combination with other technologies, and introduction under various socioeconomic and environment conditions, it takes an enormous amount of computational load and time to perform detailed numerical simulations at the reactor level for each condition. Using the development of the m-TES system as a case study, we extracted the essentials for the integration of experimental study and computer-aided system design, including experiments of material properties and their integration into numerical analysis, conceptual design of devices, demonstration experiments, and full-scale performance prediction by numerical analysis.

3. Case study in thermal energy storage

3.1. Experimentally validated computer-aided design

The requirements for the m-TES system were to be able to retrofit a heat charging system into the sugar mill and to be able to charge a large amount of heat continuously at the heat charging site, and to be able to generate process steam continuously at the heat discharging site, which had not been achieved so far. To meet these requirements, a zeolite moving bed reactor was adopted for both the heat charging (named heat charger) and discharging devices (named zeolite boiler). For the zeolite boiler, it was experimentally demonstrated that pressurized steam could be supplied continuously by adopting an indirect heat exchange process (Fujii et al., 2021). For the heat charger, a counter-current contact type moving bed was adopted, and it was demonstrated that continuous heat charging was possible (Arimoto et al., 2019). To incorporate the adsorption equilibrium, kinetics, and heat transfer model of the zeolite into the numerical analysis, a separate series of packed-bed tests were conducted. The developed numerical analysis was found to be able to simulate the temperature distribution in the packed bed, which can be used to predict the full-scale performance and optimize the design parameters. However, these numerical analyses require a lot of computational loads, and we also found that it is difficult to implement numerical analysis in both the process flow model of the plant, and all of the calculations for each case of life cycle assessment and techno-economic analysis as the system-level analysis.

3.2. Seamless connection to system-level analysis utilizing experimental data

The above simulation of the device for each condition of the scenario analysis takes a lot of computational load and time. Therefore, it is necessary for researchers of individual technologies to clarify the relationship between input parameters and output results (performance and required auxiliary power) in advance by sensitivity analysis of input parameters using the constructed numerical model. In addition, regression analysis of the input parameters and output results can reduce the computational load and time for system-level analysis. When researchers of individual technologies are studying and demonstrating concepts of new materials and devices, they can correctly reflect the information of the technology in the system-level analysis by conducting even regression analysis on a full-scale basis, instead of taking the laboratory-level results as the final result. In this way, researchers and system planners working on future system-level analysis can predict system performance based on experimentally validated numerical simulations while utilizing appropriate inventory data, seamlessly linking laboratory and future system predictions.

Figure 1 shows the schematic of the zeolite boiler and the relationship between design parameters and performance as an example. The zeolite boiler uses a moving bed and indirect heat exchange system. The heat charged zeolite is supplied from the top of the zeolite boiler, and the steam generated by the existing boiler is injected into the zeolite boiler. The zeolite adsorbs the injected steam and generates adsorption heat. The adsorption heat raises the temperature of the zeolite bed, and this heat is transferred to the feed water that is introduced from the bottom to the heat exchanger to produce pressurized steam for the process. The full-scale performance was predicted using numerical analysis, and the relationship between the design parameters and the performance was summarized by conducting sensitivity analysis with the design parameters. As an example, Figure 2 shows the sensitivity analysis for the initial water uptake on zeolite for the zeolite boiler. The lower the water uptake at input, the higher

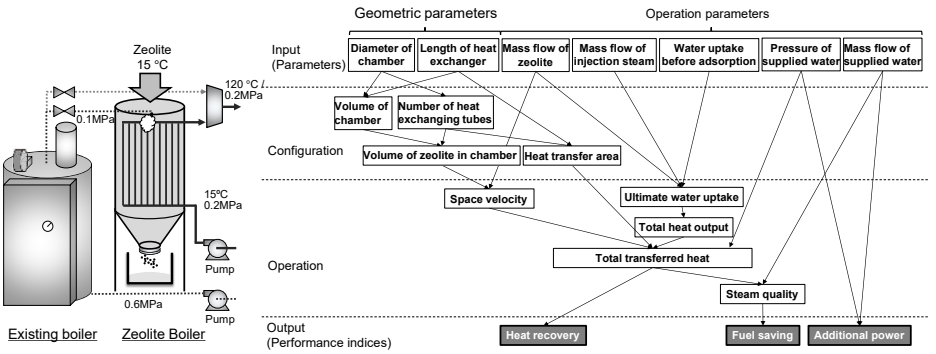


Figure 1 Schematic and linkage flow diagram of relationship among each parameter of zeolite boiler.

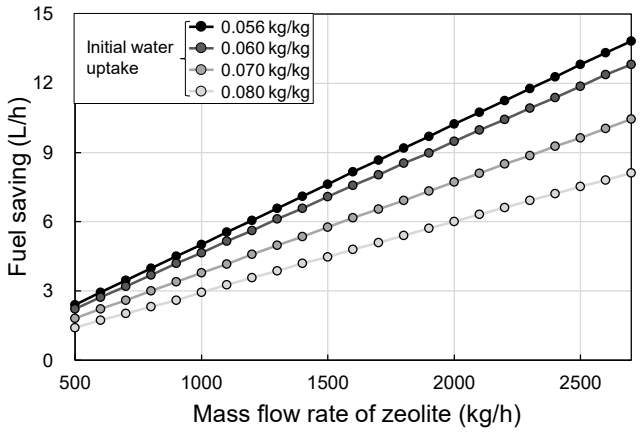


Figure 2 Effect of initial water uptake for zeolite boiler on fuel savings

the heat storage density, and thus the higher the fuel saving effect at the same scale. All point data are the result of optimizing other parameters to maximize the fuel saving based on the established design methodology using numerical analysis. The results show that the performance can be linearized with respect to scale at each initial water uptake, and this linear relationship was used in the system evaluation without going through complex numerical analysis. The same method was applied to other design parameters of the zeolite boiler and the heat charger, and the results were reflected in the process flow of the sugar mill and the heat demand for life cycle assessment and techno-economic analysis.

3.3. Research flow of integrating experimental study and system-level analysis

From the case of m-TES system, the integration of experimental study and computer-aided system design can be achieved by following procedure.

Material-level

-Testing the material properties and heat transfer characteristics

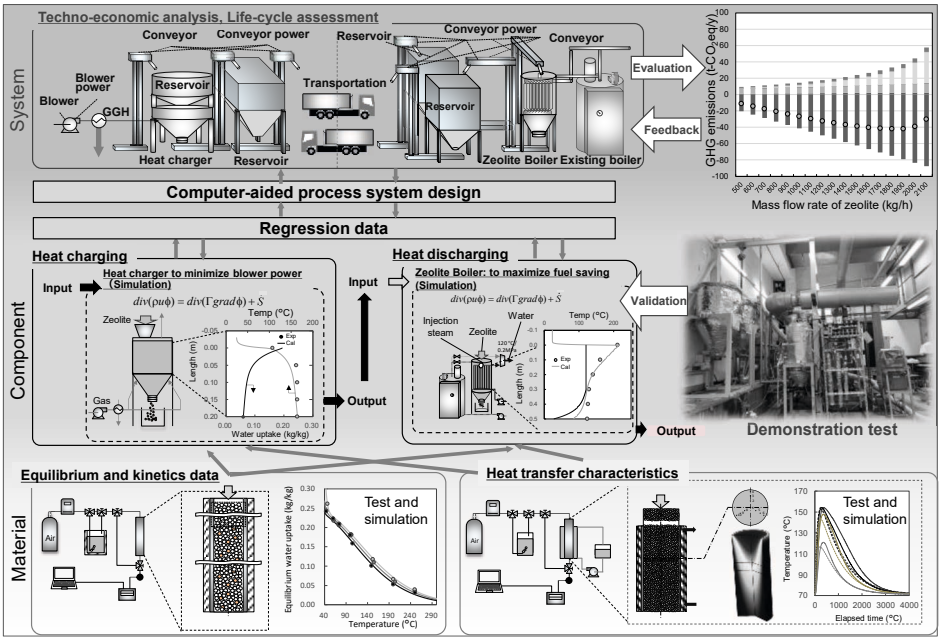


Figure 3 Research flow of seamless analysis from maerial to system level

Component-level

-Developing a numerical model that incorporates the material level information, and validate the model by demonstration tests
-Designing a full-scale equipment utilizing the validated numerical model, and predicting the performance, summarizing the relationship between design parameters and performance, and creating the regression data by sensitivity analysis

System-level

-Reflecting regression data in the process flow model of the plant
-Conducting system-level analysis such as life cycle assessment and techno-economic analysis utilizing inventory data obtained from the regression data.
-Identifying hot spots based on the results of the system-level analysis and feed them back to the target of technology in the reverse order above.

These are summarized in Figure 3. By creating this research flow, when a new material or device design is devised, the effect can be immediately evaluated from the system perspective by computer-aided system design, and it can also be used for scenario analysis in combinations of multiple technologies and various socioeconomic and environmental conditions.

4. Conclusions

In this study, a m-TES system with a zeolite water vapor ad/desorption cycle was used as a case study to extract the essentials for the integration of experimental studies and computer-aided system design. Using experimentally validated numerical analysis, sensitivity analysis, and regression analysis with design parameters of the equipment, computer-aided design (numerical analysis, process flow modelling, and future system-level analysis including environmental and socioeconomic parameters) can be

integrated with experimental studies, and each scale from material-level to system-level can be consistently linked.

Acknowledgement

This work was supported by Shinko Sugar Mill Co., Ltd., JSPS Grant-in-Aid for Research Activity Start-up (Grant Number 20K23360), JST COI-NEXT (Grant Number JPMJPF2003), and the Environment Research and Technology Development Fund (Grant Number JPMEERF20213R01 and JPMEERF20192010) of the Environmental Restoration and Conservation Agency of Japan. The activities of the Presidential Endowed Chair for "Platinum Society" at the University of Tokyo are supported by the KAITEKI Institute Incorporated, Mitsui Fudosan Corporation, Shin-Etsu Chemical Co., ORIX Corporation, Sekisui House, Ltd., East Japan Railway Company, and Toyota Tsusho Corporation.

References

N. Arimoto, Y. Abe, S. Fujii, Y. Kikuchi, Y. Kanematsu, T. Nakagaki, 2019, Performance prediction and experimental validation of heat charging device in thermochemical energy storage and transport system utilizing unused heat from sugar mill (in Japanese), Proceedings of the 24th National Symposium on Power and Energy Systems.

S. Fujii, N. Horie, K. Nakaibayashi, Y. Kanematsu, Y. Kikuchi, T. Nakagaki, 2019, Design of zeolite boiler in thermochemical energy storage and transport system utilizing unused heat from sugar mill, Applied Energy, 238, 561-571.

S. Fujii, Y. Kanematsu, Y. Kikuchi, T. Nakagaki, 2021, Effect of residual heat recovery of zeolite boiler in thermochemical energy storage and transport system, Proceedings of the 15th International Conference on Energy Storage ENERSTOCK

B. Mette, H. Kerskes, H. Druck, 2014a, Experimental and numerical investigations of different reactor concepts for thermochemical energy storage, Energy Procedia, 57, 2380-2389.

B. Mette, H. Kerskes, H. Druck, H. Muller-Steinhagen, 2014b, Experimental and numerical investigations on the water vapor adsorption isotherms and kinetics of binderless zeolite 13X, International Journal of Heat and Mass Transfer, 71, 555-561.

T. Nonnen, S. Beckert, K. Gleichmann, A. Brandt, B. Unger, H. Kerskes, B. Mette, S. Bonk, T. Badenhop, F. Salg, R. Glaser, 2016, A thermochemical long-term heat storage system based on a salt/zeolite composite, Chemical Engineering & Technology, 39, 2427-2434.

S. Ohara, Y. Kikuchi, K. Ouchida, A, Sugimoto, T. Hattori, T. Yasuhara, Y. Fukushima, 2019, Reduction of greenhouse gas emissions in the introduction of inversion system to produce sugar and ethanol from sugarcane (in Japanese), Journal of Life Cycle Assessment, Japan, 15, 86-100.

K. Ouchida, Y. Fukushima, S. Ohara, A. Sugimoto, M. Hirao, Y. Kikuchi, 2017, Integrated design of agricultural and industrial processes: A case study of combined sugar and ethanol production, AIChE J., 63, 560-581.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50350-X

Design support toolbox for renewable-based regional energy systems; The concept, data integration, and simulator development

Yuichiro Kanematsu[a*], Shoma Fujii[b], and Yasunori Kikuchi[a,b, c]

[a] *Presidential Endowed Chair for "Platinum Society", The University of Tokyo, Tokyo 113-8656, Japan*
[b] *Institute for Future Initiatives, The University of Tokyo, Tokyo 113-8654, Japan*
[c] *Department of Chemical System Engineering, The University of Tokyo, Tokyo 113-8656, Japan*
kanematsu@platinum.u-tokyo.ac.jp

Abstract

The design and implementation of renewable-based energy systems in various regions are increasingly necessitated. However, the data required are diverse and distributed across multiple ministries and local governments, and it often takes a considerable amount of time and efforts just to collect the basic data. Many studies regarding simulations of system designs for various resources and regions have been conducted, but they are not always reusable. In this study, we propose a design support toolbox integrated with databases for renewable-based regional energy systems, and the status of the development is introduced.

Keywords: Life cycle assessment, Open data, Decarbonization.

1. Introduction

Under the strong global demand for decarbonisation, there is an urgent need for action plans at national and local level. In Japan, though many municipalities have declared their carbon neutral by 2050, most of them do not have the clear plans of technology implementation. Renewable energy is the key technology for the decarbonization. The introduction of solar and wind power has been progressing, but because of their variability, energy storage and energy carriers are becoming increasingly important (Sinsel *et al.*, 2020). Biomass can be stored as the state of fuel, but there are difficulties in design of supply chain and whole energy system (Zahraee *et al.*, 2020). Toward the achievement of carbon neutral, complicated design for the combination of these technologies will be required. Many energy system analysis tools have been developed for various spatial and temporal targets (Ringkjøb et al, 2018), but very few of them can consider such emerging technologies.

In this study, a design support toolbox for renewable-based regional energy systems is proposed. Case studies of designing regional energy system were carried out to clarify the problems and barriers in design procedure. The requirement of the support toolbox is defined through building activity model and the data model of the system design. The status of tool development is also introduced.

2. Case studies of designing renewable-based regional energy systems

To clarify the problems and barriers in designing renewable-based regional energy systems, we had been carried out some case studies of design and evaluation of them. A case of designing combined cooling, heating, and power system using local woody biomass toward sustainable forestry (Kanematsu *et al*., 2017a) was analysed. As shown in Figure 1, the design and evaluation procedure mainly consist of "Data collection", "Alternative generation", "Simulation", and "Evaluation". Data collection was required for identify and quantify the local resources and energy demand. In the alternative generation, the system flow diagrams are drawn, similar to the process flow diagrams for chemical processes, for multiple alternatives. To enable the simulation of material and energy balances, simulation modules were developed for the elemental technologies, e.g., power plant, district heating and cooling system, wood chipping machine, transport, and so on. Estimations for hourly variations in energy demand of each consumer based on a published knowledge was also required for executing simulation considering supply-demand balance. As the system evaluation, life cycle assessment (LCA) was carried out using the simulation results as inventory data. As another case study, industrial symbiosis around sugar mill using the excess heat and bagasse (Kikuchi et al., 2016) was also carried out, and the case had the similar structure.

We identified the problems that such design and evaluation could be carried out as research activity, but it was time consuming, labour intensive and requires specialized knowledge and its application. In order to achieve rapid social implementation, it needs to be possible to implement it in different multiple areas in a shorter time. Just as process simulators and chemical databases have contributed to speeding up and saving labour in chemical process design, tools such as simulators and integrated databases will be essential for rapid design of regional energy system.

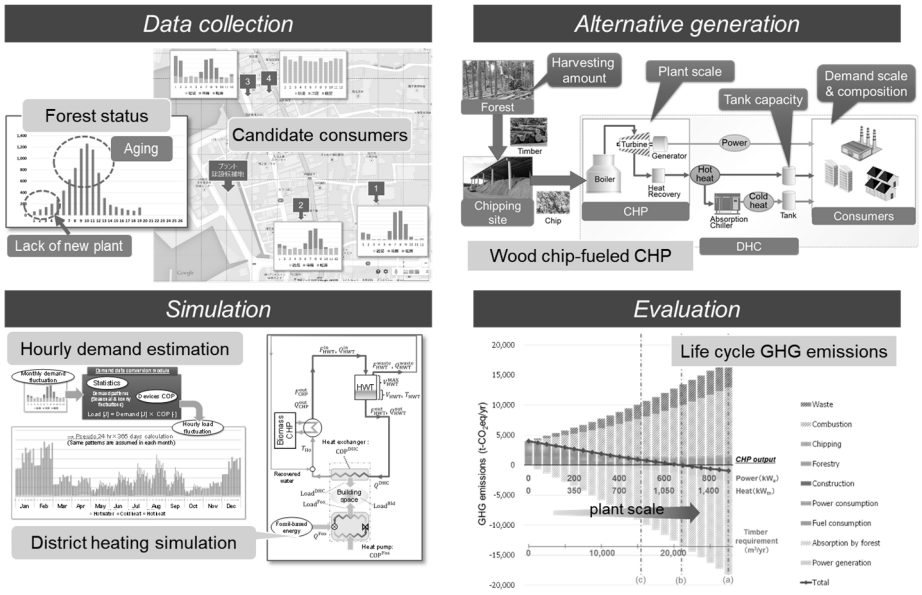


Figure 1 Overview of the case study of designing combined cooling, heating, and power system using regional woody biomass (compiled from Kanematsu et al., 2017a)

3. Requirement definition of the design support toolbox

Activity model and data model were built to define the requirements of the design support toolbox for renewable-based regional energy systems by re-analysing the abovementioned case studies.

3.1. Activity model

The designing procedure was visualized by activity model in order to clarify how we should support what activities. IDEF0 functional modelling method was employed for expressing the activity model. In IDEF0 method, an activity is represented by a box and information is represented by four types of concepts (Input, Output, Control and Mechanism), each of which is represented by an arrow with a different direction toward the activity box. The activities were expressed from the viewpoint of the designing working group as in Figure 2. The main designing tasks consists of [Examine present system], [Generate alternatives], [Simulate flows in alternatives] and [Evaluate alternatives]. Draft of proposals are created from these designing tasks, and it will be reviewed among the decision makers in the [Review proposals] activity.

In this activity modelling, we especially focused on the Mechanisms, which enable the execution of the activities including tools, human resources, and datasets. We identified the tools that should be developed in order to carry out activities with less time and effort. The required functions of these tools were identified by carefully examining the Input, Output and Control of each activity.

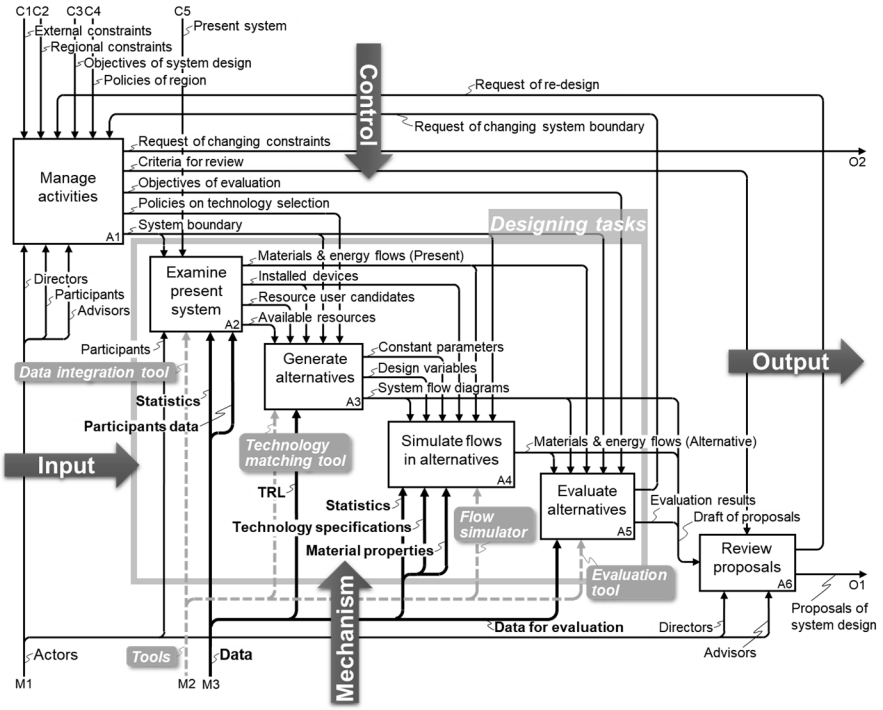


Figure 2 IDEF0 activity model: the activity of "Propose the design of renewable-based regional energy systems" (modified from Kanematsu *et al*., 2017b)

3.2. Data model for flow simulation

A data model was also built to clarify the types of data required and the linkages between them in executing the simulation. The step of simulation was focused because it was revealed to be the central activity in the design procedure from the viewpoint of data processing. Data model can also be used as the conceptual design diagram of required database for the simulator. UML class diagram was applied for expressing the data model as shown in Figure 3. The flame of "Participant" means the entities which will participate in the regional energy system to be developed, e.g., resource providers, energy suppliers, local industries, consumers, and so on. The participant is described as "Prosumer" because it can be both of consumer and producer of resources. Participant can be simply expressed which produces output from input via conversion devices. For example, sugar mill produces raw sugar, bagasse, molasses, heat, power, and other by-products from sugarcane via sugar milling processes includes in-house cogeneration system. In actual simulation, multiple modules that functionalize the participants are connected to each other, and the energy and material balance can be calculated.

3.3. The function of design support toolbox

The required functions of design support toolbox were identified and defined through the re-analysis of the activity and data models. We have developed the concept of the toolbox as in Figure 4. The toolbox consists of the tools that support the respective activities and the databases linked with these tools.

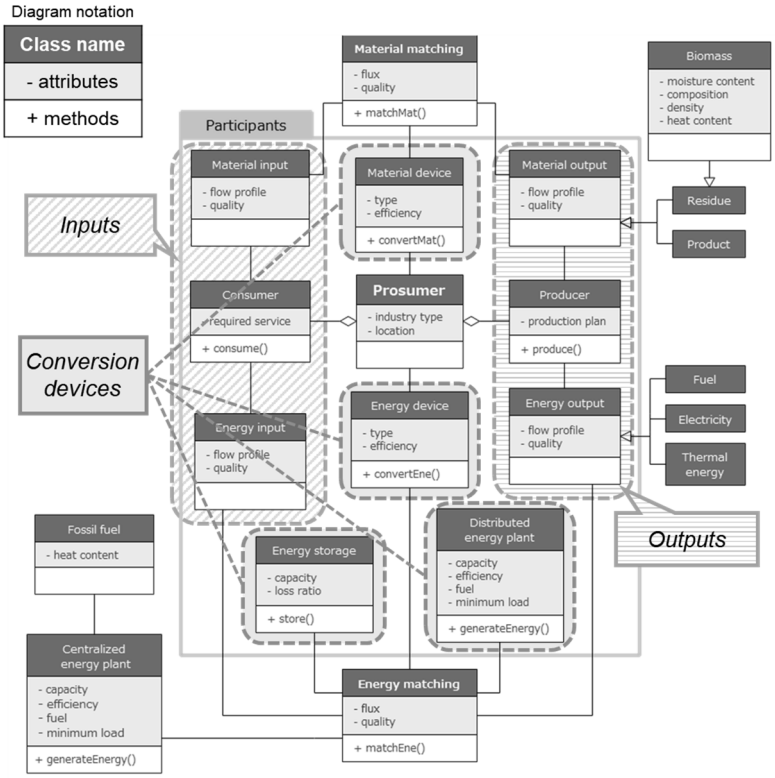


Figure 3 Data model for the simulation of renewable-based regional energy system

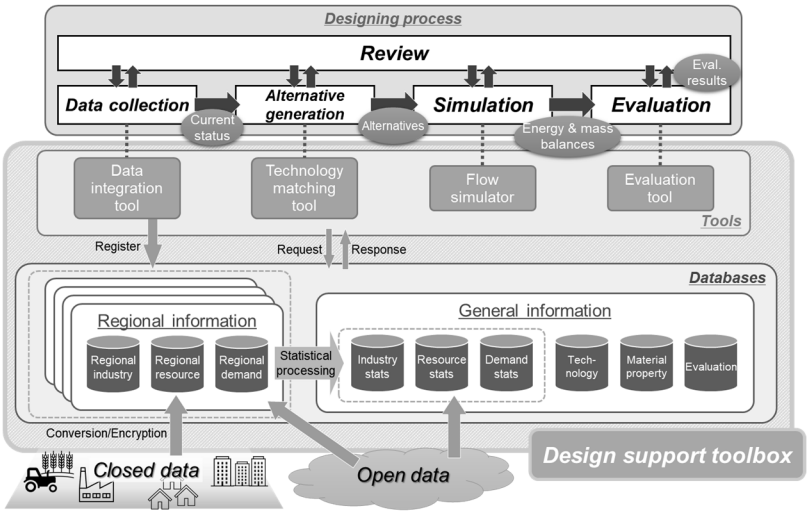


Figure 4 The concept of the design support toolbox for renewable-based regional energy systems

4. Development of modules in the toolbox

4.1. Data integration and visualization of resources and demands

A web-based application that enables the semi-automation of collection, analysis, and visualization of scattered data sources for regional resources was developed as a prototype. Open data originally organized by different ministries were integrated, such as renewable energy potential and the installation status of renewables under the Feed-in-Tariff scheme. An algorithm to estimate the energy demand for each municipality is also developed and will be combined to this application.

4.2. Matching between resources and technologies

A matching system integrated with technology database is conceptually designed under the cooperation with an association of engineering enterprises. The system can search conversion technologies that can produce value-added products from the local resources that found in the stage of data collection. This system can connect the novel technologies and local players that have not been known each other well.

4.3. Simulation of the material and energy balances

By applying the developed woody biomass simulator to other area, the time required was reduced by a factor of 20 compared to the first case. Additionally, we developed and are combining the simulators for thermal energy storage (TES) and transportation with zeolite (Fujii *et al*., 2019) and the technologies for massive installation of variable renewable energy such as battery-assisted solar-derived hydrogen production (Kikuchi *et al*., 2019, Sako *et al*., 2020) and wind-TES (Yamaki *et al*., 2020).

4.4. Environmental and socio-economic evaluations

Once the material and energy flows in the region with future installation of the new technologies could be calculated by the simulator, evaluations become executable by combining them with the data for evaluations such as life cycle inventory databases or input-output table available from national or local government. The integration mechanism of these data is under development.

5. Conclusions

The required functions of design support toolbox for renewable-based regional energy systems were defined through activity and data modelling, and part of them are developed. Because the workload of the development and continuous update will be huge, we are starting the investigation of the scheme for co-creation of this toolbox which involves technology developers, local industries, and system integrators. By co-creating this toolbox, co-creation of regional systems can be strongly supported.

Acknowledgement

This work is supported by MEXT/JSPS KAKENHI (21K17919), JST COI-NEXT (JPMJPF2003), and Environment Research and Technology Development Fund (JPMEERF20192010) of ERCA Japan. Presidential Endowed Chair for “Platinum Society” of UTokyo is supported by the KAITEKI Institute Inc., Mitsui Fudosan Co., Ltd. , Shin-Etsu Chemical Co., ORIX Corporation., Sekisui House, Ltd., the East Japan Railway Company, and Toyota Tsusho Corporation.

References

S. Fujii, N. Horie, K. Nakaibayashi, Y. Kanematsu, Y. Kikuchi, T. Nakagaki, 2019, Design of Zeolite Boiler in Thermochemical Energy Storage and Transport System Utilizing Unused Heat from Sugar Mill, Applied Energy, 238, 561-571

Y. Kanematsu, K. Oosawa, T. Okubo, Y. Kikuchi, 2017a, Designing the Scale of a Woody Biomass CHP Considering Local Forestry Reformation: A Case Study of Tanegashima, Japan. Applied Energy, 198, 160–172

Y. Kanematsu, T. Okubo, Y. Kikuchi, 2017b, Acitivity and Data Models of Planning Processes for Industrial Symbiosis in Rural Areas, Kagaku Kogaku Ronbunshu, 43 (5), 347-357

Y. Kikuchi, T. Ichikawa, M. Sugiyama, M. Koyama, 2019, Battery-assisted low-cost hydrogen production from solar energy: Rational target setting for future technology systems, International Journal of Hydrogen Energy, 44, 1451–1465

Y. Kikuchi, Y. Kanematsu, M. Ugo, Y. Hamada, T. Okubo, 2016, Industrial Symbiosis Centered on a Regional Cogeneration Power Plant Utilizing Available Local Resources: A Case Study of Tanegashima. Journal of Industrial Ecology. 20 (2), 276–288

H.-K. Ringkjøb, P. M. Haugan, I. M. Solbrekke, 2018, A review of modelling tools for energy and electricity systems with large shares of variable renewables, Renewable and Sustainable Energy Reviews, 96, 440-459,

N. Sako, M. Koyama, T. Okubo, Y. Kikuchi, 2021, Techno-Economic and Life Cycle Analyses of Battery-Assisted Hydrogen Production Systems from Photovoltaic Power, Journal of Cleaner Production, 298, 126809

S. R. Sinsel, R. L. Riemke, and V. H. Hoffmann, 2020, Challenges and solution technologies for the integration of variable renewable energy sources—a review, Renewable Energy, 145, 2271–2285

A. Yamaki, Y. Kanematsu, Y. Kikuchi, 2020, Lifecycle Greenhouse Gas Emissions of Thermal Energy Storage Implemented in a Paper Mill for Wind Energy Utilization. Energy, 205, 118056

S. M. Zahraee, N. Shiwakoti, P. Stasinopoulos, 2020, Biomass Supply Chain Environmental and Socio-Economic Analysis: 40-Years Comprehensive Review of Methods, Decision Issues, Sustainability Challenges, and the Way Forward. Biomass and Bioenergy, 142, 105777

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50351-1

Circular Economy Integration into Carbon Accounting Framework for Comprehensive Sustainability Assessment

Nasyitah Husniyah Mahbob[a], Haslenda Hashim[a,b*]

[a] *School of Chemical and Energy Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia*
[b] *Process Systems Engineering Malaysia (PROSPECT), School of Chemical and Energy Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia*
haslenda@utm.my

Abstract

Circular economy is an approach to develop economy without giving harms to the social and environment. It uses 'cradle-to-cradle' concept which design out pollution and waste while keeping materials and products in loop for as long as possible. This economy concept which is also in line with several SDG goals such as ecosystem restoration, responsible consumption, as well as climate action makes it more preferrable than the old 'cradle-to-grave' concept. There are a lot of circular economy actions that can be adopted by the industry as an effort in shifting towards circular economy. Therefore, a systematic framework to aids the industry in selecting appropriate circular economy actions is needed. In this paper, a framework for selection of circular economy actions is proposed by considering its environmental and economic impact. This framework integrates Total Circularity Index (TCI) and integrated carbon accounting and mitigation framework (INCAM) to assess the environmental impact of the actions. The practicability of the framework is illustrated through a relevant palm oil mill case study. The results show that the application of this framework enables the industry to select circular economy actions that is appropriate to its environmental and economic status.

Keywords: circular economy, carbon mitigation, sustainability, decision-making framework

1. Introduction

The world is currently facing climate change where the temperature rises leading to other consequences such as melting of glaciers and more frequent occurrence of natural disaster. Based on The Sustainable Development Goals Report 2020 released by United Nations (2020), 2019 was the second warmest year on record and the global temperature are projected to rises to 3.2 °C by 2100. The rises in Earth's temperature are due to the increasing amount of greenhouse gases emission and according to data by Climate Watch (2021), global greenhouse gases emission gradually increases each year, and it reaches 48.94 Gt CO2e in 2018.

Numerous efforts have been made to tackle climate change at various level. Paris Agreement which is a legally binding agreement that was adopted by many countries in 2015 is one of the efforts done to fight climate change at global level. This agreement

aims to limit the global warming at 1.5 °C and many countries have pledged to reduce their greenhouse gases emission through their nationally determined contributions (NDCs) to achieve this target. Apart from that, utilization of renewable energy and energy efficiency program are also part of the efforts done to tackle climate change since more than 50 % of the global greenhouse gases emission comes from energy sector (Climate Watch, 2021).

In addition, shifting towards circular economy can also be one of the ways to tackle climate change. Circular economy is an approach that uses 'cradle-to-cradle' concept which aims to decouple economic growth from consumption of finite resources. It is built to design out waste and pollution and elongate the life cycle of products and materials while regenerating natural system. Therefore, circular economy can help reduce the greenhouse gases emission by eliminating its sources (waste, pollution, and manufacturing process) and providing more carbon sinks. This statement can be proven through a study by Ramboll et al. (2020) that shows 61% of GHG emissions across European Union can be reduced by implementing circular economy actions (CE actions). Research done by Ellen MacArthur Foundation (2019) also discussed the role of circular economy in tackling climate change.

There are numerous CE actions that can be applied in an economic system ranging from utilization of renewable resources to reuse or recycling of materials and products. Therefore, a methodology to select suitable circular economy actions is required to assist the industry in decision-making process. A study conducted by Ramboll et al. (2020) has proposed a method to select and assess circular economy actions and its impact on climate change mitigation. A generic methodology to quantify the potential CO2 emission reduction was developed by integrating the life cycle analysis (LCA) and material flow modelling. The methodology is designed to be applicable to all economic sectors. However, this methodology can be challenging for some users as it requires knowledge and experiences in LCA and material flow modelling. Moreover, uncertainty and biodiversity in LCA also makes it more challenging as LCA studies depend on assumptions and scenarios.

The aim of this study is to propose an easy and systematic framework for the industry to select suitable circular economy actions by considering its economic and environmental impacts. This paper integrates the Total Circularity Index (TCI) and integrated carbon accounting and mitigation framework (INCAM) to assess the environmental impacts of the circular economy actions. The methodology proposed enables the industry to find the hot spots for carbon emissions in their plant or organization as well as selecting befitting circular economy actions based on its economic and environmental impacts.

2. Circular Economy Actions Selection Framework

According to the framework for selection of CE actions shown in Figure 1, the selection of CE actions is made based on three criteria which are the emission reduction, its circularity as well as the payback period. There are three main steps in this framework which are: (i) identifying the hot spots for GHG emissions, (ii) proposing possible CE actions, (iii) evaluating and ranking of CE actions. To identify hot spots for GHG emissions, an integrated carbon accounting and mitigation (INCAM) framework by Hashim et al. (2015) will be applied. Based on INCAM, the methodology to identify hot spots include defining Carbon Accounting Centre (CAC), developing carbon checklist, and calculating carbon emission index (CEI) and carbon emission profile (%) of each CAC. CAC with the highest carbon emission profile will be the hot spots.

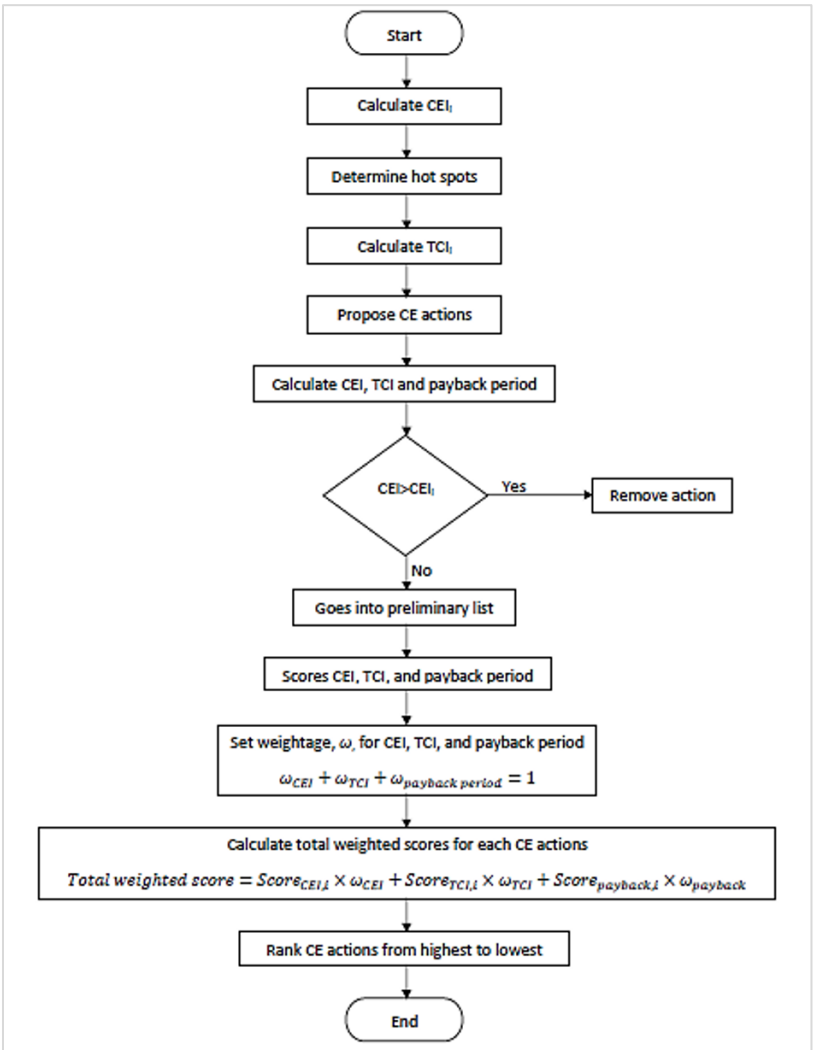


Figure 1: Framework for Selection of Circular Economy Actions

After the hot spots have been identified, total circularity index (TCI) of the hot spots will be calculated. TCI is an indicator to measure the circularity of a process route which consider both material and energy aspects. Circular Material Use (CMU) is used to calculate the circularity of the material by dividing the product produced through secondary materials with the total demand of the products. On the other hand, energy aspects will be measured through Circular Exergy Use (CEU) which is the amount of secondary exergy produced over the exergy demand. CMU and CEU will be multiplied by its respective weightage before being added up to get TCI. The TCI value will be between 0 and 1 with the latter represents full circularity of the process route. CE actions is then proposed accordingly.

CEI, TCI and payback period of each CE actions proposed will be estimated. Then, CE actions with higher CEI than CEI_i, which is the baseline, will be eliminated from the options. The CEI, TCI, and payback period of the remaining CE actions will be scored

based on scoring scales. The weighted scores of each criterion are then calculated by multiplying the scores with its respective weightage which are set according to its importance. The weighted scores of each criterion will be summed up to get the total weighted scores which will be used to rank the CE actions proposed. The priority of the CE actions is set from highest to lowest total weighted scores.

3. Case Study

The effectiveness of this framework is demonstrated through a palm oil mill case study. The mill processes 120 t fresh fruit bunch (FFB) per hour into crude palm oil (CPO). The overall process of this mill can be divided into several sections which are palm oil extraction, oil recovery, kernel recovery, effluent treatment, and boiler.

3.1. Identify Hot Spots Through INCAM

The whole process is break into five CACs, (i) oil palm extraction, (ii) oil recovery, (iii) kernel recovery, (iv) effluent treatment, (v) boiler. Each CAC have several sub-CACs. The source of emission of each CAC is identified and the monthly generation/consumption of relevant carbon performance indicator (CPI) of each CAC has been collected. The monthly emission, carbon emission profile, and CEI of each CAC has been calculated. Figure 2 and Figure 3 show the carbon emission profile of each CAC and each CPI respectively. From the analysis, the top three carbon emission are from electricity, fuel consumption and wastewater with CPI of 68 %, 28 % and 3 % respectively. The highest emission is in CAC 5 which is the boiler followed by CAC 2, palm oil extraction section, which has a lot of equipment that consumes electricity.

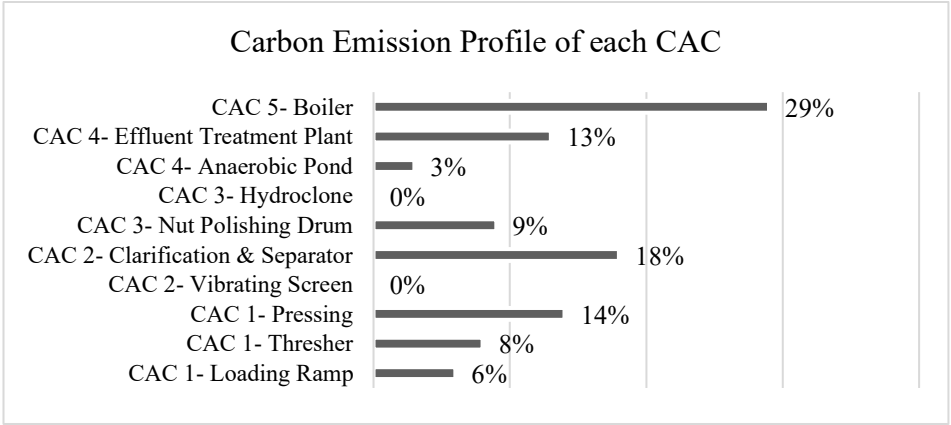


Figure 2: Carbon Emission Profile of each CAC

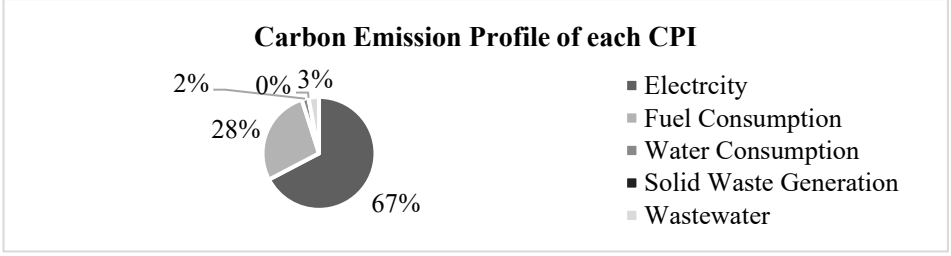


Figure 3: Carbon Emission Profile of each CPI

3.2. Propose CE Actions and Priority Ranking

Before proposing CE actions, TCI of the palm oil is calculated in order to identify any circularity actions that has been implemented in the mill. In this case study, the demand for product and exergy used in CMU and CEU calculation is the amount of product and exergy demand by the palm oil mill and its plantation. The total exergy demand is 3086 GJ/month. From the calculation, the CMU, CEU and TCI of the palm oil mill is all 0. This means that there is no circular action has been implemented in the mill. After measuring the circularity of the mill, CE actions to reduce carbon emissions are proposed according to the hot spots identified earlier. The CE actions proposed will focus on reducing electricity and fuel consumption as well as reducing emissions from effluent and boiler. There are three CE actions proposed which are electricity generation using biogas from POME, change boiler fuel from diesel to biogas, and production of fertilizer from POME. The contributions - CEI, TCI and payback period of all CE actions proposed has been estimated and summarized in Table 1. Since all actions proposed has lower CEI than the baseline CEI, all actions proposed will be considered for implementation.

The contributions of each CE actions are scored based on scoring scales. For this case study, the weightage for emission reduction, TCI and payback period has been set at 0.4, 0.2, and 0.4 respectively. This is because the mill wants to prioritize reducing emission with low costs. Table 2 shows the priority rank of the CE actions according to the total weighted scores. From the analysis, it shows that production of fertilizer from POME should be implemented first since it has the highest total weighted scores compared to other options. This option can reduce 13,325 kgCO2e/month and improves the TCI of the mill by 50 %. It also has the lowest payback period which is 0.84 years. The second highest ranking is to change boiler fuel from diesel to biogas which has average scores in all criteria. The third priority is to generate electricity using biogas from POME. Although this option can reduce the most carbon emissions, 309,725 kgCO2e/month, it requires high payback period which is 7.64 years and has the lowest TCI improvement which makes it less favorable.

Table 1: CEI, TCI and payback period of CE actions proposed

CE Actions	Emission (kgCO2e/month)			Total Circularity Index		Payback Period (year)
	Baseline	After	Reduction	Before	After	
Biogas from POME to generate energy	437,462.46	127,737.43	309,725.03	0	0.23	7.64
Change boiler fuel to biogas (POME)		313,850.36	123,612.10		0.27	1.71
POME to fertilizer		424,137.43	13,325.03		0.50	0.84

Table 2: Scores and priority ranking for CE actions proposed

CE Actions	Scores							**Priority Ranking**
	Emission Reduction	Weighted Score (0.4)	TCI	Weighted Score (0.2)	Payback Period	Weighted Score (0.4)	**Total Weighted Scores**	
Biogas from POME to generate energy	4	1.6	2	0.4	2	0.8	**2.8**	**2**
Change boiler fuel to biogas (POME)	2	0.8	2	0.4	4	1.6	**2.8**	**2**
POME to fertilizer	1	0.4	3	0.6	5	2.0	**3**	**1**

4. Conclusions

In this paper, a framework to select circular economy actions has been proposed. This framework aims to ease the decision maker to choose which CE actions to be prioritized and implemented by considering the emission reduction, circularity, and the payback period. The practicality of this framework has been demonstrated through a palm oil mill case study and the result shows that this framework is able to guide decision-maker in choosing mitigation actions that can improve carbon performance, circularity, and economy. Improvement that can be made for this framework is by considering other constraints in the selection process such as the reduction target and financial budget.

5. Acknowledgement

The authors would like to gratefully acknowledge UTM University Grant Vot. No. Q.J13000.3009.02M81 and Q.J130000.2409.08G96 for the financial in completing this project.

References

Climate Watch. (2021). Historical GHG Emissions. Retrieved from Climate Watch: https://www.climatewatchdata.org

Ellen MacArthur Foundation. (2019). Completing the Picture: How the Circular Economy Tackles Climate Change www.ellenmacarthurfoundation.org/publications

H., Hashim, M. R., Ramlan, L. J., Shiun, H. C., Siong, H., Kamyab, M. Z., Majid, & C. T., Lee (2015). An Integrated Carbon Accounting and Mitigation Framework for Greening the Industry. Energy Procedia, 2993-2998.

Ramboll, Fraunhofer, & Ecological Institute. (2020). The Decorbonisation Benefits of Sectoral Circular Economy Actions.

United Nations. (2020). The Sustainable Development Goals Report 2020. New York: United Nations Publications.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50352-3

Design and analysis of fuel-assisted solid oxide electrolysis cell combined with biomass gasifier for hydrogen production

Shih-Chieh Chen, Jyh-Cheng Jeng*

Department of Chemical Engineering and Biotechnology, National Taipei University of Technology, Taipei 10608 Taiwan
jcjeng@ntut.edu.tw

Abstract

To solve environmental problems, the development of alternative energy sources is becoming more and more important. Hydrogen is a very high potential energy carrier because it has a higher energy density compared to other energy carriers and clean in use. This study proposes a novel hydrogen production process that combines a biomass gasifier and a fuel-assisted solid oxide electrolysis cell (SOFEC). The syngas fed to the anode of SOFEC for fuel-assisted electrolysis is formed by gasification of biomass. The advantages of combining these two systems are that the operating temperatures are similar to each other, which reduces energy waste, and that part of the electrical energy used in electrolysis is replaced by the chemical energy of the fuel, which significantly reduces the demand for external power supply during electrolysis. In addition, the thermal energy of the high-temperature flow from the exhaust of SOFEC can be used to preheat the feed. In this study, we consider the effect of gasifier operating parameters, including gasification temperature, dried biomass moisture, and equivalence ratio on system. The results show that there is a positive effect on the system when the gasification temperature and the moisture content of the biomass are higher; With different gasifier operating conditions, the efficiency of the system is highest when the equivalence ratio is between 0.15 and 0.25, mainly depending on the amount of syngas production from the gasifier.

Keywords: Hydrogen production, Gasifier, Fuel-assisted solid oxide electrolysis cell, System simulation.

1. Introduction

Since the industrial revolution, fossil fuel combustion has been the main method of energy generation, contributing the largest proportion of global energy demand. However, fossil fuels are non-renewable resources, making them more and more scarce in the future and bound to face problems such as energy depletion, and greenhouse gas emissions from fossil fuel use already have a serious impact on global warming and air pollution. Therefore, governments are actively looking for alternative energy sources. Considering the economic growth and energy demand, as well as the environmental issues, the research on renewable energy and its storage has been increasing rapidly in recent years.

Although producing hydrogen through fossil fuels is the most efficient and economical way to produce hydrogen, the production results in greenhouse gas emissions that cause environmental pollution. Therefore, the generated hydrogen cannot be regarded as green hydrogen. Hydrogen production through renewable energy is the way to produce green hydrogen, and in the medium term, the solid oxide electrolysis cell is the most potential

solution for the development of green hydrogen production technology. Therefore, this study proposes a hydrogen production system using syngas generated from biomass gasifier to assist the solid oxide electrolysis cell, replacing part of the electrical energy in electrolysis with chemical energy from syngas and thus reducing the electrical energy demand in the electrolysis of hydrogen. In this study, the impact of the operating parameters of the gasifier on the system efficiency is investigated.

2. Models of Gasifier and Fuel-Assisted Solid Oxide Electrolysis Cell

2.1. Gasifier model

When fresh biomass enters the gasifier for gasification reaction, the process is divided into drying zone, decomposition zone, gasification zone, and sepearation zone. These sections are realized through different modules in Aspen Plus®, as the model shown in Figure 1. First, the biomass source will enter the RStoic reactor and dry the biomass with the temperature set at 150°C as the drying temperature. After drying, the moisture content of biomass will decrease and become dry biomass and then enter the RYield reactor to simulate the pyrolysis of biomass in which the reaction temperature is set at 550°C. After the pyrolysis reaction, biomass will enter the gasification reactor (Rgibbs) in the form of elements, and additional air will be used as the gasification agent. The reacted syngas will be separated from ash and carbon through SEPARATE unit to simulate the ash and syngas separation section of the gasifier, and H2O-SEP is used to separate the water from the biomass to calculate the dry basis of biomass syngas (Tavares et al., 2020).

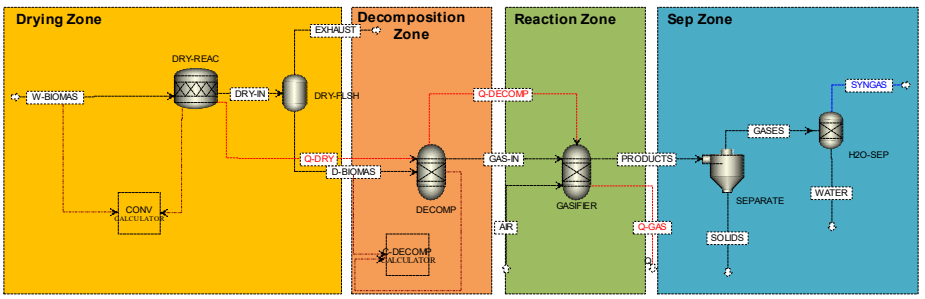


Figure 1 Gasifier Model in Aspen Plus®.

2.2. SOFEC model

The actual voltage of an electrolysis cell during electrolysis is expressed by the equivalent voltage added to its irreversible loss. Voltage is the most important parameter that affects the performance of electrolytic cells. Therefore, the amount of polarization will affect the power consumption of the entire electrolytic cell :

$$V_{SOFEC} = V_{SOFEC}^{OCV} + \eta_{act} + \eta_{ohm} + \eta_{conc} \tag{1}$$

Because the electrochemical reactions in the SOFEC cathode and anode are opposite reactions, there is no standard potentials of electrochemical reactions in the calculation of open-circuit potential, but the quotient of the component concentrations is retained according to the Nernst equation. The equilibrium voltage is as shown in Eq.(2) where R is the ideal gas constant (8.314 J/mole K), T is the operating temperature (K), and $P_{H2,ca}$, $P_{H2O,ca}$, $P_{H2,an}$, $P_{H2O,an}$ are the surface partial pressure of each component at the inlet of the cathode and anode sides, respectively (Salzano et al., 1985).

$$V_{SOFEC,Nernst} = \frac{RT}{2F}\ln\left(\frac{P_{H_2,ca}P_{H_2O,an}}{P_{H_2O,ca}P_{H_2,an}}\right) \tag{2}$$

For the calculation of the actual voltage in the electrolysis, the actual voltage is greater than the equilibrium voltage due to the irreversible loss. The irreversible losses in electrolysis are activation polarization, ohmic polarization, and concentration polarization. The activation polarization is given by Butler-Volmer equation as: (Yahya et al., 2018)

$$\eta_{act,i} = \frac{RT}{2F}\ln\left[\frac{J}{2J_{0,i}} + \sqrt{\left(\frac{J}{2J_{0,i}}\right)^2 + 1}\right] \tag{3}$$

$$J_{0,i} = \frac{RT}{2F}\left(\frac{P_{H_2O}}{P_{ref}}\right)^{1.0433}\left(\frac{P_{H_2}}{P_{ref}}\right)^{0.9653} k_i \exp\left(-\frac{E_i}{RT}\right) \tag{4}$$

Here, i can be cathode or anode, and J and $J_{0,i}$ represent the current density and exchange current density, respectively. The ohmic polarization is given by: (Ferguson et al., 1996)

$$\eta_{ohm} = J\left(\frac{d_{an}}{\sigma_{an}} + \frac{d_{ele}}{\sigma_{ele}} + \frac{d_{ca}}{\sigma_{ca}}\right) \tag{6}$$

where d and σ are the thickness and conductivity, respectively. The concentration polarizations in the SOFEC anode and cathode can be calculated by

$$\eta_{conc,SOFEC} = \frac{RT}{2F}\ln\left(\frac{P^{TPB}_{H_2,ca}P_{H_2O,ca}}{P_{H_2,ca}P^{TPB}_{H_2O,ca}}\right) + \frac{RT}{2F}\ln\left(\frac{P^{TPB}_{H_2O,an}P_{H_2,an}}{P_{H_2O,an}P^{TPB}_{H_2,an}}\right) \tag{7}$$

The power input or output to the SOFEC stack is given in Eq.(8), and the current can be calculated according to Faraday's law, which means that for every mole of water vapor electrolyzed, two moles of electrons are generated

$$P_{\mathrm{SOFEC}} = IV_{\mathrm{SOFEC}} \tag{8}$$

$$I = 2F \times \dot{n}_{steam,consumed} \tag{9}$$

3. Integrated System process description

The process system created by Aspen Plus consists of two main subsystems, the biomass gasifier and the SOFEC. The SOFEC system can be divided into a cathode and an anode side. At the cathode side of the SOFEC, the electrolysis reaction of steam is carried out. Fresh feed water is pressurized and heated to generate steam through an evaporator. The electrolysis temperature of the SOFEC is set to 800°C. In order to avoid the breakage of the SOFEC structure and the formation of oxidizing conditions, the steam is heated to a higher temperature through the heat exchanger and mixed with the partial recirculation of cathode exit stream before entering the cathode. The cathode outlet stream contains hydrogen and unreacted steam, so the water is removed from the stream by condensation and the pure hydrogen is separated as the hydrogen product of the system.

The anode side is mainly used for fuel-assisted electrolysis reaction, where the fuel source is the syngas generated by the gasifier. Firstly, the solid biomass is converted into syngas through a gasifier. In order to increase the amount of hydrogen in the syngas and to avoid carbon deposition on the surface of the SOFEC anode, the syngas will be passed through the reformer for the steam reforming reaction before entering the SOFEC anode side. The reformed fuel is heated through heat exchange and then entering the anode for the O^{2-} reduction reaction, while the unreacted fuel at the anode exit will enter the afterburner for combustion. The burned streams contain a large amount of heat enaegy that can be used to preheat the inlet streams of the SOFEC system, and at the end the excess heat is used for heat recovery applications.

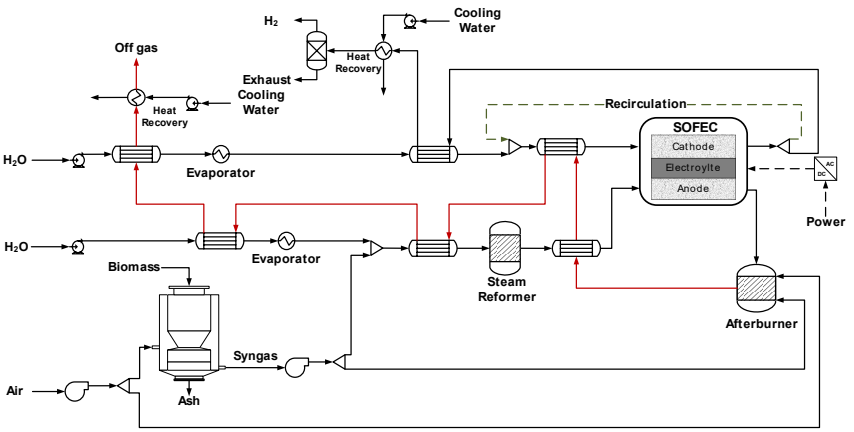


Figure 2 Gasifier-SOFEC integration system.

4. Results and Discussions

4.1. Effect of equivalence ratio on the system

Figure 3 shows the effect of changing the gasification equivalence ratio on the system efficiency. When the equivalence ratio is at 0.2, the total system efficiency is the highest at each operating steam utilization, and when the biomass gasifier equivalence ratio exceeds 0.2, the system efficiency decreases as the equivalence ratio increases because the flow rate of the syngas produced at an equivalence ratio of 0.2 is the highest, thus allowing the system to electrolyze more steam on the cathode side.

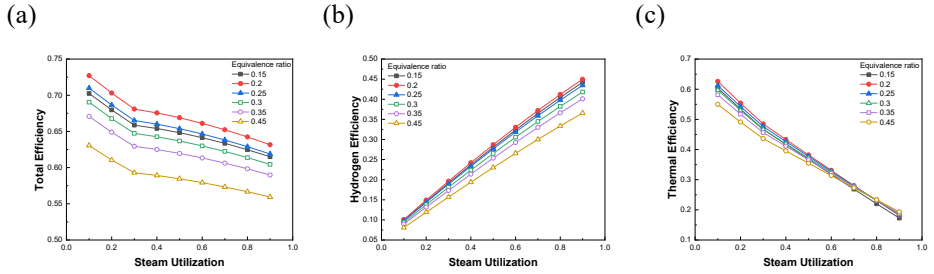


Figure 3 Effect of gasifier equivalent ratio on system efficiency (a)Total efficiency (b)Hydrogen production efficiency (c)Thermal efficiency.

The total efficiency decreases as the SOFEC steam utilization increases. The highest overall efficiency is 72.70% at a SOFEC steam utilization of 0.1 and a gasifier equivalence ratio of 0.2. As the steam utilization gradually increases, the hydrogen production efficiency gradually increases, while the thermal efficiency gradually decreases. At an equivalence ratio of 0.2 and an SOFEC steam utilization rate of 0.9, the maximum hydrogen production efficiency of the system is about 44.99%. Because the amount of syngas consumed in the electrochemical reaction is less when the SOFEC steam utilization is small, the amount of unused fuel at the anode outlet is higher and more fuel is burned in the afterburner, resulting in a large amount of heat that can be recovered at the outlet.

4.2. Effect of gasification temperature on the system

Figure 4 show shows the effect of gasification temperature on system efficiency, the maximum total efficiency occurs at the gasifier temperature of 900°C and SOFEC electrolysis rate of 0.1, which is about 68.67%, while the total efficiency is much lower than other operating temperatures when the gasifier is operated at 600°C. In the hydrogen production efficiency, it can be seen that when the temperature of the biomass gasifier is operated at 600°C, the efficiency decreases by about 10% compared to other temperatures. At other temperatures, the hydrogen production efficiencies of the three almost overlap. The main difference is in thermal efficiency because the carbon conversion rate of the gasifier is low at the gasification temperature of 600°C, resulting in less syngas production in the gasifier, which reduces the amount of fuel that can be assisted, making the amount of steam that can be electrolyzed in the cathode lower.

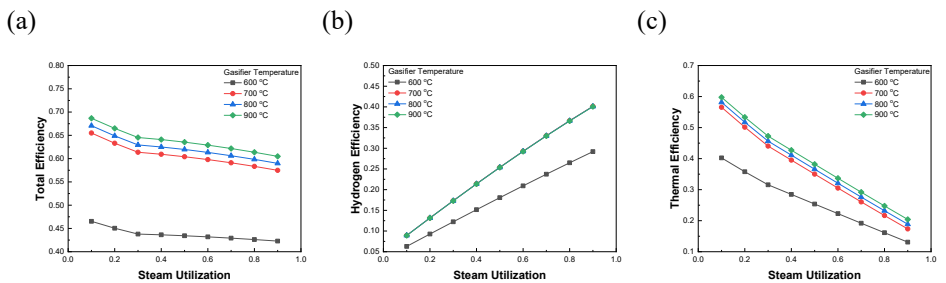


Figure 4 Effect of gasification temperature on system efficiency: (a) Total efficiency (b) Hydrogen production efficiency (c) Thermal efficiency.

4.3. Effect of different gasifier operating conditions on system efficiency

According to the previous analysis, changes in the operating parameters of the gasifier will affect the overall efficiency of the system. The higher the production of syngas in the gasifier, the higher the overall efficiency of the integrated system. By varying the moisture content of the biomass after drying, the highest efficiency of the gasifier occurs at moisture contents of 5%, 9% and 15% with equivalent ratios of 0.25, 0.2 and 0.15, respectively. From Figure 5, the maximum efficiency of the system at 800°C occurred at 15% biomass moisture content and 0.15 equivalents, with a maximum total efficiency of 75.42% for SOFEC operated at 0.1 steam utilization. The maximum hydrogen production efficiency was 46.97% at a steam utilization of 0.9. From the results, it can be confirmed that when the biomass gasifier operates at maximum efficiency, the integrated system also has the highest efficiency.

(a) (b)

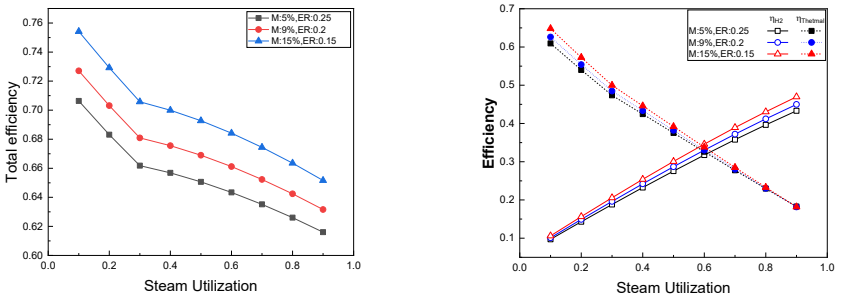


Figure 5 System efficiency under different gasification conditions: (a) Total efficiency (b) Hydrogen production and Thermal efficiency.

5. Conclusions

In the study, the effects of operating parameters of the gasifier on the overall efficiency and specific energy consumption of the gasifier-SOFEC integrated system were analyzed, and it can be learned that the key parameters of the gasifier are the equivalence ratio of the gasification agent and the gasification temperature, while the impact of the moisture content of the biomass drying on the system efficiency and specific energy consumption is not obvious. When the gasification temperature is higher than 700°C, the effect on system efficiency is reduced. When the syngas production from the gasifier is higher, the efficiency of the integrated system increases, mainly because the higher syngas production enables more water vapor to be electrolyzed, resulting in higher hydrogen production. Therefore, from the final results, the highest efficiency of the system was achieved when the gasification temperature was operated at 800C, the water content of the biomass was 15%, and the equivalence ratio was 0.15. The total efficiency was 75.42% under the SOFEC steam utilization of 0.1, and the maximum hydrogen production efficiency was 46.97% under the steam utilziation of 0.9.

6. Reference

Tavares, R., E. Monteiro, F. Tabet, and A. Rouboa, Numerical investigation of optimum operating conditions for syngas and hydrogen production from biomass gasification using Aspen Plus. Renewable Energy, 2020. **146**: p. 1309-1314.

Salzano, F.J., G. Skaperdas, and A. Mezzina, Water vapor electrolysis at high temperature: Systems considerations and benefits. International Journal of Hydrogen Energy, 1985. **10**(12): p. 801-809.

Yahya, A., D. Ferrero, H. Dhahri, P. Leone, K. Slimi, and M. Santarelli, Electrochemical performance of solid oxide fuel cell: Experimental study and calibrated model. Energy, 2018. **142**: p. 932-943.

Ferguson, J.R., J.M. Fiard, and R. Herbin, Three-dimensional numerical simulation for various geometries of solid oxide fuel cells. Journal of Power Sources, 1996. **58**(2): p. 109-122.

Proceedings of the 14th International Symposium on Process Systems Engineering -PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50353-5

Plasma-Based Pyrolysis of Municipal Solid Plastic Waste for a Robust WTE Process

Hossam A.Gabbar*, Emmanuel Galiwango, Mustafa A. Aldeeb, Sharif Abu Darda, Kiran Mohammed

Faculty of Energy Systems and Nuclear Science, University of Ontario Institute of Technology, Oshawa, ON L1G0C5, Canada.
* Hossam.Gaber@ontariotechu.ca

ABSTRACT

The process design and material charactrization for a futuristic plasma-based pyrolysis/gasification for waste to energy is presented. The direct current as the source of thermal energy and the plasma-based reactor were designed. The plastic solid waste was analyze for pyrolytic, functional groups and morphological properties. Pyrolysis results using thermogravimetric analysis showed increase in thermal energy in the system with increasing heat rate and one major degradation peak was observed between 400-550 ºC. There was no observed difference in functional groups between raw plastic solid waste and its ash. The structural morphologies of the plastic ash sample showed uniform rough surfaces. The physicochemical results of high volatile matter, low moisture and oxygen content revealved the ease of pyrolysis and postullate for quality products out of the futuristic plasma-based pyrolysis/gasification process.

Keywords: Municipal solid waste; Control; Operation; Characterization

1. Introduction

Solid wastes constitute a significant fraction of waste in the environment. The world generates 2.01 billion tonnes of municipal solid waste annually and this amount is expected to increase to 3.40 billion tonnes by 2050, and is double the population growth rate [1]. In addition, over at least 33 percent of that is extremely conservatively not managed in an environmentally safe manner [2]. Proper management and recycling of large amounts of solid waste are necessary to reduce its environment burdens and to minimize risks to human health [3, 4]. Sustainable and safe solid waste management is an under tapped field with great potential for energy production [5, 6]. Proper utilization of solid waste for conversion into energy and valuable products currently face various challenges such as heterogeneous nature of the waste, large moisture content, low calorific value, making it industrially undesirable [7-9]. Although appreciable amount of research in the field has been carried out on the conversion of solid waste to energy, there is still lack of comprehensive and efficient conversion methods in literature. Conventional conversion technologies of waste-to-energy (WTE) include biological (such as fermentation, anaerobic digestion and etc), thermochemical processes (such as incineration, pyrolysis, hydrothermal oxidation, and gasification) [10, 11]. These technologies enjoy several advantages such as biogas from biological process produce huge amounts of CH_4 and CO_2 with high energy value (1 m^3 of biogas was reported to be equivalent to 21 MJ of energy, and it could generate 2.04 kW h of electricity considering the 35% of generation efficiency [12]. Furthermore, Incinerators can reduce the volume of solid wastes up to 80–85%, and thus, significantly reduce the necessary volume for disposal. Also, biofuels from processes such as pyrolysis/gasification are positive indicators of good approach for WTE [13]. However, there setbacks in these technological advances such as pollution challenges from components like nitrogen, hydrogen sulfide

and oxygen embedded in the waste [14, 15]. In addition, the heterogeneous nature of most solid wastes makes energy recovery, product yields and quality difficult [16]. With the improvement in air emission control systems and commitment from policy makes to enforce more strict environmental regulatory rules that significantly decreasing potential human health effects, more research on efficient processes for solid waste conversion to WTE and oil is gaining more attention.

The overall goal was to contribute to the solution of this challenge, we proposed the integration of DC/Microwave driven Inductively Coupled Atmospheric pressure thermal Plasma (RF/M-ICAP) to provide energy from municipal solid waste for transportation systems such as: trains, ships, garbage trucks etc. The proposed system includes new MICAP torch and reactor chamber design to produce highly efficient plasma jet suitable to process any waste type for different applications. MICAP, not only applied to reduce the volume of solid waste, but also to produce fuel from plastic and electricity through a process called pyrolysis and gasification. The suggested process has a potential to provide solution for waste-to-energy within transportation infrastructures, with reduced waste volume and increase clean energy generation while reducing emissions of trips resulting from extra transportation trips of waste transfer. Engineering designs are proposed for the target mobile waste-to-energy to be integrated with transportation infrastructures and develop discharge mechanisms of produced clean fuel and electricity in distributed stations based on demand profiles and generation capacities. This paper introduces novel design of modular chamber torches and waste characteristics with innovative features to maximize energy efficiency and minimize losses while improving the waste conversion to energy with minimum losses. Intelligent control systems will be developed to control plasma torch, chamber, and the associated energy systems.

2. Materials and Methods

2.1. Materials and chemicals

The shredded plastic MSW (average particle size <1mm) were obtained from industrial partner (Pro-flange, ON, Canada). The samples were air dried, separated according to their uniform compositions prior to the physicochemical characterization. The ultimate analysis was analyzed in external laboratory (Biotron Experimental Climate Change Research Centre at Western University in London, Ontario, Canada). The proximate analysis were determined on site in triplicates according to ASTM D3173 (inherent moisture content), ASTM D3174 (ash), ASTM D3175 (volatile matter), ASTM D3172-07a (Fixed carbon) and bulky density according to the reference method [17].

The torches and reactor designs were designed using solid works and COMSOL simualtor.

2.2. Waste characterization

2.2.1 Thermogravimetric analysis

The samples were subjected to thermogravimetric analysis (TGA) to investigate their pyrolytic properties of the waste material. In all experimental runs, 8±2 mg of the sample was placed on a platinum thermobalance crucible and loaded to the TGA analyser (Q50 series, TA instrument). The samples were equilibrated at 30 ^{o}C for 5 minutes before heating to 800 ^{o}C at heating rates of 10, 20, 30 and 40 °C/min under continuous inert N_2 flow at 20 mL/min. The results of thermal decomposition were continuously recorded as a change in weight as a function of temperature and time.

2.2.2 Fourier Transform Infra-red spectroscopy (FTIR) analysis

The ATR-FTIR spectroscopic analysis was performed to investigate the possible structural alteration between the samples before and after pyrolysis. Plastic samples before and after pyrolysis were pressed uniformly against a diamond surface by a fixed sample holder anvil, spectra were observed using a Bruker optics vertex system with inbuilt diamond-germanium ATR single reflection crystal. Spectra were obtained over a range of 400 and 4000 cm^{-1} with 34 average numbers of scans and spectral resolution of 4 cm^{-1}.

2.2.3 Scanning electron microscope (SEM) analysis

A FlexSEM1000 Scanning Electron Microscope (SEM) operated at 5 kV; spot size of 40 was used to image the char sample. To improve conductivity and quality of image, samples were coated with Au/C using a vacuum sputter coater.

2.3 Thermal source design

The torch design featured different major parts that aid in simulation to provide optimum conditions for the pyrolysis and gasification of the waste.

3. Results and Discussion

3.1 Waste characterization

The material characteristics are presented in Table 1. The low moisture content is important in lowering the energy requirements often needed to offset the inherent high moisture content in many MSW. This makes the waste under investigation a economically a good candidate for production of valuable products and electricity. In addition, the high volatile matter and low fixed carbon is an indication of the ease for combustion of this type of plastic waste, thus economical viability of the process. The ultimate analysis results show high hydrogen to carbon ratio (0.17) with low oxygen content, hence an anticipation of good quality products with less oxygenated fractionates.

Table 1: Material physicochemical charateristics

Proximate analysis (wt.%)				
Moisture content	Ash content	Volatile matter	Fixed carbon	Bulky density (Kg/m3)
0.13 ± 0.02	0.79 ± 0.01	86.8 ± 0.4	12.28 ± 0.1	84.43 ± 1.5
Ultimate analysis (wt.%)				
C	H	N	S	O
85.44	14.34	0.02	<MRL	0.22

MRL: method reporting limit

The TGA results in Figure 1 revealed the possible isothermal conversion (α) of the waste to be > 90 wt.% at all tested heating rates. Morever, the increase in the heating rate shifted the TA curves towards higher temeperature side, an indication of increase in thermal energy to the system with increasing heating rate. The differential thermograms showed one major degradation peak, a postullation of a single component under thermal pyrolysis. Pyrolysis/gasification of this category of waste is poised to be successful across different heating rates and product distribution will be easy to understand given the DTG degradation that suggests similarity in degradation chemistry of the material under pyrolysis.

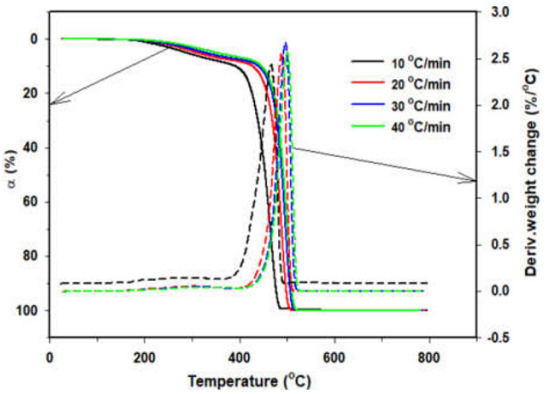


Figure 1: TGA and DTG curves for mixture of plastic waste at heating rates of 10, 20, 30 and 40 °C/min.

The study of the functional groups in Figure 2 shows characteristic similarity between the raw waste and its ash with minimal changes in the functional groups but rather more spectral intensity for the ash. These results prove the known phenomena about the material being a thermoplastic. The major vibrations between 2900-2800 cm^{-1} were assigned to the C-H and CH_2 bond vibrations characteristic to hydrocarbons present in most plastic waste. The absorption bands below 2000 cm^{-1} were associated with the stretching of the C=C, C-H, C-O, symetric and asymetric CH_2 bonds inherent in the aromatic ring of plastic polymer and the aliphatic structural makeup of the plastic [18, 19]. It is worthy to note that the absence of absorption at wave number around 3500 cm-1 a region characterstic of O-H bond vibration proves low existance of water molecule bonds as already seen from the low moisture content measurements.

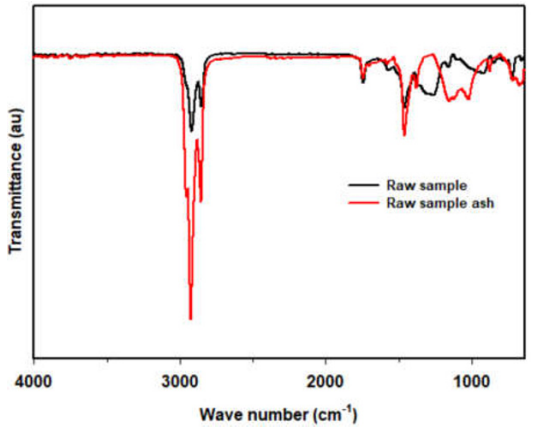


Figure 2: FTIR measurements of raw plastic sample and its ash after pyroysis at 800 °C.

The morphological studies following pyrolysis of the plastic MSW samples in Figure 3 revealed agglomerates with majorly rough surfaces and a few smooth surfaces observed X5 and X10 magnification. The micrographs depicts a good thermal interaction with the plastic waste's polymer matrix.

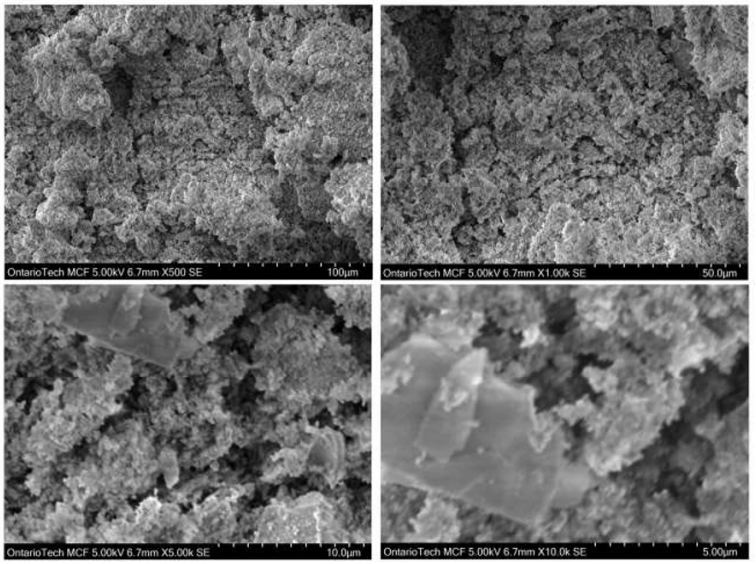


Figure 3: Plastic ash SEM images at different magnifications.

3.1 Thermal source design

Figure 4 represents the design for the DC torch and the plasma-based reactor for the futuristic process for production of renewable environmentally safe energy from waste.

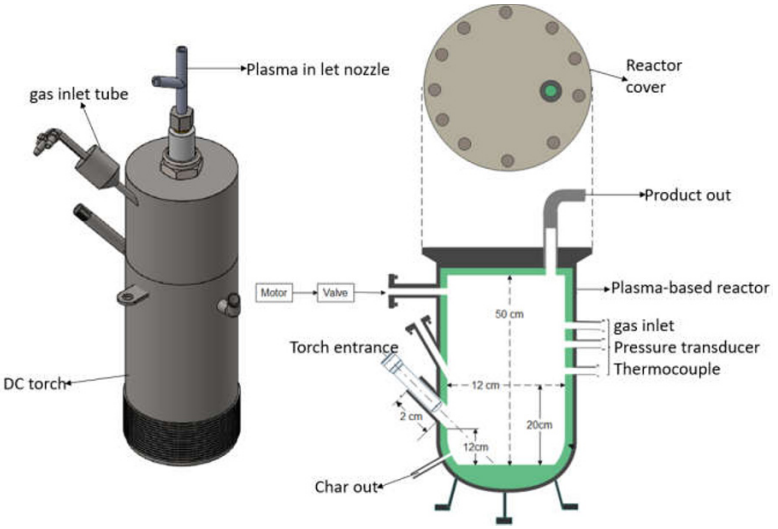


Figure 4: DC torch and plasma-based reactor drawings.

4. CONCLUSION

The conceptual design and charaterization of the plastic municipal solid waste provides a deeper understanding and new approach to waste management in an environmentally and economically friendly manner. The low moisture content and ash content combined with high voltility of the waste and H-C ratio signifies a potential ease of the futuristic plasma-

based pyrolysis/gasification process both in terms of energy requirements and reactor configuration.

Acknowledgements: Thanks to lab members of Advanced Plasma Engineering Lab (APEL). Authors would like to thank Proflange Ltd. for their funding and continuous support. The research is also supported by NSERC, OCI, and Mitacs.

References

[1] T.W. Bank, Trends in Solid Waste Management, in, 2016.
[2] M.A. Nwachukwu, M. Ronald, H. Feng, Global capacity, potentials and trends of solid waste research and management, Waste Management & Research, 35 (2017) 923-934.
[3] H. Yang, M. Ma, J.R. Thompson, R.J. Flower, Waste management, informal recycling, environmental pollution and public health, J Epidemiol Community Health, 72 (2018) 237-243.
[4] P. Alam, K. Ahmade, Impact of solid waste on health and the environment, International Journal of Sustainable Development and Green Economics (IJSDGE), 2 (2013) 165-168.
[5] S.A.T. Muawad, A.A.M. Omara, Waste to energy as an alternative energy source and waste management solution, in, IEEE, pp. 1-6.
[6] P. Prajapati, S. Varjani, R.R. Singhania, A.K. Patel, M.K. Awasthi, R. Sindhu, Z. Zhang, P. Binod, S.K. Awasthi, P. Chaturvedi, Critical review on technological advancements for effective waste management of municipal solid waste-Updates and way forward, Environmental Technology & Innovation, (2021) 101749.
[7] S. Das, S.H. Lee, P. Kumar, K.-H. Kim, S.S. Lee, S.S. Bhattacharya, Solid waste management: Scope and the challenge of sustainability, Journal of cleaner production, 228 (2019) 658-678.
[8] J. Ali, T. Rasheed, M. Afreen, M.T. Anwar, Z. Nawaz, H. Anwar, K. Rizwan, Modalities for conversion of waste to energy—challenges and perspectives, Science of The Total Environment, 727 (2020) 138610.
[9] K. Moustakas, M. Loizidou, M. Rehan, A.S. Nizami, A review of recent developments in renewable and sustainable energy systems: Key challenges and future perspective, in, Elsevier, 2020.
[10] N. Tulebayeva, D. Yergobek, G. Pestunova, A. Mottaeva, Z. Sapakova, Green economy: waste management and recycling methods, in, EDP Sciences, pp. 01012.
[11] D. Ghosal, R. Kaur, D.A. Jadhav, Utilization and Management of Waste Derived Material for Sustainable Energy Production: A mini review, Academia Letters, (2021).
[12] J.D. Murphy, E. McKeogh, G. Kiely, Technical/economic/environmental analysis of biogas utilisation, Applied Energy, 77 (2004) 407-427.
[13] T. Rasheed, M.T. Anwar, N. Ahmad, F. Sher, S.U.-D. Khan, A. Ahmad, R. Khan, I. Wazeer, Valorisation and emerging perspective of biomass based waste-to-energy technologies and their socio-environmental impact: A review, Journal of Environmental Management, 287 (2021) 112257.
[14] M.S. Qureshi, A. Oasmaa, H. Pihkola, I. Deviatkin, A. Tenhunen, J. Mannila, H. Minkkinen, M. Pohjakallio, J. Laine-Ylijoki, Pyrolysis of plastic waste: Opportunities and challenges, Journal of Analytical and Applied Pyrolysis, 152 (2020) 104804.
[15] J. Lui, W.-H. Chen, D.C.W. Tsang, S. You, A critical review on the principles, applications, and challenges of waste-to-hydrogen technologies, Renewable and Sustainable Energy Reviews, 134 (2020) 110365.
[16] M. Singh, S.A. Salaudeen, B.H. Gilroyed, S.M. Al-Salem, A. Dutta, A review on co-pyrolysis of biomass with plastics and tires: recent progress, catalyst development, and scaling up potential, Biomass Conversion and Biorefinery, (2021) 1-25.
[17] I. Obernberger, G. Thek, Physical characterisation and chemical composition of densified biomass fuels with regard to their combustion behaviour, Biomass and bioenergy, 27 (2004) 653-669.
[18] R.C. Asensio, M.S.A. Moya, J.M. de la Roja, M. Gómez, Analytical characterization of polymers used in conservation and restoration by ATR-FTIR spectroscopy, Analytical and bioanalytical chemistry, 395 (2009) 2081-2096.
[19] I. Noda, A.E. Dowrey, J.L. Haynes, C. Marcott, Group frequency assignments for major infrared bands observed in common synthetic polymers, in: Physical properties of polymers handbook, Springer, 2007, pp. 395-406.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50354-7

Hybrid Modelling Strategies for Continuous Pharmaceutical Manufacturing within Digital Twin Framework

Pooja Bhalode[a#], Yingjie Chen[b#], Marianthi Ierapetritou[b,*]

[a]*Department of Chemical and Biochemical Engineering, Rutgers University, 98 Brett Road, Piscataway, NJ 08854, USA*
[b]*Department of Chemical and Biomolecular Engineering, 150 Academy Street, Newark, DE 19716, USA*
[*]*Corresponding author – mgi@udel.edu*
[#]*These authors contributed equally to this work*

Abstract

The application of Industry 4.0 related technologies has enabled the development of digital twins to model and mirror the physical systems in a virtual construct. Despite the successful adoption of digital twins in other industries, a complete digital twin framework is yet to be developed for continuous pharmaceutical manufacturing processes. Challenges related to the specific industry include the integration of multi-scale information ranging from powder properties to process flowsheets, and the need of model adaptability to capture changes in operation. In this work, hybrid multizonal compartment models and hybrid adaptive models are developed to address such challenges. The computationally efficient and self-adaptive hybrid modelling strategies can aid the developing of digital twin for continuous pharmaceutical manufacturing.

Keywords: Hybrid modelling; Digital twin; Pharma 4.0; Continuous pharmaceutical manufacturing; multi-zonal compartment modelling.

1. Introduction

Industry 4.0 revolution has been catalysed by the advances in digitization, cyber-infrastructure, artificial intelligence, and Internet of Things. In pharmaceutical industry, this digitalization move is popularly named as Pharma 4.0, and it is incentivized by increasing market competition, along with the encouragement from regulatory agencies to develop agile, flexible, and robust manufacturing lines (O'Connor et al., 2016, Chen et al., 2020). Pharma 4.0 technologies enable the development of a digital twin, which can be defined as an integrated digitized framework consisting of virtual and physical components, with a seamless connection between the two (Chen et al., 2020). In continuous pharmaceutical manufacturing, although some efforts have been made to create data integration framework and various types of models, a fully integrated digital twin that allows for real-time process monitoring, control, and optimization is still in its infancy. From a modelling perspective, multi-scale models that can characterize powder properties and describe process flows need to be integrated, which is a challenging task. Furthermore, models developed are required to adapt to process changes, instead of being static. These challenges are bottlenecks in development of a digital twin for continuous pharmaceutical manufacturing.

We propose the use of hybrid modelling strategies to address some of these issues. In general, hybrid models can be considered as a combination of data-driven and white-box (mechanistic) models (Zendehboudi et al., 2018). The modelling components can be arranged in a serial, parallel, or combined manner (Chen and Ierapetritou, 2020). This modelling strategy provides a unique way to combine known mechanistic knowledge with data. In this work, a novel two-stage solution approach based on hybrid modelling strategies is proposed. The first stage addresses the challenge of multi-scale models by utilizing hybrid multi-zonal compartmentalization method, which is discussed in Section 2. The second stage focuses on the hybrid adaptive model framework, as discussed in Section 3. The two stages, when integrated, would be a central part of a computationally efficient and self-adaptive digital twin for continuous pharmaceutical manufacturing.

2. Hybrid Multi-zonal Compartmentalization

Multi-zonal compartmentalization methodology is a computationally efficient approach used to model complex systems (Jourdan et al., 2019). Compartmentalization methodology aims at combining systemic (unit operation) and local (particle) level information using multi-scale models and has been applied for complex process operations, such as biologics (Delafosse et al., 2014) Here, the systemic information obtained from mechanistic models is combined with local process information obtained from high-fidelity simulations such as computational fluid dynamics (CFD) or discrete element modelling (DEM). The detailed local information is compressed by dividing the simulation domain into compartments based on the process variables of interest, such that each compartment has process variables within user-defined limits based on the degree of scrutiny required for modelling the system.

In this paper, compartmentalization methodology is used to develop a predictive model for continuous powder blender (Bhalode and Ierapetritou, 2021). The process models developed in literature for blenders lack understanding and prediction of powder mixing within the systems and systemic models do not capture this level of detail (Vanarase and Muzzio, 2011). Important for assuring drug product quality, it is crucial to develop predictive models that incorporate local information concerning powder mixing (Lee et al., 2015). The details of the developed compartment model for continuous powder blender are described in Figure 1 where, the blender is broken down into periodic sections and a section is simulated using DEM. To demonstrate the proposed strategy, two types of spherical particles with 1 mm particle size are added to the section, to ensure reasonable computational times of DEM simulation, and can be easily extended for more particle types. The simulation is performed with an Intel Xeon E5-2650 v4 2.2GHz processor and 128 GB RAM using EDEM 2021 (Altair Solutions, Michigan, USA), leading to a total computation time of 8 hrs with 4 CPU cores. The simulated section is then divided into 10 equal slices along the X axis for ease of computation and each slice is further divided into 20x20 grids along Y and Z axes for post-processing. Average particle velocities along all directions (v) are extracted for all grids, and the grids are compartmentalized into zones where v is positive, negative, zero, and zone with no particles. Using these zones, radial compartment maps are developed for all slices, and these maps are overlaid on top of each other (Bhalode and Ierapetritou 2021). Using the overlaid radial map, interconnection flowrates are determined based on the interfacial area and the velocities of the respective compartments, converting the compartment map into an inter-connected network of compartments. A similar approach is adopted along the axial direction to develop network of compartments based on interconnection flow rates between the

compartments. The axial network is eventually combined with the radial to obtain the overall the multizonal compartment model for the periodic section. This model thus incorporates systemic as well as local level information, allowing detailed assessment of powder flow.

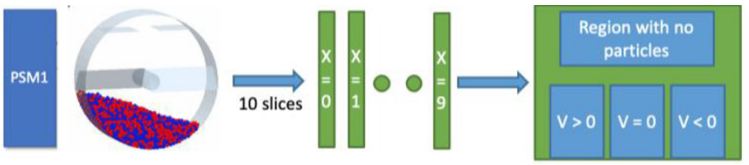


Figure 1: Zones developed along axial and radial direction for the periodic section

Following the model development, validation is performed along both radial and axial directions. For validation along the radial direction, the degree of powder mixing is compared between the compartment model and the DEM simulation. For comparison, a mixing index – relative standard deviation (RSD) is used to obtain the time at which the system has reached 95% of the total mixing in the system (T_{95}). T_{95} is evaluated for different RPMs and shown in Table 1 along with the absolute relative error. Validation along the axial direction is performed by comparing the axial flow using residence time distribution (RTD) profiles, as shown in Figure 2, where similar RTD profiles are obtained using DEM and compartment models. Thus, compartment model provides accurate prediction of powder flow compared to the computationally intensive DEM simulation, along with significant time savings (in order of 1-2 mins) Lastly, the compartment model can be extended to model the entire length of continuous blender by connecting periodic sections, as shown in Figure 3.

Table 1. T_{95} mixing times for radial validation of DEM compared to Compartment model

Blade speed	T_{95} (DEM)	T_{95} (Compartment model)	Absolute relative error
50 RPM	14.812	14.548	0.018
75 RPM	12.366	12.129	0.019
100 RPM	8.911	9.268	0.040

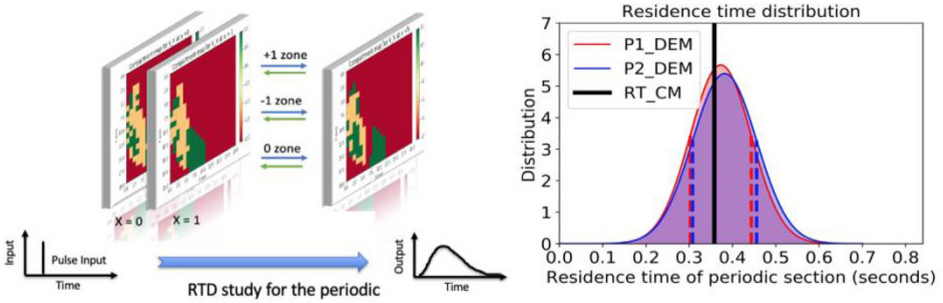


Figure 2. Axial validation performed using residence time distribution profiles.

3. Hybrid Adaptive Modelling

As mentioned in Section 1, the development of hybrid models involves the combination of data-driven and mechanistic models. Given the structure of hybrid models, some parts are data-driven and typically trained on defined sets of historical data. Therefore, the

resulting hybrid models are often time-invariant and can only reflect the system for the range of operational conditions, environmental variables, equipment status, and material properties covered in the training set (Gama et al., 2014), leading to a challenge related to model updates. Although the hybrid models may accurately describe the current state of a system, the process can shift slowly due to changes in the features mentioned above. These informed or uninformed changes can impact the performance of a process and, eventually, the critical quality attributes of a product leading to decline of model prediction accuracy (Gama et al., 2014). Appropriate model update strategies thus need to be established to resolve this issue.

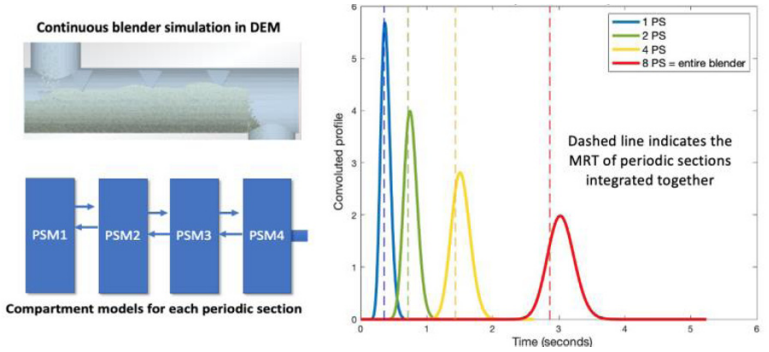


Figure 3. Compartment model extended for the entire continuous powder blender

The proposed framework for hybrid adaptive modelling is shown in Figure 4. Using historical data with adequate model selection criteria, an initial data-driven sub-model can be built and integrated into an appropriate hybrid model structure. Prediction results from the hybrid model are compared with plant outputs in a continuous manner. Based on defined criteria, the adaptive algorithm determines if an update on the hybrid model is required, and carries out the necessary steps to retrain the model. The overall computational time depends on the selection of model training and adaptive algorithms.

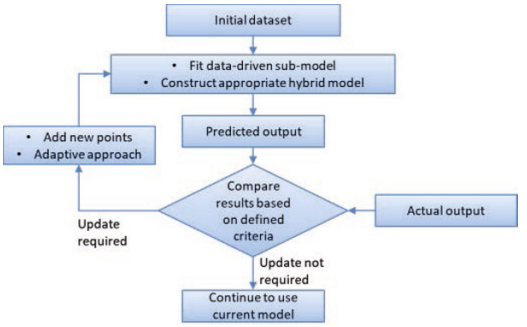


Figure 4. Hybrid adaptive modelling framework.

Two adaptive modelling algorithms are applied. The moving window method with fixed window size, as depicted in Figure 5(a), refers to a blind adaptation technique where the model is adapted using a fixed number of most recent data points, regardless if a change is observed (Gama et al., 2014). As new samples are streamed in, the window slides to include the newest samples and forgets the older ones. This strategy also implies that

model update is always required. The adaptive windowing method, as shown in Figure 5(b), is an informed adaptation technique (Gama et al., 2014), which includes a change detection component and an adaptive algorithm. The number of data points included in a window grows until a significant change is detected. The detection algorithm splits the window into two sub-windows in all possible permutations and compares the means. When the difference of two means is larger than a defined tolerance, a process change is detected, and the points before such change (i.e., the ones in the first sub-window) are discarded. Both approaches can be integrated into different data-driven modelling methodologies to enable model updates, but the adaptive windowing can be more computationally expensive because of the change detection algorithm.

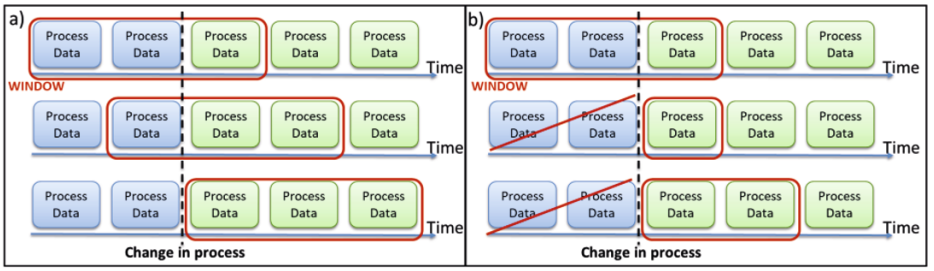


Figure 5. Schematic of (a) moving window with fixed window size, and (b) the adaptive windowing method.

Both algorithms are tested on a continuous pharmaceutical manufacturing line via direct compaction route. Such process is explained in (Wang et al., 2017). The baseline models are neural networks with the combined structure, detailed in (Chen and Ierapetritou 2020).

A case study of sudden and patterned process change is developed, for which the hybrid adaptive modelling framework experiences an uninformed change of excipient bulk density from the nominal 400 kg/m^3 to 520 kg/m^3 (and back for the patterned change), with other variables remaining unaffected. The impacts of the change onto tablet weight are monitored, and the results are shown in Figure 6.

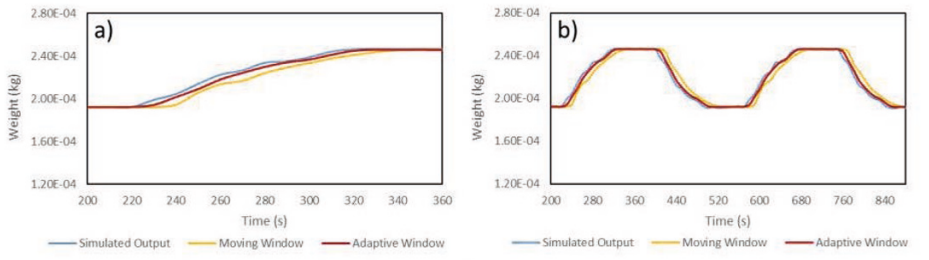


Figure 6. Comparison of results for between simulated output, hybrid adaptive model predictions based on moving window and adaptive windowing for (a) sudden, and (b) patterned change.

The hybrid adaptive models can follow the changes in process, and the adaptive windowing technique outperforms moving window. This observation is expected as the change detection algorithm in adaptive windowing makes it very sensitive to this type of sudden changes, resulting in a faster response.

4. Conclusions

The proposed article focuses on utilizing hybrid modelling strategies to address the challenges associated with developing a digital twin for pharmaceutical manufacturing. Hybrid multi-zonal compartment models are developed to integrate multi-scale information concerning powder flow properties and operation mechanics within the developed process model, demonstrated for continuous powder blender. Hybrid adaptive modelling framework is developed to facilitate model updates with process changes or newly acquired data, showcased with a case of mid-process change. The resulting hybrid and adaptive model can be used as an online predictive model in a fully integrated digital twin. The two strategies can be further integrated together to move towards an integrated digital twin in continuous pharmaceutical manufacturing.

Acknowledgement

The authors would like to acknowledge funding from U.S. Food and Drug Administration through grants DHHS-FDA-U01FD006487 and FDABAA-20-00123.

References

P. Bhalode, M. Ierapetritou, 2021, Hybrid multi-zonal compartment modeling for continuous powder blending processes, Int J Pharm, 602, 120643

Y. Chen, M. Ierapetritou, 2020, A framework of hybrid model development with identification of plant-model mismatch, AIChE Jounal, 66, 10, e16996.

Y. Chen, O. Yang, C. Sampat, P. Bhalode, R. Ramachandran, M. Ierapetritou, 2020, Digital Twins in Pharmaceutical and Biopharmaceutical Manufacturing: A Literature Review, Process, 8, 9.

A. Delafosse, M. Collignon, S. Calvo, F. Delvigne, M. Crine, P. Thonart, D. Toye, 2014, CFD-based compartment model for description of mixing in bioreactors, Chemical Engineering Science, 106, 76-85.

J. Gama, I. Zliobaite, A. Bifet, M. Pechenizkiy, A. Bouchachia, 2014, A survey on concept drift adaptation, ACM Computing Surveys, 46, 4, 1-37.

N. Jourdan, T. Neveux, O. Potier, M. Kanniche, J. Wicks, I. Nopens, U. Rehman, Y. Le Moullec, 2019, Compartmental Modelling in chemical engineering: A critical review, Chemical Engineering Science, 210, 115196.

S. Lee, T. O'Connor, X. Yang, C. Cruz, S. Chatterjee, R. Madurawe, C. Moore, L. Yu, J. Woodcock, 2015, Modernizing Pharmaceutical Manufacturing: from Batch to Continuous Production, Journal of Pharmaceutical Innovation, 10, 3, 191-199.

T. O'Connor, L. Yu, S. Lee, 2016, Emerging technology: A key enabler for modernizing pharmaceutical manufacturing and advancing product quality, Int J Pharm, 509, 1-2, 492-498.

A. Vanarase, F. Muzzio, 2011, Effect of operating conditions and design parameters in a continuous powder mixer, Powder Technology, 208, 1, 26-36.

Z. Wang, E. Sebastian, M. Ierapetritou, 2017, Process analysis and optimization of continuous pharmaceutical manufacturing using flowsheet models, Copmuters & Chemical Engineering, 107, 77-91.

S. Zendehboudi, N. Rezaei, A. Lohi, 2018, Applications of hybrid models in chemical, petroleum, and energy systems: A systematic review, Applied Energy, 228, 2539-2566.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50355-9

Determination of probabilistic design spaces in the hybrid manufacture of an active pharmaceutical ingredient using PharmaPy

Daniel Laky[a], Daniel Casas-Orozco[a], Francesco Rossi[a], Jaron S. Mackey[a], Gintaras V. Reklaitis[a], Zoltan K. Nagy[a]*

[a]*Davidson School of Chemical Engineering, Purdue University, West Lafayette, IN, 47907, USA*
**Corresponding author's e-mail: zknagy@purdue.edu*

Abstract

The pharmaceutical industry has seen more interest in shifting from traditional fully batch operation to continuous manufacturing. With these shifts in mind and the adoption of industry 4.0 standards, digital tools are required to ensure critical medicines can be manufactured with quality guarantees. PharmaPy is one such tool that can create a digital twin of a pharmaceutical manufacturing process and enable the digital design of optimized manufacturing routes. This new tool has particular strengths in modelling batch, continuous and hybrid manufacturing systems. PharmaPy is shown throughout this work to be capable of digitally addressing quality-by-design (QbD) through design space identification for processes containing a variety of operating modes.

Keywords: design space, pharmaceutical manufacturing, process systems engineering, hybrid processing

1. Introduction

The development and transition of pharmaceutical manufacturing from predominantly batch processing to end-to-end continuous or hybrid operation mode has been prompted by benefits such as operational robustness, consistency of product quality and, potentially, lower environmental impact (Içten et al. 2020). However, these benefits typically are not trivial to identify during the process design phase, especially for hybrid, dynamic process systems.

Traditionally, a process operating region must be identified in which process inputs correspond to outputs that meet product quality constraints, also known as the design space. In pharmaceutical manufacturing, the Food and Drug Administration (FDA) launched the now well-known quality-by-design (QbD) paradigm (FDA, 2004), promoting process development of optimal manufacturing pathways through design space analysis with process robustness and quality assurance in mind.

With the widespread adoption of the QbD approach, a numerical tool that handles the simulation of end-to-end batch, hybrid, or end-to-end continuous systems, as well as laying the groundwork for in silico design space identification and other process system analysis, is desirable. Recently, we have developed a simulation platform, PharmaPy (Casas-Orozco et al. 2021a), structured as an object-oriented tool based on a set of robust declarative representations (Marquardt 1992), which offers a rich suite of first-principles, dynamic models for drug substance manufacturing. In addition to the process model

library and unit operation simulation capabilities, PharmaPy allows the creation of digital twins by offering both parameter estimation and statistical analysis capabilities for uncertainty quantification and propagation.

The numerical capabilities of PharmaPy regarding process simulation, parameter estimation and uncertainty quantification make the tool useful in determining the design space for a fixed pharmaceutical flowsheet or set of candidate flowsheets, thus facilitating the analysis of the combined effect of operational variables (flowrates, inlet compositions, etc.) and parametric uncertainty (e.g., kinetics and transport phenomena) (Bano et al. 2019; Laky et al. 2019). The evaluation of the design space under uncertainty makes the mapping of feasible operating regions more comprehensive and provides a valuable tool for process validation from a regulatory perspective.

In this work, a two-reactor process flowsheet for the synthesis of a low volume/high value active pharmaceutical ingredient (API) is presented as a case study to demonstrate the simulation/design-space generation framework. Different flowsheets resulting from the combination of reactor operating modes (batch and continuous) for a pilot scale production plant are evaluated in terms of waste generation (unreacted reagents) and API productivity. To account for uncertainty, parameter estimation was performed on one reaction system using PharmaPy to identify variance/covariance information regarding model parameters for reaction kinetics. Using this model uncertainty, probabilistic design spaces, representing feasible regions for process operation, are generated via adaptive sampling combined with appropriate uncertainty propagation strategies. Specifically, the probabilistic design space for a continuous process, a plug flow reactor (PFR) followed by a continuous stirred tank reactor (CSTR), is compared with that from a semibatch reactor to demonstrate the benefits in terms of probability of adhering to critical quality attributes of an API reaction product mixture.

2. Methodology

To perform probabilistic design space analysis, model parameters must have a tangible uncertainty representation to infer confidence that some chosen operating point will guarantee that all critical quality attributes (CQAs) are maintained. For this purpose, PharmaPy was used to fit experimental spectral data in a sequential manner. First, a multivariate, Principal Component Regression (PCR) calibration model was gathered to compute species concentrations during experiments. Real-time IR spectra were then gathered during several experiments at varying temperatures and molar ratios between reactants, which allowed to capture temperature dependence of the reaction parameters for a set of representative molar concentrations. Finally, experimental data were bootstrapped (Chernick, 2008) to generate the nominal values and parametric uncertainty of those reaction kinetics parameters. The parameter estimation results were first shown in Casas-Orozco et. al. (2021b).

Uncertainty is a key component of generating probabilistic design spaces. Generation of the probabilistic design space is as follows. First, the desired or proposed continuous operating region is transformed into a discrete counterpart by dividing each operating parameter range into equally-spaced increments, generating a finite number of operating points. Then, for each operating point, the model parameters are sampled 100 times. For each of these random samples, the PharmaPy model of the process is simulated.

Once the simulation has finished, the resulting state variables can be used to evaluate the set of CQAs. If all CQAs are met, (i.e. all critical quality constraints are less than zero),

this sample is marked as successful. The number of those samples that are successful at a given operating point are tallied and once those 100 samples finish, the probability that the current operating point provides output at acceptable quality is the number of successful samples divided by the number of samples run. Performing this analysis for the complete set of discretized operating conditions and interpolating the results provides a digital representation of the probabilistic design space of the given process. After analysis is complete, review of the probabilistic design space can lead to operating conditions that guarantee quality with a quantitative confidence.

3. Case Study

For this analysis, in-house data on the synthesis of Lomustine was used. Lomustine is an API used to treat brain cancer and has been a compound of interest while developing an end-to-end optimal pharmaceutical modelling framework using PharmaPy. Lomustine synthesis is generated through two synthesis steps:

$$\mathrm{CHA} + \mathrm{ISOCN} \overset{k_1}{\rightarrow} \mathrm{INT} \tag{1}$$

$$\mathrm{INT} + \mathrm{TBN} \overset{k_2}{\rightarrow} \mathrm{Lom} + \mathrm{TBOH} \tag{2}$$

Where CHA is cyclohexylamine, ISOCN is 1-Chloro-2-isocyanate, INT is the lomustine intermediate 1-(2-Chloroethyl)-3-cyclohexylurea, TBN is tert-Butyl nitrite, Lom is Lomustine, and TBOH is tert-Butyl alcohol. For this work, the form of reaction kinetics that were fit followed the standard Arrhenius rate law shown in Eq. (3).

$$k_i = A_i exp\left(\frac{-E_{a,i}}{R} * \left(\frac{1}{T}\right)\right) \tag{3}$$

Experiments have shown that the relative speed which reaction 1 occurs is sufficiently fast and thus can be modelled accordingly by setting low activation energy $E_{a,1}$ and high preexponential factor A_1. Parameters for the second reaction were estimated for varying temperatures as previously mentioned. The nominal parameter values for preexponential factors A_i and activation energies $E_{a,i}$ are shown below in Eq. (4).

$$\{A_1, A_2\} = \left\{2.1\ \frac{L}{mol*s}, 1.877 * 10^7\ \frac{L}{mol*s}\right\}\ ;\ \{E_{a,1}, E_{a,2}\} = \left\{2.0\ \frac{kJ}{mol}, 52.52\ \frac{kJ}{mol}\right\} \tag{4}$$

The reaction 2 parameters were transformed to exhibit improved numerical behaviour from the standard Arrhenius rate law such that the log of a reference preexponential factor at temperature T_{ref} and activation energy were fit from experimental data. This transformation is shown below in Eq. (5).

$$ln(k_2) = ln(A_{2,ref}) - \frac{E_{a,2}}{R}\left(\frac{1}{T} - \frac{1}{T_{ref}}\right) \tag{5}$$

Using this transformation, uncertainty in parameters $A_{2,ref}$ and $E_{a,2}$ were evaluated through bootstrapping and are shown as covariance matrix Σ below in Eq. (6).

$$\Sigma = \begin{bmatrix} 9.797 * 10^{-5} & -1.176 * 10^{-3} \\ -1.176 * 10^{-3} & 4.839 * 10^{-1} \end{bmatrix} \tag{6}$$

Using these reaction kinetics, two processing routes were explored, described below in Figure 1: Schematic of both process alternatives for the productions of Lomustine. On the left, the continuous flowsheet with two unit operations. On the right, the one-pot semibatch reaction. The first is a continuous reaction train using a PFR followed by a CSTR. Under normal operation, optimal design for the continuous reactors is a PFR volume of approximately 695 mL with a flowrate of 16.66 L/h (Laky et al., 2021). The CSTR volume is set at 20 L to ensure high residence time to achieve adequate conversion of intermediate to Lomustine. A schematic of the continuous process is shown below.

Also shown is the second processing route, a semibatch reactor. Here an initial loading of CHA and ISOCN are set to react for a small amount of time (approximately 30 minutes). Then, pure TBN is added over the course of 15 minutes. The system continues reacting until a total batch time of 4 hours is reached. It is known that Lomustine is unstable above 40 °C. Thus, all reactors begin at 25 °C and are heated to 35 °C to increase reaction rate while mitigating thermal degradation of the API. Throughout this work, both processes were modelled using PharmaPy.

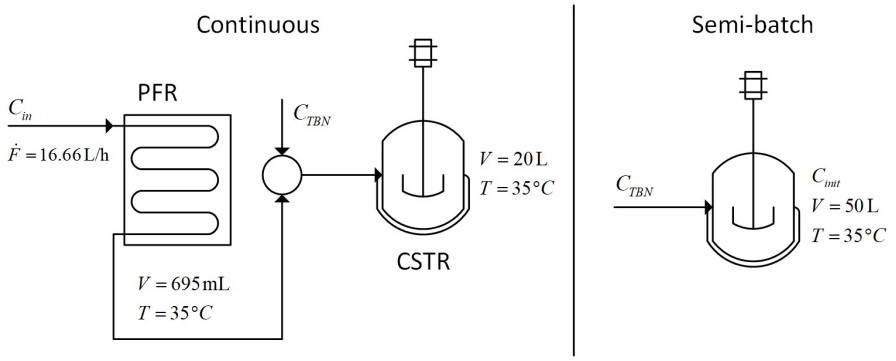


Figure 1: Schematic of both process alternatives for the productions of Lomustine. On the left, the continuous flowsheet with two unit operations. On the right, the one-pot semibatch reaction.

There were two operating parameters explored in this work. First is the inlet concentration or initial concentration of CHA and ISOCN in the first reactor, shown as C_{in} in Figure 1: Schematic of both process alternatives for the productions of Lomustine. On the left, the continuous flowsheet with two unit operations. On the right, the one-pot semibatch reaction. Second is the concentration of TBN for the second reaction step, shown as C_{TBN}. For this study, the concentration of CHA and ISOCN were fed or initialized in an equimolar ratio. Inlet concentration C_{in} has bounds from 0.04 M to 0.08 M and C_{TBN} from 0.06 M to 0.12 M. Both operating parameters were discretized with 11 points for a total of 121 total discrete operating points over the design space.

Purification of Lomustine using crystallization is the next processing step. With this in mind, CQAs can be defined as below:

$$C_{Lom} - \alpha C_{Lom,Sat} \leq 0 \tag{7}$$

$$C_{CHA} + C_{ISOCN} + C_{INT} + C_{TBN} - C_{Lom} \leq 0 \tag{8}$$

$$4.0\,C_{INT} - C_{Lom} \leq 0 \tag{9}$$

Eq. (7), ensures that no premature crystallization of Lomustine occurs. Eq. (8), ensures that all reagents are sufficiently converted to Lomustine. Eq. (9), ensures that the intermediate, which is known to crystallize, adequately converts to Lomustine.

4. Results

Using the processes and discretization schemes described above, 100 samples at each operating point were taken in the model parameter space, and the probabilistic design spaces were generated. These resulting design spaces are shown below in Figure 2: Contour plot of the probabilistic design space generated for both operating modes. Contours represented by confidence that an operating point adheres to CQAs as a fraction of 1.0 where lighter tones represent higher confidence of feasibility..

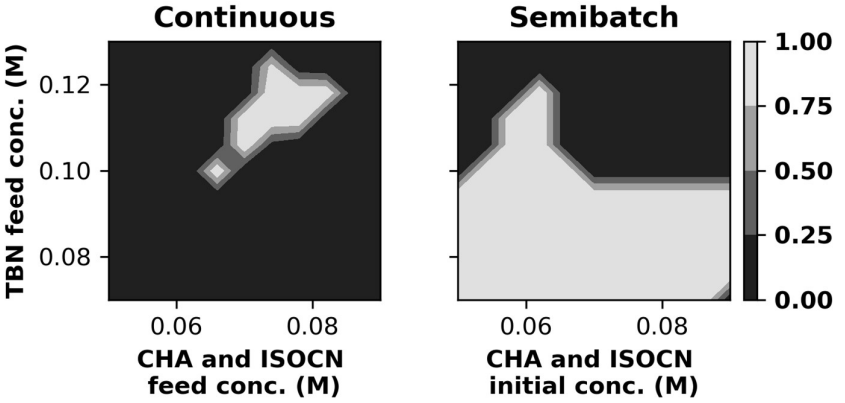


Figure 2: Contour plot of the probabilistic design space generated for both operating modes. Contours represented by confidence that an operating point adheres to CQAs as a fraction of 1.0 where lighter tones represent higher confidence of feasibility.

Generation of the design space for the continuous process took 243 minutes, whereas the semibatch design space took about 46 minutes (Macbook Pro 2016, 2.6 GHz Quad-Core i7 processor, 16GB 2133 MHz LPDDR3 RAM). The difference in simulation time for the continuous case results from the numerical solution of PDE models versus ODE models in the semibatch case. From the figure, the semibatch operating mode provides a much wider range of feasible operation than the continuous case. Nearly half of the explored operating space results in feasible operation, whereas only a small region is feasible in the continuous case.

The discrepancy in probabilistic design space between the two designs highlights the advantage of the high conversion of batch systems. In the continuous case, limitations on the residence time and volume of the CSTR lead to lower overall conversion. The integration of two unit operations with their individual design spaces, also leads to a smaller overall feasible design space for the integrated process. If lower flowrates were possible, the residence time of the CSTR could increase and lead a larger feasible operating region. The results indicate that while continuous integrated processes present numerous advantages compared to batch or semibatch systems, the overall robust design space for these tend to be smaller, and the implementation of suitable control systems can be critical to keep the process within the robust operating space.

5. Conclusions

In this work, PharmaPy was used to generate probabilistic design spaces of API synthesis routes using statistical sampling of experimentally determined uncertainty in the relevant model parameters. PharmaPy has strengths in easily modelling varying operational modes, from fully batch, to hybrid, to fully continuous operation. Using these strengths, the synthesis of a cancer drug API, Lomustine, was analysed for both continuous and semibatch operation. It was found that in the case of batch processing, when reaction time is long enough, conversion of raw materials to the API allows for a larger feasible operating space than in the continuous case. Also, these design spaces consider no control action, and likewise the feasibility region of the continuous manufacturing process will necessarily increase when control action is implemented.

Overall, PharmaPy is an emerging tool in leveraging digital twins and digital analysis in the pharmaceutical manufacturing space. It can be seamlessly called in many relevant digital process analysis techniques, such as those facilitating QbD and QbC. This is just one example highlighting these capabilities.

6. Acknowledgement

This project was supported by the United States Food and Drug Administration through grant U01FD006738. The views expressed by the authors do not necessarily reflect the official policies of the Department of Health and Human Services; nor does any mention of trade names, commercial practices, or organization imply endorsement by the United States Government.

References

G. Bano, P. Facco, M. Ierapetritou, F. Bezzo, and M. Barolo. 2019, Design Space Maintenance by Online Model Adaptation in Pharmaceutical Manufacturing, Computers and Chemical Engineering, 127, 254–71.

D. Casas-Orozco, D. Laky, V. Wang, M. Abdi, X. Feng, E. Wood, C. Laird, G. V. Reklaitis, and Z. K. Nagy, 2021a, PharmaPy: An object-oriented tool for the development of hybrid pharmaceutical flowsheets, Computers and Chemical Engineering, 153

D. Casas-Orozco, D. Laky, J. Mackey, A. Mufti, G. V. Reklaitis, and Z. Nagy, 2021b, Chemometric techniques for the combined calibration/parameter estimation of pharmaceutical drug substance manufacture, AIChE Annual Meeting 2021, Boston, MA

M.R. Chernick, 2008. Bootstrap Methods: A Guide for Practitioners and Researchers, 2nd ed. Wiley - Interscience, New Jersey.

FDA 2004, Pharmaceutical CGMPs for the 21s Century-A Risk-Based Approach, Rockville, MD

E. Içten, A. J. Maloney, M. G. Beaver, D. E. Shen, X. Zhu, L. R. Graham, J. A. Robinson, S. Huggins, A. Allian, R. Hart, S. D. Walker, P. Rolandi, and R. D. Braatz, 2020, A Virtual Plant for Integrated Continuous Manufacturing of a Carfilzomib Drug Substance Intermediate, Part 1: CDI-Promoted Amide Bond Formation, Organic Process Research & Development, 24, 10, 1861-1875

D. Laky, S. Xu, J. S. Rodriguez, S. Vaidyaraman, S. García Muñoz, and C. Laird. 2019. An Optimization-Based Framework to Define the Probabilistic Design Space of Pharmaceutical Processes with Model Uncertainty, Processes, 7 (2)

D. Laky, D. Casas-Orozco, G. V. Reklaitis, C. Laird, and Z. Nagy. 2021, Simulation-optimization framework for grey-box optimization using PharmaPy, AIChE Annual Meeting – Boston 2021

W. 1992. An Object-Oriented Representation of Structured Process Models, Computers and Chemical Engineering 16: S329–36

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50356-0

Hybrid Modelling of CHO-MK Cell Cultivation in Monoclonal Antibody Production

Kozue Okamura[a], Sara Badr[a*], Sei Murakami[b], Hirokazu Sugiyama[a]

[a]*Department of Chemical System Engineering, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, 113-8656, Tokyo, Japan*
[b]*Manufacturing Technology Association of Biologics, 2-6-16, Shinkawa, Chuo-ku, 104-0033, Tokyo, Japan*
badr@pse.t.u-tokyo.ac.jp

Abstract

Monoclonal antibodies (mAbs) are essential as drug substances and are getting increased attention from industry and academia. In order to improve mAb production efficiency, optimising the process is necessary. Optimising the cell cultivation step where mAbs are produced by host cells is especially important due to its influence on cost and time. Adequate process models are essential to perform the needed simulations. This work aims to propose a new modelling approach that avoids the failures of previously developed kinetic models with new cell lines. A new hybrid modelling approach is introduced, which combines a mechanistic module with a data-driven one to account for cell phases and varying environmental conditions. The developed hybrid approach was tailored for a newly developed Chinese Hamster Ovary (CHO) cell line (CHO-MK 9E-1). The hybrid approach gave a higher accuracy in the depiction of lactate and glucose concentrations compared to the mechanistic model alone. Insights are gained through analysing the results of the data-driven module. Such insights can be used as feedback in an opportunity to develop more versatile mechanistic models.

Keywords: Cell culture; Chinese hamster ovary (CHO) cells; Hybrid modelling; Principal component analysis; Biopharmaceutical production

1. Introduction

Monoclonal antibodies (mAbs) are essential as an active pharmaceutical ingredient. They are produced using Chinese Hamster Ovary (CHO) cells through the process shown in Figure 1. Process models are necessary to perform simulations for process optimisation and reduce the number of required experiments. In particular, models for the cell cultivations are required due to the high contribution of this step to the overall production cost and time (Yang et al., 2019).

Many efforts have been made to develop a kinetic cultivation model that represents the cultivation phenomena well. The recent works are well summarised by Tang et al. (2020). One of the latest works is done by Badr et al. (2021), and the proposed Monod-type-based model is validated with both the most common cell line (CHO-K1) and a newly established cell line (CHO-MK CL1001) cultivation.

However, despite of the efforts of many researchers in this field, developing a versatile kinetic model has been a challenge due to the complexity of the cell cultivation phenomena and variations in the cell lines. For example, previous models were not successful in depicting lactate concentrations for a new cell line (CHO-MK 9E-1).

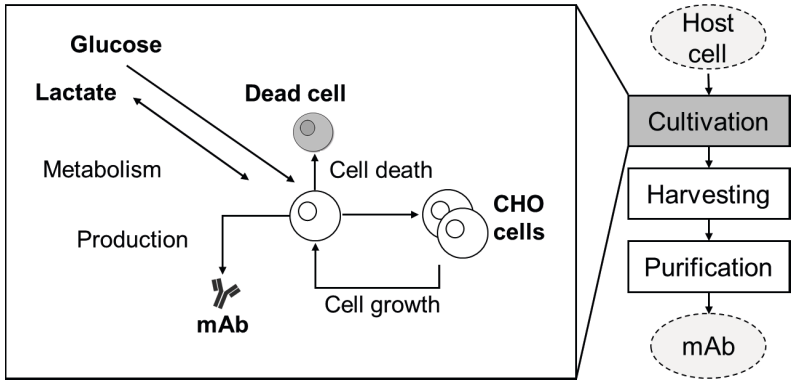


Figure 1: A monoclonal antibody (mAb) production process flow and cultivation phenomena.

This cell line is known to have a faster metabolism and is accordingly sensitive to lactate concentrations, which affects other aspects of cell viability and growth.

To overcome the challenge of needing to describe the impact of behaviour at different cell phases and environmental conditions, a hybrid modelling approach is introduced in this work. In this hybrid approach, a mechanistic module involving kinetic models of cell metabolism is combined with a data-driven one. Mechanistic models are easier to interpret but require a comprehensive understanding of the underlying mechanisms. On the other hand, data-driven models are easy to develop and apply even when the underlying mechanisms are not fully understood. However, they require a high volume of experimental data for model development and are more difficult to interpret or extrapolate. Hybrid modelling approaches aim to combine the advantages of both modelling techniques (Hong et al., 2018). The applications of hybrid modelling in pharmaceutical bioprocesses are well summarised by Narayanan et al. (2019).

This work proposes a new hybrid modelling approach that drastically improved the lactate modelling accuracy of CHO-MK 9E-1. A discussion is provided for the insights gained from the data-driven module to help improve the mechanistic understanding of the underlying phenomena. The developed model can subsequently be further exploited to achieve a better description of cell death and subsequent impurity generation, such as host cell proteins (HCP) and DNA.

2. Methodology

2.1. Experimental Data

Experimental data for model validation was obtained from the Kobe GMP consolidated lab of Manufacturing Technology Association of Biologics. Two cultivation modes were represented: the fed-batch mode, where nutrients continue to be fed to the cells throughout the operation, and the perfusion mode, where nutrients are continuously fed to the reactor and outlet streams are constantly removed to achieve continuous operations. Data from four experiments are used in this work: two experiments of CHO-MK 9E-1 cell in fed-batch mode (a 50 L stirred tank and a 2 L glass vessel, respectively); one experiment of CHO-K1 cell in fed-batch mode (a 50 L stirred tank); one experiment of CHO-MK CL1001 in perfusion mode (a 2 L glass vessel).

2.2. Cultivation Model

In this study, a hybrid modelling approach was introduced to improve the accuracy of lactate concentration modelling. The mechanistic module involved a kinetic model for fundamental cell metabolism. The model developed by Badr et al. (2021) was used in this work (Eqs.(1)-(7)). Badr et al. (2021) validated the model with the same experimental data of CHO-K1 cell in fed-batch mode and CHO-MK CL1001 in perfusion mode.

$$\frac{dVX_{\mathrm{V}}}{dt} = (\mu - \mu_{\mathrm{d}})VX_{\mathrm{V}} - F_{\mathrm{bleed}}X_{\mathrm{V}} \tag{1}$$

$$\frac{dVP}{dt} = Q_{\mathrm{P}}VX_{\mathrm{V}} - (F_{\mathrm{harvest}} + F_{\mathrm{bleed}})P \tag{2}$$

$$\frac{dV[\mathrm{GLC}]}{dt} = -\left(\frac{\mu-\mu_{\mathrm{d}}}{Y_{X_{\mathrm{V}}/\mathrm{glc}}} + m_{\mathrm{glc}}\right)VX_{\mathrm{V}} + F_{\mathrm{in}}c_{\mathrm{in}} + F_{\mathrm{suppl}}c_{\mathrm{suppl}} - (F_{\mathrm{harvest}} + F_{\mathrm{bleed}})[\mathrm{GLC}] \tag{3}$$

$$\frac{dV[\mathrm{LAC}]}{dt} = Y_{\mathrm{lac/glc}}VX_{\mathrm{V}} - (F_{\mathrm{harvest}} + F_{\mathrm{bleed}})[\mathrm{LAC}] \tag{4}$$

$$\frac{dV}{dt} = F_{\mathrm{in}} + F_{\mathrm{suppl}} - F_{\mathrm{harvest}} + F_{\mathrm{bleed}} \tag{5}$$

$$\mu = \mu_{\max}\left(\frac{[\mathrm{GLC}]}{K_{\mathrm{glc}}+[\mathrm{GLC}]}\right)\left(\frac{KI_{\mathrm{lac}}}{KI_{\mathrm{lac}}+[\mathrm{LAC}]}\right) \tag{6}$$

$$\mu_{\mathrm{d}} = k_{\mathrm{d}}\left(\frac{[\mathrm{LAC}]}{KD_{\mathrm{lac}}+[\mathrm{LAC}]}\right)\left(\frac{KD_{\mathrm{glc}}}{KD_{\mathrm{glc}}+[\mathrm{GLC}]}\right) \tag{7}$$

where, X_{V} is the viable cell density. V is the solution volume inside bioreactor. F_{bleed}, F_{harvest}, F_{in}, and F_{suppl} are flow rates of bleeding, harvesting, feeding, and supplementary glucose solution feeding respectively. P, $[\mathrm{GLC}]$, and $[\mathrm{LAC}]$ represent concentrations of mAb, glucose (GLC), and lactate (LAC), respectively. c_{in} and c_{suppl} are glucose concentrations in the feeding media and supplementary glucose solution. respectively. μ and μ_d are cell growth and death rates. μ_{max} and k_d are their maximum values. Q_P is the specific mAb production rate. Y is the yield coefficient. m_{glc} is the glucose consumption coefficient for cell maintenance. K is Monod parameter.

The mechanistic module is followed by a data-driven one based on principal component analysis followed by linear regression (PCR), where LAC and GLC concentrations are calculated according to Eq. (8).

$$[Y_j] = \sum_{i=1}^{N} a_i PC_i + C \tag{8}$$

where $[Y_j]$ represents the concentration of component *j* representing in this work (LCA or GLC). PC_i represents the i_{th} principal component derived from the principal

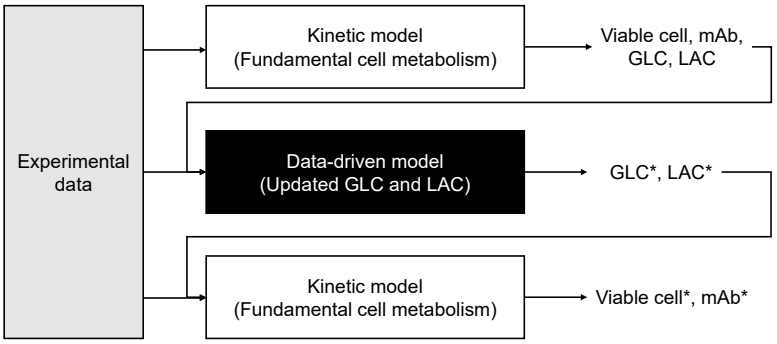


Figure 2: Outline of the hybrid modelling flow. * represents updated calculation results.

component analysis, a_i represents the coefficient of i_{th} principal component, N represents the appropriate number of principal components, and C is a constant.

Figure 2 shows how the two modules are combined in the hybrid approach. After fitting the kinetic model to the experimental data, the model parameters are obtained with initial predictions of viable cell density, and mAb, GLC, and LAC concentrations. These concentrations along with the estimated Monod parameters are fed as input to the data-driven module. Experimental conditions not included in the mechanistic model are also added to the data-driven module (e.g. dissolved oxygen (DO), pH, Temperature). The model input also includes time and information about the cell line and operating mode. The data-driven module then gives updated LAC and GLC concentrations as an output. These concentrations are fed back to the kinetic model to update the values of fitted parameters and obtain updated VCD and mAb values.

3. Results and Discussion

3.1. Cell Cultivation Modelling Results

Figure 3 shows part of the results of applying the model to predict concentrations in one experiment (CHO-MK 9E-1 in fed-batch mode with a 50L stirred tank). Lactate concentrations were better predicted by the hybrid model compared to the kinetic model alone. Other concentration profiles (GLC, viable cell, and mAb) were also well predicted. Similar results were obtained when using the model with each of the other experiments in the training set, with the hybrid model outperforming the kinetic model. A leave-one-out-cross-validation is also carried out to confirm model stability.

There is still a gap in viable cell toward the end of cultivation even though LAC modelling was improved. It might be interpreted as a need to further modify Eqs.(6) and (7), which define the effect of GLC and LAC on cell growth rate and death rate.

3.2. Insights from PCA loading trends

Figure 4 shows the cumulative explained variance ratio by each principal component (PC) and the categorised variable composition of each PC. With 9 PCs, the cumulative explained ratio reached 0.96. In order to make it easier to see the overall trend, the input variables were categorised into four groups: cell line and measurement factors; cell metabolism factors; environmental factors; operating factors.

In the case of PC_2 and PC_3, cell metabolism factors were the dominant ones. This indicates that such factors have a relatively high influence on the variance of lactate

profiles. They also indicate different cell behaviour at different operating conditions of each experiment despite using similar cell lines. For example, incorporating an additional mechanistic model describing lactate consumption by the cells, it might be possible to further describe the complex behaviour of lactate metabolism. The variables in this category are the outputs of the kinetic model. The data-driven module thus uses both experimental data and such initial kinetic parameters to close the gap between the experimental data and calculated results of the kinetic model.

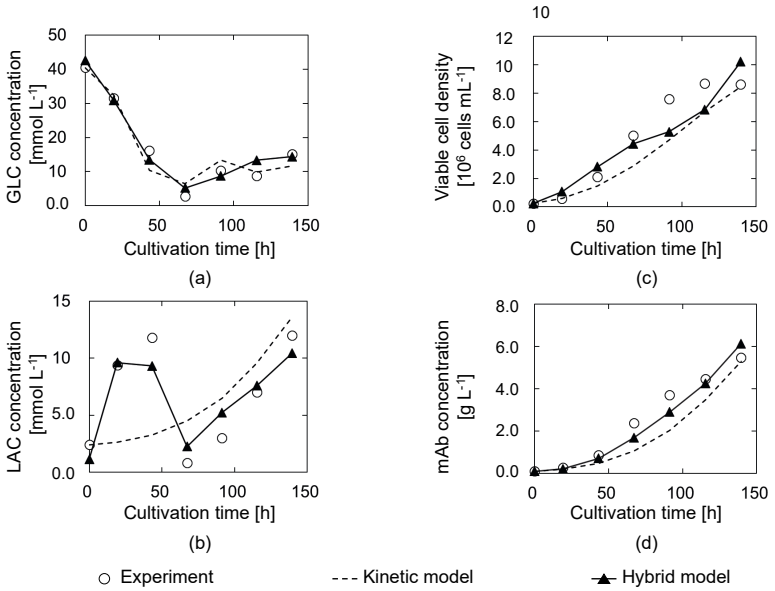


Figure 3: Modelling results of CHO-MK 9E-1 in fed-batch mode (a 50L stirred tank), showing (a) GLC concentration, (b) LAC concentration, (c) viable cell density, and (d) mAb concentration. The accuracy of the hybrid model are as follows: (a) GLC (R^2=0.961), (b) LAC (R^2=0.863), (c) viable cell (R^2=0.851), and (d) mAb (R^2=0.942)

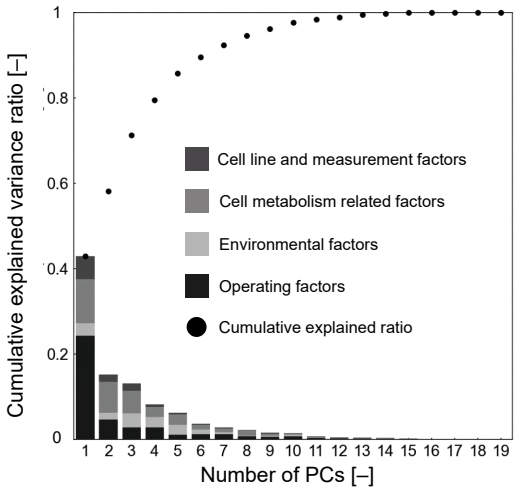


Figure 4: Cumulative explained variance ratio and categorised variable composition of each PC.

On the other hand, the ratio of environmental factors toward latter PCs becomes relatively larger along with cell metabolism factors. Here, by looking into the details of PC loadings, some possible relationships could be observed, such as between cell growth or death rate and DO or pH. This kind of insight could tell us the critical factors which have not been considered in the kinetic model yet. Utilising such feedback in kinetic model development might contribute to deeper mechanistic understanding towards obtaining more versatile kinetic models.

4. Conclusions

This work presents a new hybrid modelling approach for cell metabolism. The approach worked well with experimental data of different tested cell lines, including CHO-MK 9E-1, especially to achieve higher LAC modelling accuracy (R^2=0.863). Also, opportunities to utilise the insights obtained from the data-driven modelling part for further development of kinetic models were presented. Analysing the biological meaning of PC loading compositions and providing feedback to kinetic model development might contribute to a deeper understanding of cell cultivation phenomena. Experiments to further confirm the validity of the modelling approach with further cell lines and experimental conditions are still required. Additionally, exploration of different data-driven approaches is still required to enhance the modelling versatility and performance.

Acknowledgements

This work was supported by the Japan Agency for Medical Research and Development (AMED) [grant No. JP21ae0121015, JP20ae0101064, JP20ae0101058].

References

S. Badr, K. Okamura, N. Takahashi, V. Ubbenjans, H. Shirahata, H. Sugiyama, 2021, Integrated design of biopharmaceutical manufacturing processes: Operation modes and process configurations for monoclonal antibody production, Comput Chem Eng, 153, 107422

M. S. Hong, K. A. Severson, M. Jiang, A. E. Lu, J. C. Love, R. D. Braatz, 2018, Challenges and opportunities in biopharmaceutical manufacturing control, Comput Chem Eng, 110, 106-114

H. Narayanan, M. Sokolov, M. Morbidelli, A. Butté, 2019, A new generation of predictive models: The added value of hybrid models for manufacturing processes of therapeutic proteins, Biotechnology and Bioengineering, 116, 10, 2540-2549

O. Yang, S. Prabhu, M. Ierapetritou, 2019, Integrated Design of Biopharmaceutical Manufacturing Processes: Operation Modes and Process Configurations for Monoclonal Antibody Production, Comparison between Batch and Continuous Monoclonal Antibody Production and Economic Analysis, Ind. Eng. Chem. Res., 58, 15, 5851-5863

P. Tang, J. Xu, A. Louey, Z. Tan, A. Yongky, S. Liang, Z. J. Li, Y. Weng, S. Liu, Kinetic modeling of Chinese hamster ovary cell culture: factors and principles, Crit. Rev. Biotechnol., 40, 2, 265-281

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50357-2

Multimodal modelling of uneven batch data

Atli Freyr Magnússon[a,b], Jari Pajander[b], Gürkan Sin[b], Stuart M. Stocks[b]

[a]*Technical University of Denmark, Søltofts Plads 227, Kgs. Lyngby 2800, Denmark*
[b]*Leo Pharma A/S, Industriparken 55, Ballerup 2750, Denmark*
flydkl@leo-pharma.com

Abstract

This work explores the application of a novel tri-linear regression methodology known as Shifted Covariates REgression Analysis for Multi-way data (SCREAM) to predict the quality of a fed-batch process. The SCREAM model shows promise as it is the only known multilinear regression tool that can directly handle three-way data arrays of different lengths. Thus, it provides an alternative modelling tool that does not require complicated time warping methods as a preprocessing step. The model was tested on a simulated fed-batch dataset based on industrial simulation of penicillin production. Variations were intentionally included in the simulations to create uneven data arrays. The SCREAM model outperforms traditional staples of multivariate models like NPLS and UPLS when warping is not considered and thus shows promise for application in fed-batch processes.

Keywords: Fed-Batch, Multimodal Modelling, PLS, Multivariate analysis

1. Introduction

The biochemical industry readily uses fed-batch processes for the production of various chemicals and pharmaceuticals. It is common in the industry to utilize a recipe driven approach, where operations are not adjusted to accommodate the variations in feed or initial blending of the batch. Biochemical processes by their very nature will introduce batch-to-batch variations if the system is not very tightly controlled. Knowledge of how the system variations may affect the final batch yield or quality can be vital for proper recovery of the final product. Most industries monitor key process parameters throughout the process, resulting in large amount of data. This has given rise to assumption free modelling (Westad et al, 2015). Data driven approaches can be utilized to establish correlations between captured data and the process quality.

For most data-driven regression methods, it is required that the data is available as two-dimensional matrix. However, batch data consisting of K batches is a three-way dataset $\underline{\boldsymbol{X}}(I\ x\ J\ x\ K)$ where each batch is measured on I process variables at J time points, while quality variables may be measured only after the conclusion of the batch. For a regression type problem, the most common method of modelling three-way array is to reshape the 3-way array into a 2-D matrix in a process known as unfolding. If preserving the batch mode, the resulting matrix will be $\boldsymbol{X}(KxIJ)$. The most common multivariate analysis is to use Partial Least Squares (PLS) on the unfolded matrix, the overall process is referred to as UPLS (Wold et al, 1987). There also exist tri-linear models that work directly with 3-way arrays. They are useful if one expects the multilinear structure to affect the overall system variation. For regression purposes the easiest one to use would be the multilinear PLS or NPLS (Bro, 1996). Direct trilinear

models are more strict than working with unfolded data and thus usually do not necessarily perform better when looking at prediction errors. They are still useful as they preserve the multi-linear structure and have more interpretative properties.

However, unfolding the matrix or utilizing NPLS requires that the lengths of the batch data matrix $(I\ x\ J)$ be consistent across all batches. Variations of batch operation and initial blend will create uneven data with shifts and shape changes, where different events or peaks in data take place at different times. The simplest way to handle uneven lengths of batches is to identify the shortest runtime of a batch and remove excess time points from all others, a process known as cut-to-shortest. In case of severe shifts or variation in batch length, it is possible to synchronize the data structure in a process known as time warping (Gonzáles-Martinéz et al, 2018).

Alternatively, the Shifted Covariates REgression Analysis for Multi-way data (SCREAM) method (Marini and Bro, 2013) can be used. This method is the multi-modal version of the Principal Covariates Regression (PovR), but it utilizes a Parallel Factor Analysis 2 (PARAFAC2) decomposition method (Kiers 1991) instead of the two-way equivalent Principal Component Analysis (PCA). PARAFAC2 is a decomposition method that allows for shifts and uneven lengths in a single mode. This makes SCREAM an interesting candidate for modelling batches as it can directly model the data without unfolding the three-way structure and requires no time warping as a preprocessing step.

2. Materials and Methods

2.1. SCREAM Model

The SCREAM model utilizes a PARAFAC2 fitting algorithm based on an Alternating Least Squares approach. PARAFAC2 models are expressed as

$$\boldsymbol{X}_k = \boldsymbol{A}\boldsymbol{D}_k\boldsymbol{B}_k^T + \boldsymbol{E}_k \qquad k = 1, \dots K \tag{1}$$

Here $\boldsymbol{X}_k$ is a single slab of the entire three-way structure $\underline{\boldsymbol{X}}$, or in this case the data from a single batch. For a PARAFAC2 model with F components, the matrix $\boldsymbol{A}$ is a matrix $(I\ x\ F)$ of loadings in the I direction. For batch data, this is usually the variable loadings. $\boldsymbol{D}_k$ is a diagonal matrix $(I\ x\ F)$ containing the k'th row of the matrix $\mathbf{C}(K\ x\ F)$ which contains the loadings in the K or batch direction. $\mathbf{C}$ is similar to a score matrix in ordinary 2-way PCA. Finally, $\boldsymbol{B}_k$ is the loadings in the J direction or the time point direction. Generally, $\boldsymbol{B}_k$ hold the loadings where the shifts happen. Finally, $\boldsymbol{E}_k$ contains the residuals. PARAFAC2 models are made unique by the constraint that the cross-product of each $\boldsymbol{B}_k$ is the same i.e. $\boldsymbol{B}_k\boldsymbol{B}_k^T = \boldsymbol{H}$ for all $k = 1, \dots K$. The standard PARAFAC does not have a unique loading matrix in the J direction for each slab $\boldsymbol{X}_k$ but rather uses a single $\boldsymbol{B}$ for the entire three-way structure. Different $\boldsymbol{B}_k$ loadings allow PARAFAC2 to directly model three-way arrays of batch data of different lengths but also makes it more flexible when handling shifts in batch data.

Fitting a PARAFAC2 model is the least squares minimization of the following loss function

$$\sum_{k=1}^{K} ||(\boldsymbol{X}_k - \boldsymbol{A}\boldsymbol{D}_k\boldsymbol{B}_k^T)||^2 \tag{2}$$

Note that $\mathbf{C}$ is a 2D matrix and a direct multi-linear regression onto $\mathbf{Y}$ is possible. This would be the multimodal equivalent to Principal Component Regression (PCR). However, there is no guarantee that the score matrix $\mathbf{C}$ is predictive of $\mathbf{Y}$ as it is attempting to summarizing the entire $\underline{\boldsymbol{X}}$ array. Thus, changes in $\underline{\boldsymbol{X}}$ that may have no significance on the output $\mathbf{Y}$, will still affect the $\mathbf{C}$ matrix.

For prediction purposes it is sought to seek a score matrix $\mathbf{C}$ that is relevant for predicting $\mathbf{Y}$. For a single dependent variable $\mathbf{y}$, it is achieved by minimizing the following

$$||\boldsymbol{y} - \boldsymbol{Cr}||^2 \quad (3)$$

where $\mathbf{r}$ is a vector of regression coefficients. Making a predictive model that is relevant for both $\underline{\boldsymbol{X}}$ and $\boldsymbol{y}$ requires the minimization of both equations (2) and (3). This is the same setup as in the two-way Principal Covariate Regression (PCovR) where a weighing parameter $\alpha\ 0\ \leq \alpha\ \leq 1$ is introduced. This parameter controls to what degree the fitting should summarize $\underline{\boldsymbol{X}}$ or predict $\boldsymbol{y}$. The SCREAM model is then fitted by minimizing the following

$$\alpha \sum_{k=1}^{K} ||(\boldsymbol{X}_k - \boldsymbol{A}\boldsymbol{D}_k\boldsymbol{B}_k^T)||^2 + (1-\alpha)||\boldsymbol{y} - \boldsymbol{Cr}||^2 \quad (4)$$

The PARAFAC2 direct fitting algorithm (Kiers et al 1999) is utilized to solve this minimization problem while maintaining the uniqueness constraint.
This modelling techniques has two hyperparameters that must be selected, the number of factors F and the value of the weighing parameter α. Improper selection of these parameters leads to models that do not predict or overfit on $\boldsymbol{y}$. For most practical applications, optimization of the hyperparameters is done via cross-validation. A common method is minimizing the Root Mean Square Error of Cross-Validation.

2.2. Preprocessing

Preprocessing is important for any multivariate or multi-modal methods. Centering removes constant offsets in the data, while the obvious reason for scaling is to adjust for scale differences between variables measured in different units. Since the goal is for the model to capture the variations between different batches, the centering will be across the batch mode. This is done by computing a separate mean for the measured variable for each time point across batches and then subtracting the mean from each measurement.
Scaling is performed on the centered data. There are three types of scaling for three-way data. Column scaling, single-slab scaling, and double-slab scaling. This study will utilize the single-slab scaling technique as it has shown better performance in regression modeling of batch data. (Mears et al, 2016). For single-slab scaling all time points for a single variable are scaled to unit root-mean-square.

3. Case Study: Industrial Simulation of Penicillin Fermentation

Exploring the applicability of the SCREAM method for modelling batch fermentations a dataset is needed. For this purpose, the industrial simulation model (Birol et al, 2002s) is used for simulating an industrial fed-batch fermentation for the production of penicillin. 100 batches are simulated with a recipe driven approach given in Table 1, Batch-to-batch variations are created by varying the initial conditions of biomass and glucose in each batch, the total runtime of each batch is also varied, as well as the time to start the feed. For simplification variables that are controlled such as pH and Temperature are not simulated but a random Gaussian noise is added around the set point for each time index.
In the generated dataset, a total of 11 process variables are assumed to be monitored continuously on each batch which will be utilized for multi-modal modelling. The

collection of the 11 process variables at equal timepoint distances of one hour will create the three-way array $\underline{\boldsymbol{X}}$.

Table 1 Recipe used in the fed-batch simulation for penicillin production

Recipe Setup	Value	Potential variations
Initial Biomass	0.5 g/L	±0.05 g/L
Initial Glucose	40 g/L	±2 g/L
Initial Volume	100 m^3	±100 L
Feed Rate Set Point	200 L/h	±1 L/h
Feed Glucose Concentration	600 g/L	±10 g/L
Agitation	Maintain kla of 400 h^{-1}	±1 on kla
Aeration	1200 m^3/h	±20 m^3/h
Temperature Set Point	298 K	±0.05 K
pH Set Point	5	±0.01
Fermentation Duration	400 h	±10 h
Feed Start Time	50 h	±4 h

The quality variable $\boldsymbol{y}$ of interest is the total harvested Penicillin at the end of the batch. Figure 1 shows the batch-to-batch variation of penicillin harvest. Because of the variation in runtime of up to 20 hours, the three-way array $\underline{\boldsymbol{X}}$ is uneven which is common in industrial batch processes.

4. Results and Discussions

The SCREAM model was built to predict the penicillin harvest based on monitored variables. Because the data is simulated the data there are observable effects of tank geometry or start time of batches, thus the partition into modelling and test set is done randomly. 30 Batches are held over for testing while the model is built on 70 batches. Utilizing a 10-fold venetian blinds cross validation it is found that setting F=5 and $\alpha = 0.3$ provided a decent regression without overfitting.

Model predictions are shown in Figure 2 where predicted harvest from SCREAM model and measured harvest from the simulation are reported. The Root Mean Square Error of the test set was calculated to be 13.01 which results in a less than 5% absolute error when predicting the harvest of new batches.

For comparison the traditional three-way NPLS model is built as well as the most common type of multivariate model UPLS for batch data. To make the batches even for these types of models the cut-to-shortest time warping method is used. The performance of these models is shown in Table 3. Both NPLS and UPLS started overfitting the data with 3 components but with only 1 component it failed to establish any correlation between process data and penicillin harvest, thus 2 components were used for building the models.

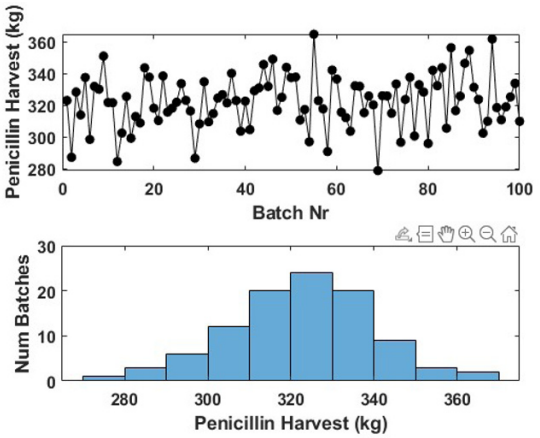


Figure 1 Batch-To-Batch variations in the penicillin harvest of the simulated dataset

However, SCREAM outperforms both when looking at the test set. This may be due to NPLS being too strict to handle the shifts, but unlikely as the UPLS is more flexible but has the same problems when predicting the penicillin harvest. The most likely issue is that the data at the end of the batch contains crucial information for establishing the correlations which is missing with a cut-to-shortest time warping. Dynamic Time Warping (DTW) or Correlation Optimized Warping (COW) may improve the performance of both UPLS or NPLS models but are difficult and time-consuming methods to employ. However, SCREAM models do not require these treatments and thus are able to predict the test-batches with more accuracy.

Table 2: Comparison of different regression models on the simulated dataset.

Model Type	RMSE (Calibration Set)	RMSE (Test Set)
SCREAM	11.23	13.01
NPLS	6.98	17.61
UPLS	7.82	18.00

5. Conclusions

A regression method that can directly model uneven batch data was tested on a simulated industrial fed-batch dataset. The SCREAM model was developed as a modification to Multivariate covariate regression, which allows it to handle shifts in dataset and uneven three-way matrices. The fed-batch simulation is based on an industrial model and disturbances were introduced in the recipe to create the variations in runtime and shifts the data, which is common in industry. The SCREAM model outperforms both NPLS and UPLS when used on the data when complex warping techniques are not utilized with no overfit and little bias. There is even room for improvement by utilizing non-derivative optimization method for the model hyperparameters, it may well be possible to find a set of hyperparameters that outperforms the model reported here.

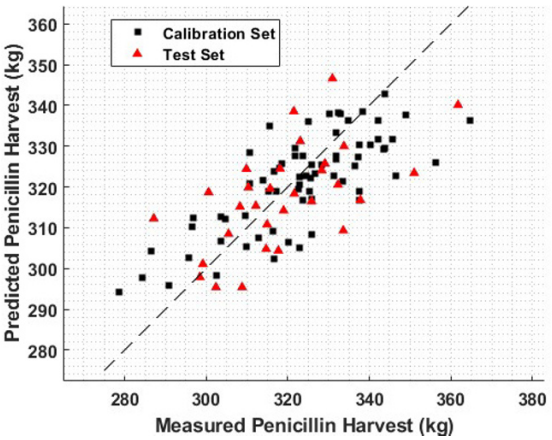


Figure 2: Predicted Penicillin harvest from the SCREAM model vs the actual simulated value.

Further work will be analysing the loadings and diagnostics of the model to see if the model can identify the key variables and time points of an industrial fermentation to extract information without access to the original simulation. Overall, the SCREAM model technique is a promising alternative in the modelling fed-batch systems.

References

G. Birol, C. Ündey, A. Çinar, 2002, A modular simulation package for fed-batch fermentation: penicillin production, Computers and Chemical Engineering, 26, 11 1553-1565

R. Bro, 1996, Multiway calibration. Multilinear PLS, Journal of Chemometrics, 10, 1, 47-61

J.M. Gonzáles-Martínez, J. Camarcho, A. Ferrer, 2018, MVBatch: A matlab toolbox for batch process modeling and monitoring, Chemometrics and Intelligent Laboratory Systems, 183, 122-133

H. Kiers, 1989, An alternating least squares algorithm for fitting the two- and three-way dedicom model and the idoscal model, Psychometrika, 54, 3, 515-521

H. Kiers, J. Ten Berge, R. Bro, 1999, PARAFAC2 – Part I. A direct fitting algorithm for the PARAFAC2 model, Journal of Chemometrics, 13, 3-4, 275-294

F. Marini, R. Bro, 2013, SCREAM: A novel method for multi-way regression problems with shifts and shape changes in one mode, Chemometrics and Intelligent Laboratory Systems, 129, 64-75

L. Mears, R. Nørregård, G. Sin, K. Gernaey, S. Stocks, M. Albaek, K. Villez, 2016, Functional Unfold Principal Component Regression Methodology for Analysis of Industrial Batch Process Data, AiChE Journal, 62, 6, 1986-1994

F. Westad, L. Gidskehaug, B. Swarbrick, G. Flåten, 2015, Assumption free modeling and monitoring of batch processes, Chemometrics and Intelligent Laboratory Systems, 149, 66-72

S. Wold, P. Geladi, K. Esbensen, J. Öhman, 1987, Multi-way principal components-and PLS-analysis, Journal of Chemometrics, 1, 1, 41-56

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50358-4

Application of MHE-based NMPC on a Rotary Tablet Press under Plant-Model Mismatch

Yan-Shu Huang[a*], M. Ziyan Sheriff[a], Sunidhi Bachawala[b], Marcial Gonzalez[b,c], Zoltan K. Nagy[a], Gintaras V. Reklaitis[a]

[a]*Davidson School of Chemial Engineering, Purdue University, West Lafayette, IN 47907, USA*
[b]*School of Mechanical Engineering, Purdue University, West Lafayette, IN 47907, USA*
[c]*Ray W. Herrick Laboratories, Purdue University, West Lafayette, IN 47907, USA*
huan1289@purdue.edu

Abstract

Active control strategies play a vital role in modern pharmaceutical manufacturing. Automation and digitalization are revolutionizing the pharmaceutical industry and are particularly important in the shift from batch operations to continuous operation. Active control strategies provide real-time corrective actions when departures from quality targets are detected or even predicted. Under the concept of Quality-by-Control (QbC), a three-level hierarchical control structure can be applied to achieve effective setpoint tracking and disturbance rejection in the tablet manufacturing process through the development and implementation of a moving horizon estimation-based nonlinear model predictive control (MHE-NMPC) framework. When MHE is coupled with NMPC, historical data in the past time window together with real-time data from the sensor network enable model parameter updating and control. The adaptive model in the NMPC strategy compensates for process uncertainties, further reducing plant-model mismatch effects. The frequency and constraints of parameter updating in the MHE window should be determined cautiously to maintain control robustness when sensor measurements are degraded or unavailable. The practical applicability of the proposed MHE-NMPC framework is demonstrated via using a commercial scale tablet press, Natoli NP-400, to control tablet properties, where the nonlinear mechanistic models used in the framework can predict the essential powder properties and provide physical interpretations.

Keywords: pharmaceutical manufacturing; continuous manufacturing; process control; nonlinear model predictive control; moving horizon estimation.

1. Introduction

Several factors currently drive the transition of the pharmaceutical manufacturing industry from batch to continuous process operation. These include potential improvement in both product quality homogeneity and process controllability. Quality control traditionally followed a Quality-by-Testing (QbT) approach, wherein product quality was tested at the end of each batch processing step. However, with improved product and process understanding, a Quality-by-Design (QbD) approach was adopted to enable systematic design of the operating space using mechanistic models. More recently, there has been a desire to adopt a Quality-by-Control (QbC) approach, wherein quantitative and predictive understanding can be leveraged for active process control and aid robust process design and operation, thereby enabling smart manufacturing (Su et al., 2019).

An important part of any real-time process monitoring and control strategy is the ability to identify and manage the impact of plant-model mismatch (PMM). PMM can arise in the continuous manufacture of oral solid dosage for numerous reasons, e.g., disturbances that affect critical material attributes (CMAs) such as the bulk density can be introduced during the feeder refill step (Destro et al., 2021). As this can result in a deviation in the critical quality attributes (CQAs), PMM needs to be identified and handled appropriately. Several approaches have been developed in order to identify and assess the impact of PMM, e.g., based on mutual-information (Chen et al., 2013) or autocovariance (Wang et al., 2017). Stringent regulations placed by regulatory bodies make it essential to track CQAs and CMAs online, but they may be unmeasurable in practice as process analytical technology (PAT) sensing methods may not be available to track these states or parameters, e.g., bulk density. Therefore, this work proposes the use of an on-line, real-time parameter estimation approaches to accurately track model parameters online, to guide operating decisions. It is important to note that most work in the continuous manufacturing domain utilize linear model predictive control (MPC) strategies, that are derived from the linearization of the nonlinear system and may not be adequate for nonlinear process models and unit operations such as the rotary tablet press (Ierapetritou et al., 2016).

A recent in-silico study by (Huang et al., 2021) demonstrated that a combined MHE-NMPC framework could satisfy the dual requirement of efficient estimation and control. Unfortunately, there are no case studies in the literature that demonstrate the application of the proposed framework to real data from a continuous pharmaceutical manufacturing process. Therefore, the primary objective of this work is to validate the practical applicability of the proposed framework using a Natoli NP-400 rotary tablet press.

2. Methodology

The moving horizon estimation-based nonlinear model predictive framework (MHE-NMPC) aims to satisfy the dual requirement of estimation and control, by combining the effective estimation capabilities of MHE with the control performance provided by NMPC. Given a nonlinear state-space model:

$$\dot{x} = g(x, u, \theta, w) \tag{1}$$

$$y = l(x, u, \theta, v) \tag{2}$$

where x, u, θ, and y are vectors that represent the state variables, input variables, model parameters, and measurements, respectively. Process and measurement noise are denoted by w and v, respectively. In this work, the model is described by a set of explicit algebraic equations with no differential states, and f and h will represent these algebraic equations. MHE can then be formulated as follows (López-Negrete and Biegler, 2012):

$$\min_{\hat{\theta}_k} J = \sum_{t=k-N_{past}}^{k} (\epsilon_t)^T W_E \, \epsilon_t + \left(\hat{\theta}_k - \hat{\theta}_{k-1}\right)^T W_\theta \left(\hat{\theta}_k - \hat{\theta}_{k-1}\right) \tag{3a}$$

subject to

$$\hat{x}_{k-N_{past}+j+1} = f\left(\hat{x}_{k-N_{past}+j}, u_{k-N_{past}+j}, \hat{\theta}_k\right) \tag{3b}$$

$$\hat{y}_{k-N_{past}+j} = h(\hat{x}_{k-N_{past}+j}) \tag{3c}$$

$$\epsilon_{k-N_{past}+j} = y_{k-N_{past}+j} - \hat{y}_{k-N_{past}+j} \tag{3d}$$

$$\hat{x}_{k-N_{past}+j+1} \in \mathbb{X}, \qquad \epsilon_{k-N_{past}+j} \in \Omega_{\epsilon}, \qquad \hat{\theta}_k \in \Omega_{\theta} \tag{3e}$$

$$j = 0, 1, \dots, N_{past} \tag{3f}$$

where $\hat{\theta}_k$ are estimated uncertain parameters, bounded in compact set Ω_{θ}. y_t and u_t are measurements of output and input variables at time *t*, respectively; $\hat{y}_t$ and $\hat{x}_t$ are estimated output and state values, respectively; ε_t are output disturbances, bounded in compact set Ω_{ϵ}; and W_E and W_{θ} are weighting matrices. Once the MHE optimization problem is solved at time $t = k$, the estimated state $\hat{x}_{k-N_{past}+1|t=k}$ is chosen as the initial state value for the next time step $t = k + 1$, i.e., $\hat{x}_{k-N_{past}+1|t=k+1} = \hat{x}_{k-N_{past}+1|t=k}$.

This study utilizes the median of the error distribution in the past time window to represent output disturbances ζ_k at time $t = k$, i.e.,

$$\zeta_k = median\left\{\epsilon_{k-N_{past}+j}\right\}, \quad \text{for } j = 0, 1, \dots, N_{past} \tag{4}$$

The NMPC framework at time $t = k$ is defined as follows:

$$\min_{\Delta u_t} J = \sum_{t=k}^{k+N_p} (\hat{y}_t - y_{sp})^T \, W_y \left(\hat{y}_t - y_{sp}\right) + \sum_{t=k}^{k+N_c-1} (\Delta u_t^T W_{\Delta u} \Delta u_t) \tag{5a}$$

subject to

$$\hat{x}_{k+j+1} = f\left(\hat{x}_{k+j}, \hat{u}_{k+j}, \hat{\theta}_k\right) \tag{5b}$$

$$\hat{y}_{k+j} = h\left(\hat{x}_{k+j}\right) + \zeta_k \tag{5c}$$

$$\Delta u_{k+j} = \hat{u}_{k+j+1} - \hat{u}_{k+j} \tag{5d}$$

$$\hat{x}_{k+j} \in \mathbb{X}, \qquad \hat{u}_{k+j} \in \mathbb{U}, \qquad \Delta u_{k+j} \in \Omega_{\Delta \mathrm{u}} \tag{5e}$$

$$j = 0, 1, \dots, N_p - 1 \tag{5f}$$

where N_c is the length of the control time window, and y_{sp} are the setpoints of the output variables. W_y and $W_{\Delta u}$ are weighting matrices. Control movements Δu are constrained in compact set $\Omega_{\Delta u}$. A detailed discussion of the MHE-NMPC framework including its computational feasibility is provided in (Huang et al., 2021).

3. Case Study

3.1. Tablet press model

The tablet press is responsible for the formation of solid tablets via mechanical compression. The weight of a convex tablet *W* and the tablet production rate $\dot{m}_{tablet}$ are given by the following relationships (Huang et al., 2021):

$$W = \rho_b V_{fill} \left(1 - \xi_1 \frac{n_T}{n_F} + \xi_2 \frac{H_{fill}}{D}\right) \tag{6}$$

$$\dot{m}_{tablet} = W n_T N_{station} \tag{7}$$

where *D*, V_{fill}, H_{fill}, ρ_b, n_T, and n_F, are the diameter of the die, volume of the die cavity, dosing position, powder bulk density, turret speed, and feed frame speed, respectively. $N_{station}$ refers to the number of stations in the tablet press. ξ_1 and ξ_2 are empirical model parameters that are estimated from experimental data. The volume of the die cavity for the D-type tooling is provided by (Huang et al., 2021).

The pre-compression force F_{pc} and the main compression force F_{punch} can be computed as follows:

$$F_{pc} = \frac{\pi D^2}{4b}\left[\frac{\rho^{pc}-\rho_c}{\rho^{pc}(a-1)+\rho_c}\right] \tag{8}$$

$$\rho^{pc} = \frac{W}{V^{pc}\rho_t} \tag{9}$$

$$F_{punch} = \frac{\pi D^2}{4b}\left[\frac{\rho^{in-die}-\rho_c}{\rho^{in-die}(a-1)+\rho_c}\right] \tag{10}$$

$$\rho^{in-die} = \frac{W}{V^{in-die}\rho_t} \tag{11}$$

where parameters a and b are Kawakita constants, which represent the maximum degree of compression and the reciprocal of the pressure applied to attain this degree of compression, respectively. ρ^{pc} and ρ^{in-die} are the pre-compression and in-die relative densities, respectively. ρ_t refers to the true density of the powder. The pre-compression volume, V^{pc}, and in-die volume V^{in-die} are provided by (Huang et al., 2021).

3.2. Applying MHE to tablet press: Comparison of fixed model and adaptive model

To investigate state estimation and parameter updating, an experiment was performed via open-loop control. Setpoint changes of input variables were introduced to the tablet press, and corresponding measurements of output variables and model predictions are recorded and shown in Figure 1 (a) with fixed model parameters and Figure 1 (b) with adaptive model parameters.

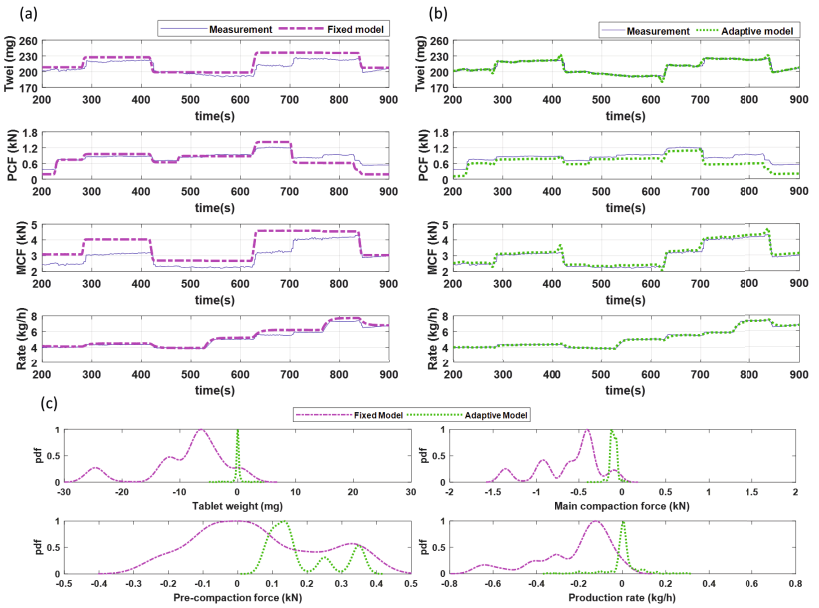


Figure 1. Real-time monitoring of PMM when uncertain parameters are (a) fixed or (b) adaptive with (c) error distribution of estimated output variables.

In this study, since the mathematical model is represented by a set of explicit algebraic equations the MHE only updates two uncertain parameters: (1) the bulk density (ρ_{bulk}),

to compensate for the effects of disturbance on the estimated value of the tablet weight, which further affects estimated values of pre-compression force, main compression force, and the production rate, and (2) the critical relative density (ρ_c) to provide flexibility to estimate the pre-compression and main compression forces more accurately in the adaptive model compared to the case of only bulk density being updated. The adaptive model predicts output variables more accurately compared to the fixed model.

To quantify model accuracy and precision, error distributions of the estimated output variables are provided in Figure 1 (c), where the probability density functions (pdf) of the error distributions are rescaled (meaning that the area under the density curve is not 1). Once MHE is applied to update the bulk and critical relative densities, the absolute values of the median error and error spans of all output estimations are significantly reduced, as shown in Figure 1 (c). However, an exception is found in pre-compression force, whose error span is reduced from 0.75 kN to 0.37 kN, while median error is increased from 0.07 kN to 0.22 kN. Since pre-compression force and main compression force share the same model parameters as shown in Equation (8-11), there exists the need to establish a compromise between the accuracy of these two output variables.

3.3. Experimental verification of MHE-NMPC

Control profiles for a representative experimental run of the 4 input variables, 4 output variables, and 2 uncertain model parameters are shown in Figure 2 (a), (b), and (c), respectively. Offsets in the output variables are observed as open-loop control is applied at the start of operation (highlighted in red). When the MHE-NMPC algorithm is implemented from t = 200 s, offset free control is achieved. Additional setpoint changes are introduced for the tablet weight at t = 600 s, 800 s, 1500 s, the main compression force at t = 1100 s, and the production rate at t = 1000 s.

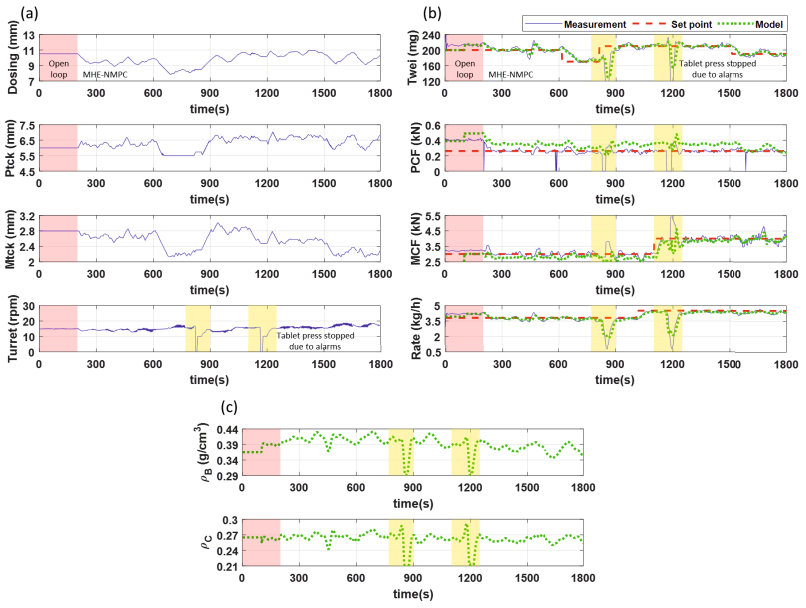


Figure 2. MHE-NMPC control performance of the tablet press with (a) input variables, (b) output variables, and (c) uncertain parameters.

All results of setpoint tracking are satisfactory as shown in Figure 2 (b). During the experimental run, an internal alarm stopped the tablet press twice at t = 820 s and t = 1160 s (highlighted in yellow) forcing the turret speed to drop to 0 rpm. The machine stop prevents the distributed control system (DCS) from collecting in-house tablet weight and production rate measurements, where the time delay can be attributed to the 10 s moving average window. The strength of the MHE-NMPC algorithm can once again be noted, as offset-free control is quickly achieved once the tablet press resumes operation. As the uncertain parameters are updated in real-time as shown in Figure 2 (c), the mismatch can be mitigated as presented in Figure 2 (b). While mismatch for pre-compression force cannot be completely mitigated, as it shares the same parameters as the main compression force, the disturbance term used in controller model described in Equation 5c still guarantees offset-free control of the pre-compression force.

4. Conclusions

Real-time process monitoring and control are essential to enable continuous operation of modern pharmaceutical manufacturing processes. The MHE-NMPC framework demonstrates satisfactory control performance and parameter updating in the rotary tablet press to handle plant-model mismatch (PMM). Future work will include sensor fusion studies to incorporate at-line measurements with long sampling time to the framework. Accurate estimation is required to enable the control of critical quality attributes such as tensile strength, which need to be predicted from soft sensors due to limited availability of real-time measurements because of the destructive nature of the testing methods used.

Acknowledgement

This work was supported by the United States Food and Drug Administration under grant 1U01FD006487-01. The authors would like to thank Natoli Engineering Company for the availability of the tablet press and Carmelo Hernandez-Vega for his technical support.

References

Chen, G., Xie, L., Zeng, J., Chu, J., Gu, Y., 2013. Detecting Model–Plant Mismatch of Nonlinear Multivariate Systems Using Mutual Information. Industrial & Engineering Chemistry Research 52, 1927–1938.

Destro, F., García Muñoz, S., Bezzo, F., Barolo, M., 2021. Powder composition monitoring in continuous pharmaceutical solid-dosage form manufacturing using state estimation – Proof of concept. International Journal of Pharmaceutics 605, 120808.

Huang, Y.-S., Sheriff, M.Z., Bachawala, S., Gonzalez, M., Nagy, Z.K., Reklaitis, G. v., 2021. Evaluation of a Combined MHE-NMPC Approach to Handle Plant-Model Mismatch in a Rotary Tablet Press. Processes 9, 1612.

Ierapetritou, M., Muzzio, F., Reklaitis, G., 2016. Perspectives on the continuous manufacturing of powder-based pharmaceutical processes. AIChE Journal 62, 1846–1862.

López-Negrete, R., Biegler, L.T., 2012. A Moving Horizon Estimator for processes with multi-rate measurements: A Nonlinear Programming sensitivity approach. Journal of Process Control 22, 677–688.

Su, Q., Ganesh, S., Moreno, M., Bommireddy, Y., Gonzalez, M., Reklaitis, G. v, Nagy, Z.K., 2019. A perspective on Quality-by-Control (QbC) in pharmaceutical continuous manufacturing. Computers & Chemical Engineering 125, 216–231.

Wang, S., Simkoff, J.M., Baldea, M., Chiang, L.H., Castillo, I., Bindlish, R., Stanley, D.B., 2017. Autocovariance-based MPC model mismatch estimation for systems with measurable disturbances. Journal of Process Control 55, 42–54.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50359-6

Gray-box modelling of pharmaceutical roller compaction process

Shuichi TANABE[a*], Shubhangini AWASTHI[b], Daiki KAKO[a], and Srikanth R. GOPIREDDY[b]

[a]*Formulation Technology Research Laboratories, Daiichi Sankyo Co., Ltd., Kanagawa 254-0014, Japan*
[b]*Formulation Technology, Daiichi Sankyo Europe GmbH, Pfaffenhofen 85276, Germany*
[*]*Corresponding author's email: tanabe.shuichi.h3@daiichisankyo.co.jp*

Abstract

A novel method was developed that allows the rolling theory of granular solids (Johanson, 1965) to predict the ribbon density accurately. In this study, a gray-box model of roller compaction process was developed based on the rolling theory of granular solids to demonstrate a practical application of the accurate and descriptive roller compaction model for process development. A placebo formulation composed of mannitol, microcrystalline cellulose, and magnesium stearate was used to generate 26 samples of roller compaction experiment. Compressibility factor and elastic recovery rate were predicted using regression models to consider the dependence on material attribute and process parameter (PP). The gray-box model composed of the modified rolling theory proposed by Reynolds et al. (2010) and the complemental regression models showed a better prediction performance compared to the white-box model reported by Reynolds et al. The root mean square error of cross validation of the ribbon density in white-box and gray-box models were 0.07 g/cc and 0.04 g/cc, respectively. With the gray-box model the effect of PP on the ribbon density and mass throughput was visualized, which is beneficial to identify the target and the acceptable ranges of PP for manufacturing.

Keywords: Dry granulation, Modeling and Simulation, Ribbon density, Control strategy.

1. Introduction

The granulation process is a critical process that impacts the quality of pharmaceutical products such as the dissolution and the uniformity of active ingredient content of tablets, capsules, powder filled bottles, etc. Roller compaction is a dry granulation processes for producing granules from powder blends. In roller compaction powder blends are continuously compressed by the two counter-rotating rolls and subsequently the generated ribbons are milled to obtain the granules. In general, ribbon density is considered as a critical material attribute (CMA), i.e., a factor of drug product quality and therefore it needs to be controlled via process parameter (PP) such as roll force, roll speed, and roll gap. The pressure on the powder blends in the roller compactor directly correlates to the ribbon density. Due to the difficulty in direct pressure measurement, process models that use the measurable PP to predict process outputs are needed to control the drug product quality.

Johanson (1965) developed the rolling theory of granular solids, which correlates the powder compaction in roller compaction process with the raw material properties and the PP. In the rolling theory, the space in between the rolls are divided into three different regions, i.e., the slip region, the nip region, and the release region. A sketch of the three different regions are provided in Figure 1a. In the slip region, powder blends slip along with the rotation of rolls whereas in the nip region, the powders are trapped by the rolls and move at the same speed as the roll surface, which results in the compaction of the powders to form roller compacted ribbons. In the release region, the ribbons show elastic recovery due to the release from compaction force by the rolls. The nip angle α is the transition angle from slip to nip region.

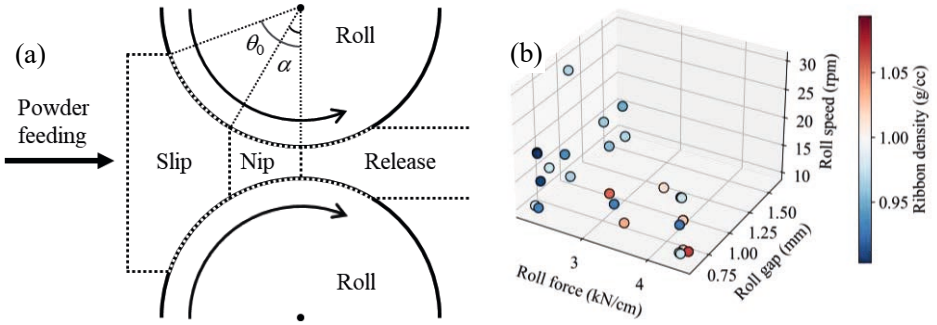


Figure 1 (a) Sketch of the three different regions in roller compaction. (b) Scatter plot of experimental runs with different PP setting.

The maximum pressure at the minimum separation of rolls, which is defined as roll gap, is estimated from the pressure distribution equation as a function of roll angle. The output parameters such as ribbon density and mass throughput are calculated based on the estimated maximum pressure. Due to the difficulty in measuring the required pre-consolidation pressure, limited works have been reported with focus on the practical application of the rolling theory for process understanding and control. Reynolds et al. (2010) proposed a modified rolling theory as a practical compromise with an assumption that the pre-consolidation pressure is 1 MPa. Based on this assumption, the pre-consolidation density and the compressibility factor are considered constant regardless of the PP, and were estimated from uniaxial experiments to predict ribbon density and mass throughput. While their approach provided accurate prediction of ribbon density, it has a limitation in predicting ribbon density accurately because their approach does not consider the effect of the PP on the pre-consolidation pressure. As presented by Reynolds et al., the pre-consolidation density and compressibility factor are different between the uniaxial measurement and the roller compaction process, which indicates that the pre-consolidation density and compressibility factor depend on the PP. Besides, the rolling theory and the variations do not consider the elastic recovery of the ribbons in release region even though the ribbon thickness is larger than the roll gap as reported in previous studies (Shi and Sprockel, 2016; Souihi et al., 2013). Since the off-line ribbon density measurement uses samples collected from release region, the relationship between the PP and the elastic recovery rate should be considered for higher prediction accuracy.

Gray-box model is one of the practical solutions to complement prediction performance of theoretical models with the use of data while keeping the theoretical structure

(Ahmad et al., 2020). The gray-box model is expected to present a higher interpretability compared to the black-box model, i.e., data-driven model. The gray-box models have been successfully applied to chemical and pharmaceutical processes (Van sprang et al., 2005; Ahmad et al., 2020). However, application of the gray-box model to the roller compaction process is not reported so far. In this study, a gray-box model was demonstrated on the rolling theory for the first time to show its practical applicability to the process development and control strategy setting.

2. Materials and Methods

A placebo formulation composed of Pearlitol 100SD as mannitol (Roquette, France), Ceolus UF-711 as microcrystalline cellulose (Asahi Kasei, Japan), and HyQual 5712 as magnesium stearate (Mallinckrodt, USA) in a weight ratio percentage of 79:20:1 was used for this study. A 5-L V-blender was used to manufacture the powder blends for roller compaction in 2-kg/batch scale. A roller compactor FP90 (Freund Turbo corporation, Japan) that have a roll diameter of 90 mm and a roll width of 30 mm with the textured roll surface was used for producing ribbons. 26 runs of roller compaction experiments were performed using two different lots of powder blends. Figure 1b shows the scatter plot of experimental runs with different PP setting.

2.1. Physical testing of powder blends and ribbons

The bulk and tapped density of powder blends were measured using a graduated cylinder. A quantity of the powder blends was poured to a graduated cylinder. The net weight of the material and the bulk volume were recorded. The tapped volumes were recorded using a tapping apparatus SZ-02 (Rinkan Kogyo Co., Ltd., Japan). The flow properties of the powder blends were measured using FT4 powder rheometer (Freeman Technology, UK). A 25 mm vessel in diameter was used to assess effective angle of internal friction (EAIF) and wall friction angle (WFA). The ribbon density was measured using GeoPyc 1365 and AccuPyc 1340 (Micrometrics, USA). A sample chamber with an internal diameter of 25.4 mm was used for GeoPyc 1365. The ribbon thickness was measured immediately after the sampling of ribbons during experiments using thickness gauge.

2.2. The modified rolling theory of granular solids and the gray-box model

The rolling theory of granular solids modelled the stress gradient in the slip and the nip region along with the roll angle θ as shown in Eq.(1) and Eq.(2).

$$\left(\frac{d\sigma}{d\theta}\right)_{slip} = \frac{4\sigma\left((\pi/2)-\theta-v\right)\tan\delta}{D/2\,(1+S/D-\cos\theta)\left(\cot\left(\left((\theta+v+(\pi/2))/2\right)-u\right)-\cot\left(((\theta+v+\pi/2)/2)+u\right)\right)} \tag{1}$$

$$\left(\frac{d\sigma}{d\theta}\right)_{nip} = \frac{K\sigma_\theta(2\cos\theta-1-S/D)\tan\theta}{(D/2)\left((1+S/D-\cos\theta)\cos\theta\right)} \tag{2}$$

where

$$v = \left(\pi - \sin^{-1}\left(\frac{\sin\varphi}{\sin\delta}\right)-\varphi\right)/2 \tag{3}$$

$$u = (\pi/4)-(\delta/2) \tag{4}$$

K is compressibility factor, δ is EAIF, $\emptyset$ is WFA, S is roll gap, and D is roll diameter. The nip angle α at which the pressure gradients for the slip and nip regions were equal is determined by equating Eq.(1) and Eq.(2). In the Reynolds' modified rolling theory, the pressure distribution between the rolls was used to relate the roll force R_{f}, which is a common and a measurable PP with the peak pressure P_{max} applied at the minimum separation of rolls as given in Eq.(5) and Eq.(6).

$$R_{\mathrm{f}} = \frac{P_{\mathrm{max}} WDF}{2} \tag{5}$$

where

$$F = \int_{\theta=0}^{\theta=\alpha(\delta,\emptyset,K)} \left[\frac{S/D}{(1 + S/D - \cos\theta)\cos\theta}\right]^{K} \cos\theta \, d\theta \tag{6}$$

The Eq.(5) and Eq.(6) represent a relationship between PP (R_{f}, S), geometric parameters (D, roll width W), and material properties of powder blends (δ, $\emptyset$, and K). Furthermore, the relationship between the material density and P_{max} was defined in Eq.(7) based on the assumption that the pre-consolidation pressure P_0 equals 1 MPa.

$$\gamma_{\mathrm{R}} = \gamma_0 \left(\frac{P_{\mathrm{max}}}{P_0}\right)^{\frac{1}{K}} \tag{7}$$

where γ_{R} is relative ribbon density, γ_0 is relative pre-consolidation blend density.

The elastic recovery rate β is defined as a ratio of ribbon thickness to roll gap. In the proposed gray-box model, K and β are assumed to be PP and material attribute dependent to capture unconsidered relationship in the modified rolling theory. The pre-consolidation density was assumed equal to bulk tapped density ρ_0 to provide further practicability. Utilizing regression models for K and β the ribbon density ρ_R and mass throughput M_{R} were calculated based on the Eq.(8) and Eq.(9).

$$\rho_{\mathrm{R}} = \rho_0 (P_{\mathrm{max}})^{1/K} / \beta \tag{8}$$

$$M_{\mathrm{R}} = \pi DNWS\rho_0 (P_{\mathrm{max}})^{1/K} \tag{9}$$

where N is roll speed. Gaussian process regression (GPR) was used as a regression method to cope with nonlinearity. In the regression analysis, material attributes of powder blends (bulk loose density, bulk tapped density, EAIF, WFA) and PP (roll force, roll gap, roll speed) were used as input parameters to predict K and β. The input and output parameters were centered by subtracting mean values and scaled by dividing by sample standard deviation (SD), which is a so-called auto-scaling. In the regression analysis, root mean squared errors of cross validation (RMSECV) and coefficient of determination (R2) for the ribbon density in leave-one-out cross validation (LOOCV) was used to show the validity of the gray-box model. The RMSECV and R2 for the white-box model according to the Reynolds' modified rolling theory were also evaluated using the mean K and β in the calibration set in LOOCV to compare the prediction performance.

3. Results

Figure 2 shows the prediction performances of the white-box and gray-box models in LOOCV. The gray-box model improved prediction accuracy of the white-box model in terms of the RMSECV (0.07 g/cc for white-box model and 0.04 g/cc for gray-box model) and R2 values (0.19 for white-box model and 0.54 for gray-box model). With the consideration of the sampling error in normal process variation, the prediction performance of the gray-box model for the ribbon density was practically sufficient for process development. The lower prediction accuracy in the white-box model could be derived from variations of the K and β in the calibration set. The RSD% of the K and β in the calibration set were 11.1% and 12.8%, respectively.

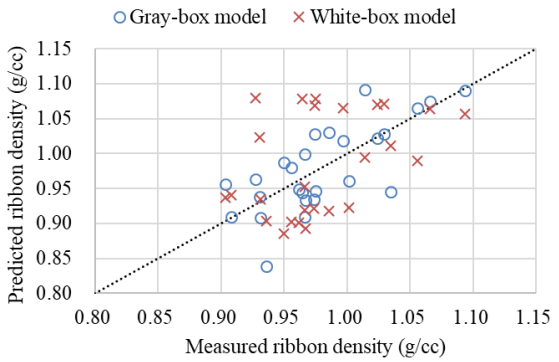


Figure 2 Scatter plot of measured and predicted ribbon density

To understand the relationship between the input and output parameters captured in the gray-box model, the effect of PP on the ribbon density and mass throughput was visualized into the contour plot as a function of roll gap and roll speed, see Figure 3a. The contour plot suggested that roll gap had a positive impact on both ribbon density and mass throughput. On the other hand, roll speed affected mass throughput only and had little impact on ribbon density. Same analysis can be performed to observe the effect of roll force on ribbon density and mass throughput. The reliability of the estimated impacts was also visualized using the normalized SD of the GPR prediction as shown in Figure 3b. The color progression from white to black represents the SD of the expected values in GPR, i.e., the predicted values of K and β. The highest SD is represented with black and color tones become closer to white as the SD decreases. In principle, the expected values with high SD in GPR prediction suggested that there are few data points in the calibration dataset around the predicting points and therefore the prediction is less reliable. Therefore, the PP with the higher SD of predicted K and β suggested that the predicted impacts are less reliable and will need additional experiments to clarify its actual responses. The threshold of SD to adopt the expected values and the gray-box model outputs would be determined considering actual variance of the observed data. These estimations provide a clear view on the control of PP to achieve target outputs and will be a justification for the future experiments.

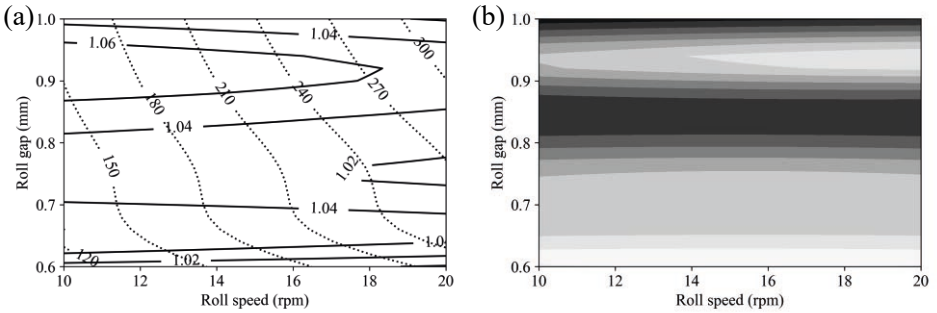


Figure 3 (a) Contour plots of the predicted ribbon density (solid line, g/cc) and mass throughput (dotted line, g/min). (b) Contour plot of the SD of the expected values in GPR. The color progression from white to black represents the SD of the expected values.

Overall, the applicability of the gray-box model for the roller compaction process was demonstrated. The gray-box model showed a higher prediction accuracy at the observed data points compared to the white-box model. A robust estimate of the effect of PP on the process outputs was presented by accounting for the SD of the expected value in the GPR model introduced in the gray-box model. In conclusion, the proposed gray-box model provides excellent applicability to the roller compaction process development.

4. Conclusions

The gray-box model composed of the rolling theory of granular solids for roller compaction process and the complemental regression models was presented to achieve high prediction accuracy and interpretability. The approach was found to provide reliable predictions for the process outputs such as ribbon density and mass throughput based on the material attributes, PP, and the geometric parameters. With the gray-box model the criticality of the PP were successfully visualized together with the reliability of prediction by means of the SD of the expected values, which would contribute to the process development activities such as risk assessment and design space setting.

References

I. Ahmad, A. Ayub, M. Kano, I.I. Cheema, 2020, Gray-box soft sensors in process industry: current practice, and fugure prospects in era of big data, Processes, 8, 243.

J.R. Johanson, 1965, A rolling theory for granular solids, Journal of Applied Mechanics, 32, 4, 842–848.

G. Reynolds, R. Ingale, R. Robert, S. Kothari, B. Gururajan, Practical application of roller compaction process modeling, Computers and Chemical Engineering 34, 1049-1057.

W. Shi, O.L. Sprockel, 2016, A practical approach for the scale-up of roller compaction process, 106, 15-19.

N. Souihi, M. Josefson, P. Tajarobi, B. Gururajan, J. Trygg, 2013, Design space estimation of the roller compaction process, Industrial & Engineering Chemistry Research, 52, 12408-12419.

E. N. M. Van sprang, H.-J. Ramaker, J.A. Westerhuis, A.K. Smilde, 2005, Statistical batch process monitoring using gray models, AIChE Journal, 51, 3 931-945.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50360-2

Multi-objective optimisation for early-stage pharmaceutical process development

Mohamad H. Muhieddine[a], Shekhar K. Viswanath[b], Alan Armstrong[c], Amparo Galindo[a], Claire S. Adjiman[a*]

[a]*Department of Chemical Engineering, The Sargent Centre for Process Systems Engineering, and Institute for Molecular Science and Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom*
[b]*Lilly Research Laboratories, Indianapolis, IN 46082, USA*
[c]*Department of Chemistry and Institute for Molecular Science and Engineering, Imperial College London, Molecular Sciences Research Hub, White City Campus, Wood Lane, London W12 0BZ, United Kingdom*
c.adjiman@imperial.ac.uk

Abstract

The pharmaceutical industry is under constant pressure to deliver its products quickly and effectively while minimising development costs and pursuing green pharmaceutical manufacturing methods. Given the many considerations in process development, a model-based method that takes multiple performance metrics into account is proposed for early process development. Several key performance indicators are identified, namely environmental footprint, cost, and conversion, selectivity, and yield. We employ multi-objective optimisation to assess the trade-offs between capital cost as one objective, and selectivity or conversion as a second objective, while exploring the interdependencies between all performance indicators. The approach is applied to two multiphasic reactions, each occurring in a 6-stage cascade CSTR: the hydrogenation of 4-Isobutylacetophenone (4-IBAP) to 1-(4-Isobutylphenyl)ethanol (4-IBPE) and the carbonylation of 4-IBPE to Ibuprofen (IBP).

Keywords: pharmaceutical process development, continuous manufacturing, multi-objective optimisation, Ibuprofen.

1. Introduction

The pharmaceutical industry faces increasing pressure to reduce development costs and time, and to meet increasingly stringent environmental regulations (Montes et al. 2017). Process Systems Engineering methods can help to improve process design/performance, enhance process understanding, and select optimal processing materials and operating conditions while exploring trade-offs between conflicting performance objectives. However, they have not yet been fully deployed for process development within the pharmaceutical and fine chemicals industries (Papadakis et al. 2018).

The complexity of the relationships between key process performance indicators (KPIs) such as feasibility, productivity, economics and environmental impact make it essential to adopt a holistic approach to process design. These KPIs are strongly linked to process-wide decisions such as process structure, unit size, operating conditions and even to molecular-level decisions (Adjiman et al. 2014). Pharmaceutical process development is a complex activity involving multiple pharmaceutically relevant

objectives that need to be satisfied (Nicolaou and Brown, 2013). The coupling of process modelling with multi-objective optimisation allows for the quantification of process metrics that are of interest to pharmaceutical manufacturing, and the evaluation of trade-offs between conflicting KPIs. The use of model-based approaches to process development is especially relevant in continuous manufacturing, an increasingly important area, given the potential technical and economic benefits of implementing flow technology (McWilliams et al. 2018).

In this work, several relevant KPIs are identified and their use within a multi-objective design framework that can be employed in early-stage process development is explored. The approach is illustrated on the modelling and multi-objective optimisation of two multiphase reactors involved in the production of Ibuprofen via the Hoechst pathway (Elango et al. 1991), with a focus on moving to continuous production. Case Study 1 consists of a catalytic hydrogenation reactor to convert 4-Isobutylacetophenone (4-IBAP) to 1-(4-Isobutylphenyl)ethanol (4-IBPE), while Case Study 2 is focused on a homogeneous carbonylation reactor for the subsequent production of Ibuprofen (IBP).

2. Methodology

The methodology can be summarized as follows. First, a set of KPIs is defined and a reactor configuration is chosen. Next, a conceptual process model is developed using kinetic rate and gas solubility equations obtained from the literature or experimental investigations. Finally, a multi-objective optimisation problem involving KPIs as objectives and constraints is formulated and solved.

2.1. Selected KPIs and process configuration

For the optimal design of pharmaceutical processes, it is important to identify, quantify, and assess the interdependencies between pharmaceutically relevant KPIs. In this work, the KPIs of interest include the capital cost C_R (\$) of the reactor system, the reaction selectivity S, the overall conversion X of raw materials, the yield Y of the pharmaceutical compound, and the Environmental factor (E-factor), defined as the mass ratio of waste product to desired product, and used to quantify environmental footprint. Mathematical expressions of these KPIs for Case Study 1 are shown in Table 1, where the expression of C_R is obtained from Douglas (1988). All symbols are defined in Table 2, where species j includes IBAP, IBPE, IBEB, H_2O and oligomers. Furthermore, the successful implementation of continuous flow technology for multiphase reactions requires efficient phase mixing and long residence times. While tubular reactors equipped with static mixers offer these features, they necessitate high volumetric flow rates which may be incompatible with slow reactions. Chapman et al. (2017) developed a multistage continuous-stirred tank reactor (CSTR) suitable for multiphasic reaction systems. Additionally, employing multiple CSTRs in series enhances system performance, minimising the total reactor volume required to achieve a specific conversion. Accordingly, a multi-stage cascade CSTR configuration is chosen for the studied multiphasic reactions. A schematic of such cascade reactor system with N reactors in series is shown in Figure 3.

2.2. Conceptual process model

A model of the cascade reactor system is developed to give a best case assessment of its performance. At this early stage of design, mass transfer limitations are neglected and thermodynamic equilibrium is assumed. The reactors are treated as isothermal.

2.2.1. Obtaining kinetic rate and gas solubility equations

Kinetic data for the hydrogenation of 4-IBAP over Pd/SiO_2 catalyst in *n*-decane as the solvent, following the scheme in Figure 1, are taken from Thakar et al. (2007). The solubility of hydrogen in the liquid phase is estimated using the Henry's law constant of Trinh et al. (2015). Kinetic and thermodynamic data for the carbonylation step in butanone (Figure 2) are taken from Seayad et al. (2003).

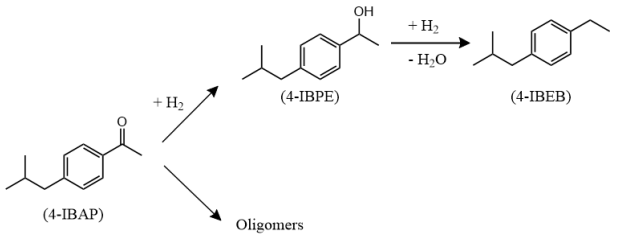


Figure 1 Reaction Scheme of the Catalytic Hydrogenation of 4-IBAP

The carbonylation of 4-IBPE is achieved using a homogeneous palladium complex catalyst dissolved in a water/butanone solvent mixture, as follows:

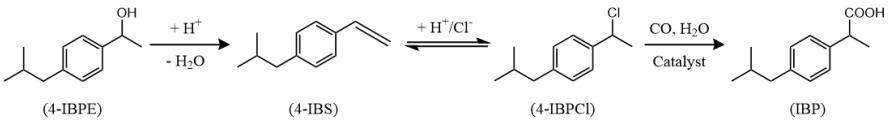


Figure 2 Reaction Scheme of the Homogeneous Carbonylation of 4-IBPE

2.2.2. Deriving flow reactor material balance equations of all reaction species

The hydrogenation system is modelled as a series of 6 triphasic CSTRs with side-stream addition of pure hydrogen. Similarly, the carbonylation reactor is modelled as a series of 6 biphasic CSTRs with side-stream addition of pure carbon monoxide. The steady-state material balance equations of all reaction species are then derived for each case study.

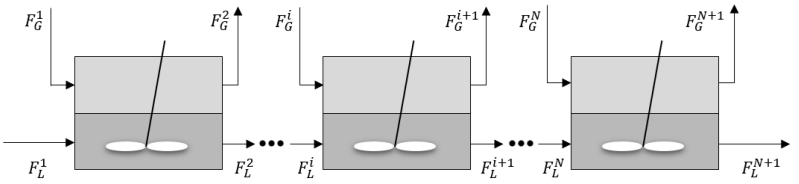


Figure 3 Gas-Liquid Cascade CSTR System

2.3. Formulating and solving the multi-objective reactor optimisation problems

For Case Study 1, the objectives of the optimisation problem are minimising cascade reactor capital cost and maximising overall conversion. For Case Study 2, the objectives are minimising reactor cost and maximising selectivity for all $X \geq 90\%$. The optimisation problems are solved using the ε-constraint method of Haimes et al. (1971). The second objective of Case Study 1 is transformed into a constraint, lower-bounding conversion by a series of ε values ranging between 50% and 98%. Similarly, the second objective of Case Study 2 is transformed into a constraint, lower-bounding selectivity by a series of ε values ranging between 70% and 98%. The identified KPIs are reported for every solution. Bounds on decision variables are imposed based on the experimental conditions of the corresponding kinetic study. The formulated problem is nonlinear and is solved using the CONOPT solver in the GAMS software, version 25.0.3. The formulation for carbonylation is similar, with the ε-constraint imposed on selectivity.

Table 1 Reactor KPIs for the Hydrogenation Case Study

KPI	Mathematical Expression	
C_R	$\frac{M\,\&\,S}{280} \times 482.37 \times \left[2.18 + F_m\left(0.0255P_{H_2}^2 + 0.0387P_{H_2} + 1.0136\right)\right] \times V_T^{0.6287}$	(1)
X	$\frac{F_{IBAP}^1 - F_{IBAP}^{N+1}}{F_{IBAP}^1}$	(2)
S	$\frac{F_{IBPE}^{N+1}}{F_{IBPE}^{N+1} + F_{IBEB}^{N+1} + F_{H_2O}^{N+1} + F_{oligomers}^{N+1}}$	(3)
Y	$\frac{F_{IBPE}^{N+1}}{F_{IBAP}^1}$	(4)
$E-factor$	$\frac{\dot{m}_{n-decane} + \dot{m}_{H_2,L}^{N+1} + \dot{m}_{IBAP}^{N+1} + \dot{m}_{IBEB}^{N+1} + \dot{m}_{H_2O}^{N+1} + \dot{m}_{oligomers}^{N+1}}{\dot{m}_{IBPE}^{N+1}}$	(5)

Table 2 Nomenclature for the hydrogenation case study

Symbol	Description	Units
$M\&S$	Marshall and Swift equipment index	dimensionless
F_m	Material correction factor	dimensionless
P_{H_2}	Inlet hydrogen pressure	MPa
V_T	Total reactor volume	m^3
F_{IBAP}^1	Inlet molar flow rate of IBAP to first reactor	kmol/sec
F_j^{N+1}	Outlet molar flow rate of species j from reactor N	kmol/sec

$\dot{m}_{n-decane}$	Mass flowrate of n-decane	kg/s
$\dot{m}_{H_2,L}^{N+1}$	Outlet mass flowrate of dissolved H_2 from reactor N	kg/s
$\dot{m}_j^{N+1}$	Outlet mass flowrate of species j from reactor N	kg/s
F_G^i	Inlet molar flow rate of gas stream to reactor i	kmol/sec
F_L^i	Inlet molar flow rate of liquid stream to reactor i	kmol/sec

3. Results and discussion

The variation of normalised cascade reactor capital cost and other reactor performance metrics is shown for the two case studies (Figure 4) for different values of ε. The hydrogenation reaction mechanism involves three reactions, with 4-IBPE being hydrogenated upon its formation to produce 4-isobutylethylbenzene (4-IBEB) and water side products. Furthermore, the condensation of 4-IBAP produces oligomers as additional side products. This explains the decrease in 4-IBPE selectivity with increasing 4-IBAP conversion. The reaction yield increases across a wide range of conversions, but this trend is reversed at higher conversions where the concentration of side products exceeds that of 4-IBPE. On the other hand, the E-factor generally decreases with conversion due to the consumption of reacting materials, but increases at higher conversions due to side product formation and the reduction in the concentration of 4-IBPE. The carbonylation mechanism also involves three reactions, with IBP being produced in the third reaction. Once again non-monotonic behaviour is observed for the E-factor as a function of conversion, whereas selectivity and conversion are now found to follow the same trends. Generating performance metric plots such as those in Figure 4 can assist pharmaceutical manufacturers in understanding the trade-off and synergies that exist for a given reaction route, and make it possible to consider alternatives across a range of criteria.

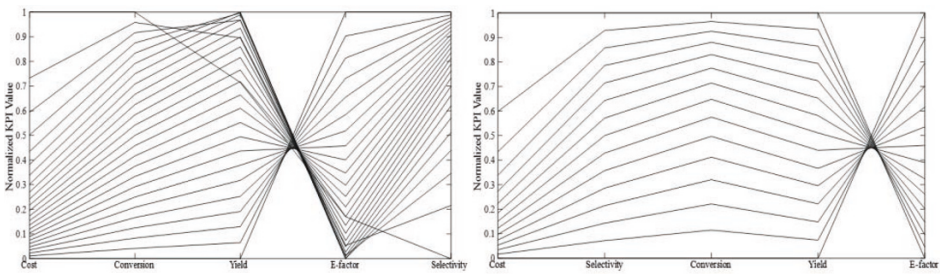


Figure 4 Trade-offs between KPIs for the hydrogenation (left) and carbonylation (right) case studies, shown as parallel coordinate plots. The two objectives are shown as the first two (leftmost) coordinates in both cases.

4. Conclusions

We have presented a set of KPIs for the assessment of process designs at the early stage, namely reactor capital cost, selectivity, E-factor, conversion, and yield. These can be used within a multi-objective optimisation framework to assess the trade-offs between different performance metrics, hence enabling more informed decisions on possible routes for pharmaceutical process design. This approach was illustrated by investigating two continuous multiphase reactors involved in Ibuprofen synthesis. The approach is currently being extended to include separation processes.

Acknowledgments

Funding from Eli Lilly and Company and the UK EPSRC, through the PharmaSEL-Prosperity Programme (EP/T005556/1), is gratefully acknowledged. AG acknowledges funding from the Royal Academy of Engineering and Eli Lilly and Company for support of a Research Chair (Grant RCSRF18193).

References

C.S. Adjiman, A. Galindo, and G. Jackson, 2014, Molecules matter: the expanding envelope of process design, Computer Aided Chemical Engineering, 34, pp. 55-64

M.R. Chapman, M.H. Kwan, G. King, K.E. Jolley, M. Hussain, S. Hussain, I.E. Salama, C. González Niño, L.A. Thompson, M.E. Bayana, A.D. Clayton, B.N. Nguyen, N.J. Turner, N. Kapur, and A.J. Blacker, 2017, Simple and versatile laboratory scale CSTR for multiphasic continuous-flow chemistry and long residence times, Organic Process Research & Development, 21(9), pp.1294-1301

J.M. Douglas, 1988, Conceptual design of chemical processes, 1110, McGraw-Hill New York

V. Elango, M. Murphy, B.L. Smith, K.G. Davenport, G.N. Mott, E.G. Zey, and G.L. Moss, 1991, Method for producing ibuprofen

Y. Y. Haimes, L. Lasdon, and D. Wismer, 1971, On a Bicriteria Formulation of the Problems of the Integrated System Identification and System Optimization, IEEE Transaction on Systems, Man, and Cybernetics, SMC-1, pp. 296-297

J.C. McWilliams, A.D. Allian, S.M. Opalka, S.A. May, M. Journet, and T.M. Braden, 2018, The evolving state of continuous processing in pharmaceutical API manufacturing: a survey of pharmaceutical companies and contract manufacturing organizations, Organic Process Research & Development, 22(9), pp.1143-1166

F. Montes, K.V. Gernaey, and G. Sin, 2017, Uncertainty and Sensitivity Analysis for an Ibuprofen Synthesis Model Based on Hoechst Path, Computer Aided Chemical Engineering, 40, pp. 163-168

C.A. Nicolaou and N. Brown, 2013, Multi-objective optimization methods in drug design, Drug Discovery Today: Technologies, 10(3), pp. 427-435

E. Papadakis, J.M. Woodley, and R. Gani, 2018, Perspective on PSE in pharmaceutical process development and innovation, Computer Aided Chemical Engineering, 41, pp. 597-656

A.M. Seayad, J. Seayad, P.L. Mills, and R.V. Chaudhari, 2003, Kinetic modeling of carbonylation of 1-(4-isobutylphenyl)ethanol using a homogeneous $PdCl_2(PPh_3)_2$/TsOH/LiCl catalyst system, Industrial & engineering chemistry research, 42(12), pp. 2496-2506

N. Thakar, R.J. Berger, F. Kapteijn, and J.A. Moulijn, 2007, Modelling kinetics and deactivation for the selective hydrogenation of an aromatic ketone over Pd/SiO_2, Chemical Engineering Science, 62(18-20), pp.5322-5329

T.K.H Trinh, J.C. de Hemptinne, R. Lugo, N. Ferrando, and J.P. Passarello, 2015, Hydrogen solubility in hydrocarbon and oxygenated organic compounds, Journal of Chemical & Engineering Data, 61(1), pp.19-34

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50361-4

Quality by design and techno-economic modelling of RNA vaccine production for pandemic-response

Zoltán Kis[a,b*], Kyungjae Tak[a], Dauda Ibrahim[a], Simon Daniel[a], Damien van de Berg[a], Maria M Papathanasiou[a], Benoît Chachuat[a], Cleo Kontoravdi[a], and Nilay Shah[a]

[a] *Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom*
[b] *Department of Chemical and Biological Engineering, The University of Sheffield, Mappin St, Sheffield S1 3JD, UK*
**Corresponding Author's E-mail: z.kis@sheffield.ac.uk*

Abstract

Vaccine production platform technologies have played a crucial role in rapidly developing and manufacturing vaccines during the COVID-19 pandemic. The role of disease agnostic platform technologies, such as the adenovirus-vectored (AVV), messenger RNA (mRNA), and the newer self-amplifying RNA (saRNA) vaccine platforms is expected to further increase in the future. Here we present modelling tools that can be used to aid the rapid development and mass-production of vaccines produced with these platform technologies. The impact of key design and operational uncertainties on the productivity and cost performance of these vaccine platforms is evaluated using techno-economic modelling and variance-based global sensitivity analysis. Furthermore, the use of the quality by digital design framework and techno-economic modelling for supporting the rapid development and improving the performance of these vaccine production technologies is also illustrated.

Keywords: techno-economic modelling; Quality by Design (QbD) modelling; process development; RNA vaccine production.

1. Introduction

Process Systems Engineering tools have a lot to offer and are not applied to their full potential in vaccine and biopharmaceutical product-process development, and during production process operation. Over the past decades, substantial progress has been made in the field of computational modelling and mechanistic, dynamic, machine learning and hybrid models have been successfully implemented in various manufacturing fields, outside of vaccine and biopharmaceutical manufacturing. These digital tools have been used to create a digital replica (or digital twins) of the manufacturing process. Vaccine and biopharmaceutical production are lagging behind in digitalisation, because vaccines and biopharmaceuticals are conventionally produced using cell-based processes that, due to their inherent complexity and variability, have been challenging to model. In addition, vaccine manufacturing is highly regulated, and improvements are not implemented rapidly to avoid the risk of negatively impacting product quality, safety, and efficacy. Modelling of complex biological systems, digitalisation, real-time monitoring, process control, automation, and knowledge-rich regulatory submissions are hindered by the lack

of real-time or near-real-time hardware sensors for measuring vaccine and biopharmaceutical quality attributes. This is because several critical quality attributes (CQAs) and parameters are difficult, time-consuming, or expensive to measure or estimate in real-time, drastically limiting the information available for developing computational models. To overcome these limitations, software sensors are being developed.

To our knowledge, the quality of vaccines is currently assured without taking advantage of digitalisation and is currently tested after every production batch. Batches that fail to yield the product quality specifications are discarded, wasting valuable resources. Quality assurance could be improved by real-time product quality monitoring and by using model-predictive control. Assuring product quality with such digital tools would fit perfectly into the QbD framework (CMC-Vaccines Working Group, 2012). The use of the QbD framework is supported by regulatory authorities for systematic co-development of the vaccine product, vaccine production processes and of the process control strategies, based on sound science and quality risk management (ICH Expert Working Group, 2009). As far as we know, a full QbD framework has not been implemented for this purpose. The QbD framework combined with digital tools is also referred to as the Quality by Digital Design (QbDD) framework and this has the potential to replace quality by testing with assuring product quality by the design and operation of the production process.

Besides QbD modelling, techno-economic modelling also offers a valuable tool for assessing the productivity and cost profile of the holistic production process (Ferreira and Petrides, 2021; Ferreira et al., 2021; Kis et al., 2021a, 2021b, 2020; Pereira Chilima et al., 2020). Moreover, this process-cost modelling approach also helps to identify the production bottlenecks, and then de-bottlenecking approaches are evaluated to increase process performance. Techno-economic modelling is also used to evaluate various scenarios, for example different downstream configurations, at different production scales to identify the process configuration that leads to maximum productivity and lowest cost (Kis et al., 2021a, 2021b, 2020). Additionally, uncertainty and sensitivity analysis is performed combined with techno-economic modelling, to identify how the co-variation of many uncertainties would impact production throughputs and resource requirements (Kis et al., 2021b).

In this work, we showcase the use of the QbDD framework together with techno-economic modelling to guide the development and operation of new vaccine production platform technologies, such as the messenger RNA (mRNA), self-amplifying RNA (saRNA) and adenoviral vectored (AVV) vaccine platforms.

2. The AVV, mRNA and saRNA vaccine production platform technologies

The AVV production process was modelled based on the manufacturing of the replication-deficient chimpanzee adenovirus-vectored (ChAdOx1) vaccine which was co-developed by Oxford University and AstraZeneca plc (Kis et al., 2021b). The ChAdOx1 production process starts with preparing the HEK293 cell seed train and the adenovirus inoculum seed train. For this, the HEK293 cells are cultured at increasing volumes until the culture amounts required for the production bioreactor (commonly at 2000 L working volume) scale are obtained. These cells are then infected with the adenovirus which was genetically modified to express the SARS-CoV-2 spike protein. Following virus replication in HEK293 cells in the bioreactor, the virus culture and cell culture enter the downstream purification, whereby cells are initially lysed then the larger

impurities are removed using microfiltration. Next, tangential flow ultrafiltration/diafiltration is carried out, followed by an ion-exchange chromatography step. After this, the adenoviral vector solution is sterile filtered, and the buffer can be exchanged for the formulation buffer, using tangential flow ultrafiltration/diafiltration. Subsequently, the adenovirus vaccine drug substance (active ingredient) is formulated and filled into vials or other containers, often at a different facility / location (Kis et al., 2021b).

The mRNA and saRNA vaccines (collectively referred to as RNA vaccines) are synthesised using the T7 RNA polymerase based on a DNA template in the *in vitro* transcription reaction, which is usually completed in 2 hours, substantially faster than AVV production (Kis et al., 2021b). Following RNA synthesis, the plasmid DNA is digested using the DNAse I endonuclease enzyme. Next, the RNA is purified out of the reaction mix using a series of conventional filtration- and chromatography-based unit operations (Kis et al., 2021a, 2020). These can include tangential flow ultrafiltration and diafiltration combined with one or two of the following chromatography techniques: ion-exchange, reverse-phase, oligo dT affinity, hydroxyapatite, hydrophobic interaction, multimodal hydrogen bonding and anion exchange, cellulose-based, and multimodal core-beads. After the RNA is purified out of the enzymatic reaction mix, the RNA is encapsulated in lipid nanoparticles. For this, the four lipid components contained in an ethanol stream are mixed with the RNA contained in an aqueous stream (e.g. in citric acid buffer). The mixing of the lipids with RNA can be achieved using a mixing device based on: microfluidics, T-junction, impingement jet, vortex, or pressurised stainless-steel tanks. Following formulation, the solution is sterile filtered and shipped to the fill-to-finish site for filling into glass vials or other containers (Kis et al., 2021b).

3. Techno-economic modelling of the AVV, mRNA and saRNA platforms

Rapid and global response to pandemics by mass vaccination is currently limited by the rate at which vaccine doses can be manufactured on a global scale. Here techno-economic modelling is presented for the AVV, mRNA and saRNA vaccine production platform technologies that were deployed during the COVID-19 pandemic. Unlike AVV and mRNA vaccines, several of which were approved by the regulatory authorities, the saRNA platform is not yet deployed at commercial scale for vaccines, with saRNA vaccines still undergoing clinical development. Herein, a combination of techno-economic modeling and variance-based global sensitivity analysis (GSA) is applied. This quantifies the performance of each platform in terms of their productivity and resource requirements, subject to key design and operational uncertainties, cf. Figure 1. GSA was carried out by interfacing SobolGSA with SuperPro Designer via MatLab and Excel Visual Basic for Applications (VBA). For GSA, 10,000 simulations were performed for each of the three platform technologies. For these simulations, model inputs were quasi-randomly sampled from a seven dimensional input space using Sobol sequences, as previously described (Kis et al., 2021b). These seven model inputs are: the scale of the production process, batch failure rate, titre/yield in the production bioreactor, cost of raw materials, cost of labour, drug substance amount per dose and cost of quality control (Kis et al., 2021b).

Cost and productivity results from the techno-economic modelling and GSA are shown below in Figure 2. The ranges and probability distributions of the number of drug substances and finished drug product doses that can be produced based on a one billion USD investment in operating expenses (OpEx) for the three platform technologies are

shown in the violin plots in Figure 2A. A one billion USD investment in OpEx will produce a median of 2.66 (IQR=2.44-2.83) billion AVV drug product doses, a median of 0.95 (IQR=0.74-123) billion mRNA drug product doses and a median of 2.48 (IQR=2.36-2.58) billion saRNA drug product doses. OpEx includes the annualised capital costs, however it is worth noting that investment in facilities must be made upfront, because constructing, equipping, validating and starting up production can take several years. The ranges and probability distribution of the cost per dose for the drug substance and finished drug product for the three platform technologies is shown in Figure 2B. The drug product manufacturing cost per dose is 0.38 (IQR=0.35-0.41), 1.05 (IQR=0.81-1.35), 0.4 (IQR=0.38-0.42) for AVV, mRNA and saRNA vaccines, respectively.

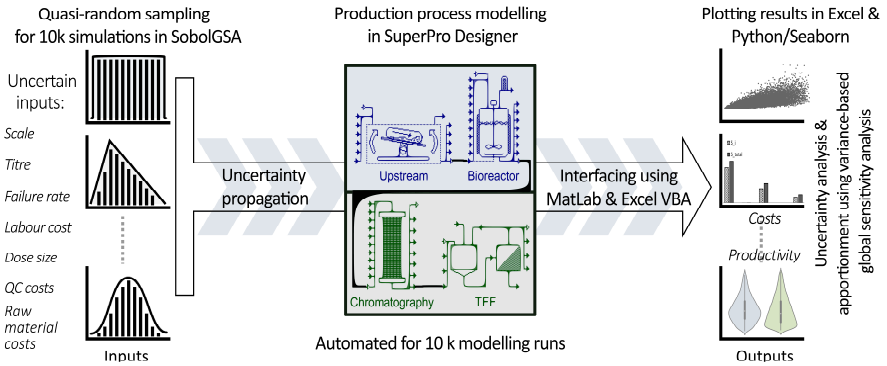


Figure 1. A computational framework for uncertainty quantification for AVV, mRNA and saRNA production. The uncertainty is propagated from the inputs via the model to the outputs. In addition, the sensitivity of the model output key performance indicators (KPIs) is attributed to the individual inputs, to determine the degree to which individual inputs impact the output KPIs. Modified from (Kis et al., 2021b).

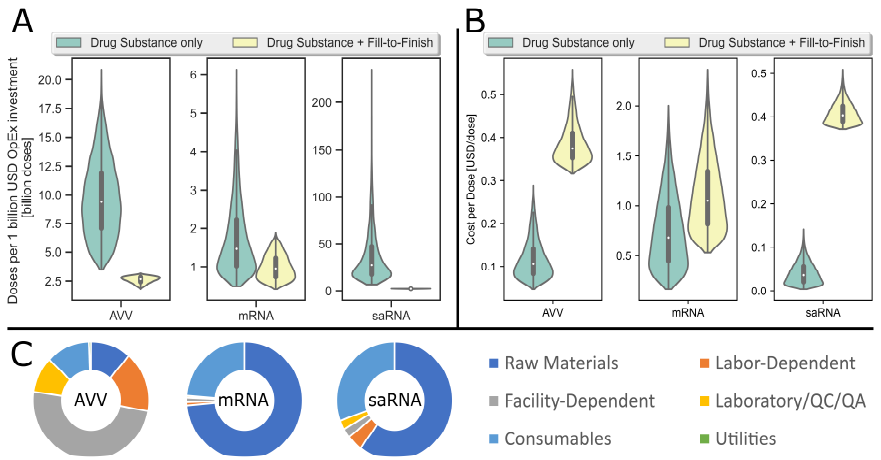


Figure 2. Cost distributions associated with AVV, mRNA and saRNA vaccine production. **A.** Violin plots showing the distribution of the estimated number of doses produced based on a 1 billion USD investment in operating expenses (OpEx). The OpEx contains the annualised facility costs. **B.** Violin plots showing the distribution of cost per dose values for AVV, mRNA and saRNA vaccine production. **C.** Doughnut charts showing the distribution of OpEx, the annualised capital costs are included in the facility-dependent costs.

In the centre of all the violin plots, box and whisker plots are shown with the median values indicated by the white dots; the 25th and 75th percentiles with the top and bottom of the boxes; and minimum and maximum values, excluding outliers, with the ends of the whiskers. The width of the violin plots represents the probability distributions. Figure 2C shows the breakdown of the annual production costs for the baseline scenarios (c.f. (Kis et al., 2021b)) for these three vaccine platform production technologies. Fixed costs dominate the AVV production costs, whereas mRNA and saRNA vaccines production is driven by variable costs. This implies that maintaining surge capacity based on the RNA platform will be more cost effective than based on the AVV platform. Fill-and-finish was modelled with 10-dose vials for AVV vaccines and 5-dose vials for mRNA and saRNA vaccines.

4. Integration of QbD and techno-economic modeling with the RNA platform

The mRNA vaccine production platform technology has been proven clinically successful during the COVID-19 pandemic. The cell-free nature and consequently the relative simplicity (compared to cell-based vaccine production) makes the RNA platform technology ideal for digitalisation and advanced automation with the QbDD framework. The integration of the QbDD framework with the RNA platform will accelerate product-process development, enhance production rates and production volumes, reduce costs, and assure high product quality. Moreover, the RNA vaccine production platform and the QbDD framework will form a powerful synergy as both tools are disease agnostic. This synergy will use prior platform knowledge, experimental data, clinical data, quality risk management and digital tools to accelerate product and process development. This will also accelerate and streamline the regulatory approval process based on knowledge-rich regulatory submissions and demonstrated product knowledge and process understanding.

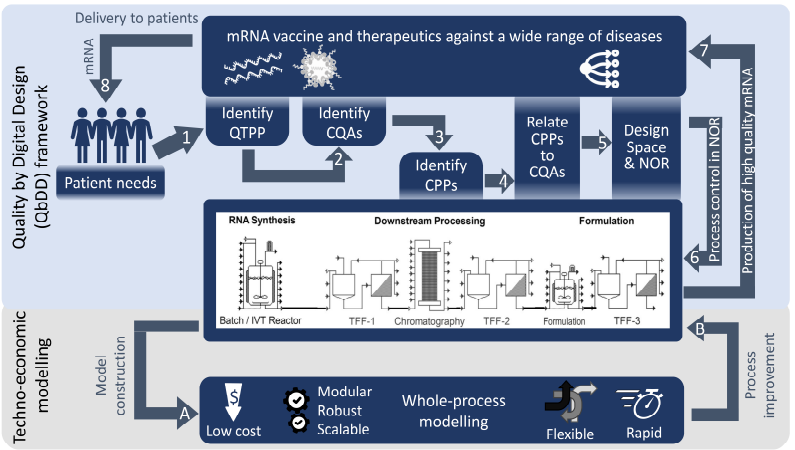


Figure 3. Integration of techno-economic and QbDD modelling with the RNA vaccine production process. Abbreviations: QbDD - Quality by Digital Design, QTPP - Quality Target Product Profile, CQAs – product Critical Quality Attributes for safety and efficacy, CPP - Critical Process Parameters, NOR - Normal Operating Range, within the design space.

The product specifications will be based on product performance instead of batch history, and the focus will shift from reproducibility to process-product robustness (Kis et al., 2020; van de Berg et al., 2021). In addition, the QbDD framework and the digital tools

will be used to automate RNA vaccine manufacturing, building quality into the design and operation of the process. These features of the QbDD framework will also support scale-up and technology transfer. In addition, techno-economic modelling will guide cost-reduction, de-bottlenecking and improved process performance. The interplay between the RNA vaccine platform technology, the QbDD framework and techno-economic modelling is shown above in Figure 3.

5. Conclusions

In conclusion, on top of having surge vaccine manufacturing capacity available for future outbreaks, modelling tools such as those presented here can further accelerate vaccine development and improve the performance of the vaccine manufacturing processes. The combination of vaccine platform technologies and these disease-agnostic modelling tools provides a powerful approach for rapid-response vaccine deployment against currently known and unknown diseases.

Acknowledgements: This research was funded by the Department of Health and Social Care using UK Aid funding and is managed by the Engineering and Physical Sciences Research Council (EPSRC, grant number: EP/R013764/1). The views expressed in this publication are those of the author(s) and not necessarily those of the Department of Health and Social Care. Funding from UK Research and Innovation (UKRI) via EPSRC grant number EP/V01479X/1 on COVID-19/SARS-CoV-2 vaccine manufacturing and supply chain optimisation is thankfully acknowledged.

References

CMC-Vaccines Working Group, 2012. A-Vax: applying quality by design to vaccines.

Ferreira, R., Petrides, D., 2021. Messenger RNA (mRNA) Vaccine Large Scale Manufacturing – Process Modeling and Techno-Economic Assessment (TEA) using SuperPro Designer. https://doi.org/10.13140/RG.2.2.12780.28800

Ferreira, R.G., Gordon, N.F., Stock, R., Petrides, D., 2021. Adenoviral Vector COVID-19 Vaccines: Process and Cost Analysis. Processes. https://doi.org/10.3390/pr9081430

ICH Expert Working Group, 2009. ICH harmonised tripartite guideline on pharmaceutical development Q8 (R2) (No. August), ICH guideline Q8 (R2) on pharmaceutical development. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH), Geneva, Switzerland.

Kis, Z., Kontoravdi, C., Dey, A.K., Shattock, R., Shah, N., 2020. Rapid development and deployment of high-volume vaccines for pandemic response. J. Adv. Manuf. Process. 2, e10060. https://doi.org/10.1002/amp2.10060

Kis, Z., Kontoravdi, C., Shattock, R., Shah, N., 2021a. Resources, Production Scales and Time Required for Producing RNA Vaccines for the Global Pandemic Demand. Vaccines. https://doi.org/10.3390/vaccines9010003

Kis, Z., Tak, K., Ibrahim, D., Papathanasiou, M.M., Chachuat, B., Shah, N., Kontoravdi, C., 2021b. Pandemic-response adenoviral vector and RNA vaccine manufacturing. medRxiv. https://doi.org/10.1101/2021.08.20.21262370

Pereira Chilima, T.D., Moncaubeig, F., Farid, S.S., 2020. Estimating capital investment and facility footprint in cell therapy facilities. Biochem. Eng. J. 155, 107439. https://doi.org/10.1016/j.bej.2019.107439

van de Berg, D., Kis, Z., Behmer, C., Samnuan, K., Blakney, A., Kontoravdi, C., Shattock, R., Shah, N., 2021. Quality by Design modelling to support rapid RNA vaccine production against emerging infectious diseases. NPJ Vaccines 6, 1–10. https://doi.org/10.1038/s41541-021-00322-7

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50362-6

Design of Value Function Trajectory for State of Control in Continuous Manufacturing System

Tomoyuki TAGUCHI [a*], Toshiyuki WATANABE [b*], Shigeru KADO [b*] and Yoshiyuki YAMASHITA [c]

[a]*Chiyoda Corporation, Pharmaceutical Engineering Department, 4-6-2, Minatomirai, Yokohama, Kanagawa 220-8765, JAPAN*
[b]*Chiyoda Corporation, Research and Development Center, 3-13, Moriya-cho, Yokohama, Kanagawa 221-0022, JAPAN*
[c] *Department of Chemical Engineering, Tokyo University of Agriculture and Technology, Tokyo 184-8588, JAPAN*
taguchi.tomoyuki@chiyodacorp.com

Abstract

Recently, pharmaceutical manufacturing has been aimed at incorporating more efficient production systems for easy scale-up, higher quality and lesser usage of solvent, and observation of the operation state. A continuous manufacturing system can enable a system design that maintains the desired process conditions, with suitable devices and measurements. The process systems engineering (PSE) approach is very helpful for the visualization of the operation state, and the compensation of system features of process dynamics and disturbances. The investigated design space provides valuable information regarding the process and control strategy. The combination of steady-state data sets among control, process, and objective variables can visualize the range of the allowable operation zone, and desired objective control state. The proposed value function trajectory is an attractive method for designing tracking control as a multivariable function of the Hamiltonian, which connects the consistent approach of the design of process equipment, structure, and quality control.

Keywords: Process Design and Control, Continuous, Design Space, Optimization.

1. Introduction

Lee (2015) stated the kinds of features of continuous manufacturing, such as small footprints, short supply chain, stable operation with good monitoring, combination of processes with no-stop handling, easy scale-up, new synthetic routes, safe operation, and efficient high-throughput production. The understanding of the process features in the early design stage is a way to achieve agility, flexibility, and robustness of continuous manufacturing. Myerson (2014) listed the needs of technologies for innovating manufacturing systems: comprehension of process steady-state and dynamics, design of monitoring and control systems, development of systems integration, and data analysis to understand the design space, considering disturbances, nonlinearities, constraints, and uncertainties. It was an effective approach for reducing various operational risks. Diab (2020) described examples of visualization for certain active pharmaceutical ingredients and unit operations, under the consideration of structuring a continuous process. This method was highly attractive for both process design and operation philosophy.

To realize a continuous pharmaceutical manufacturing system, the comprehension of material and reaction properties, a priori knowledge of transport phenomena, and the innovation of devices and instruments must be closely integrated based on proven experimental data. Such basic data should be transferred to the control design and operational support through system design.

In this study, a novel design concept of using the steady-state calculation was proposed. To understand process characteristics, the design space was determined according to the process design and operational philosophy. In this process, the objective control state and possible operational zone could be defined by introducing a value function as the potential energy of the system characteristics. Through the implementation of this approach, The Hamiltonian expression combining the potential energy as the value function, and the kinetic energy as the process dynamics, could be derived from the case study on steady-state calculation, which consequently resulted in the structuring of a feedback control system.

2. Process control with a priori knowledge

The issues of control and process optimization are strongly influenced by the chemistry, process, control, and operation. Plutschack et al. (2018) illustrated the types of heat and mass transport measurements, and the numerous examples of instruments related to various chemistries. The comprehension of system features in the wide range of materials and process attributes was achieved, a process design was conducted using mathematical modelling and simulation. Larsson (2000) reviewed the plantwide control that imposed structural design: selection of controlled variables, manipulated variables, measurements, control configuration, and type of controller were the issues of determination. They strongly depended on the process structure, followed by the degrees of freedom; therefore, the focus was on the process-oriented approach. Shalifzadeh (2013) summarized the integration approach for process design and control. The traditional approach for process design and control is sequential, that is, the control design is only conducted for rigid manufacturing systems. The decision-making of system design between process and control, and the competition between efficiency and controllability are the main issues from the perspective of the designer and planner. The proposed method of trajectory tracking control was aimed at combining the acquisition of process a priori knowledge, and the design of control structure, like the simultaneous approach. The objective control state, and surrounding allowable operation zone were achieved as the dimensional reduction information through the constitution of the design space. To incorporate such a priori knowledge into a feedback control system, a data bank that included the value function and process state relationship, was implemented by the Hamiltonian. They were extracted from the steady-state simulation model for using a process model in an environment of low computational cost, and an easy-to-handle mathematical model.

3. Reactor Simulation Model

The reaction process plays a major role in the pharmaceutical and chemical industries. Over-reaction and by-product generation are the conventional problems of organic synthesis reaction. The following first-principle model is provided to illustrate the proposed approach:

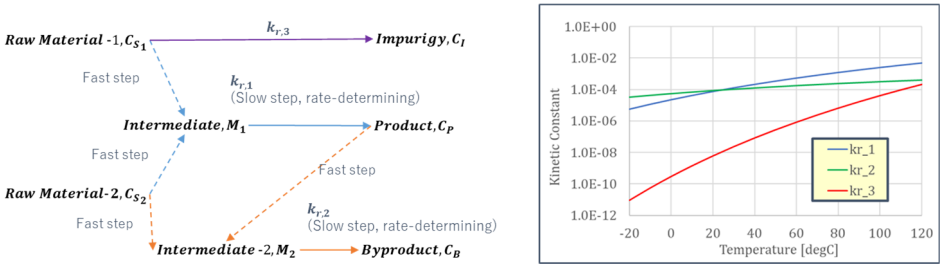


Figure 1: Synthesis route and temperature dependence of the three kinetic constants

3.1. Constraints of the reaction system

1) Reaction paths
The following three reaction paths were provided to consider the typical non-linearity of the organic synthesis route, as shown in Figure 1.
   A) The second-order reaction of two raw materials for generating the main product.
   $$r_P = k_{r_1} C_{S_1} C_{S_2} \quad (1)$$
   B) The second-order reaction of the main product and one of the raw materials for generating the by-product of over-reaction.
   $$r_B = k_{r_2} C_{S_2} C_P \quad (2)$$
   C) The first-order reaction of one of the raw materials for generating the impurity.
   $$r_I = k_{r_3} C_{S_1} \quad (3)$$
2) Kinetic constants
Three kinetic constants whose characteristics were dependent on the temperature change, are shown in Figure 1.

3.2. Reactor configuration, size and flow condition

1) Reactor configuration
A typical plug flow reactor, with a jacket to handle the liquid-phase reactants, was considered as the test bed for this study (Figure 2). The temperature of the jacket colorant was assumed to be 40 °C, and UA = 200 W/m^2-K.
2) Size
Internal diameter = 30 mm, Length = 500 mm were assumed.

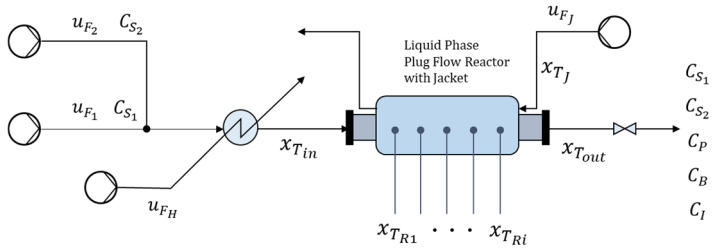


Figure 2: Typical plug flow reactor with inlet heat exchanger and colorant jacket for temperature control

Table 1: Study range of volume flow, molar concentration and temperature

Volume Flow Rate [mL/min.]		Molar Concentration [mol/m^3]		Temp. [°C]
u_{F_1}	u_{F_2}	C_{S_1}	C_{S_2}	$x_{T_{in}}$
50~500	50~500	100~500	100~500	20~100

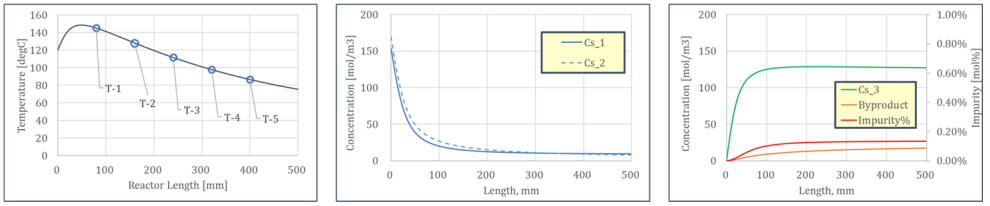


Figure 3: Typical trends of reactants and products

3) Flow balance
The following ranges (Table 1) of volume flow, molar concentration, and temperature were assumed.

3.3. Typical trends of reactants and products

Figure 3 shows an example of the simulation results. The inlet flow condition and temperature set-point strongly affected the resultant by-product and impurities.

4. Design Space

4.1. Determination of value function

To consider the non-linear constrains of reactants, the following value function, V was employed to design the objective control state. Therein, ψ_i is the weight factor, ξ_c is the conversion of substrates, ξ_y is the product yield, ξ_i is the impurity in mol%, with an upper limit of 0.01 mol%, and z_p is the assumed target production rate, in kg/h.

$$V = \psi_1(\xi_c - 1.0)^2 + \psi_2(\xi_y - 1.0)^2 + \psi_3\xi_i + \psi_4(z_p - 2.0)^2 \quad (4)$$

In case, ξ_i is higher than the assumed upper limit of 0.01 mol%, ψ_3 becomes zero. Then, the optimized value function becomes zero under the corresponding control variables, $u_{F_{opt}}$, and state variables, $x_{T_{opt}}$.

4.2. Comprehension of wide-range process attributes

The overall comprehension of the possible process variable changes during plant operation was investigated. By changing the control variables listed in Table 1, followed by the state variables, the corresponding reactant results could be achieved through the simulation case studies. Figure 4 shows the results of the value function with respect to residence time, and inlet temperature. Quick conversion, and less generation of the by-product and impurity occurred under the high-temperature feed condition, and subsequently, the value function was low.

4.3. Robust design space against process disturbances

The design space was further investigated based on the comprehension of the overall process features. Figure 5 shows the enlarged versions of the candidate zones (1), and (2), shown in Figure 4. It was found that the candidate zone (2) had a wider zone of low-value function. If a disturbance regarding two process variables occurs, candidate zone (2) would be expected in the robust control state.

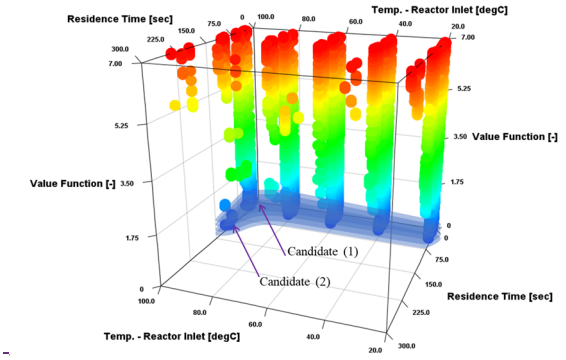


Figure 4: Wide-range three-dimensional plot of the value function

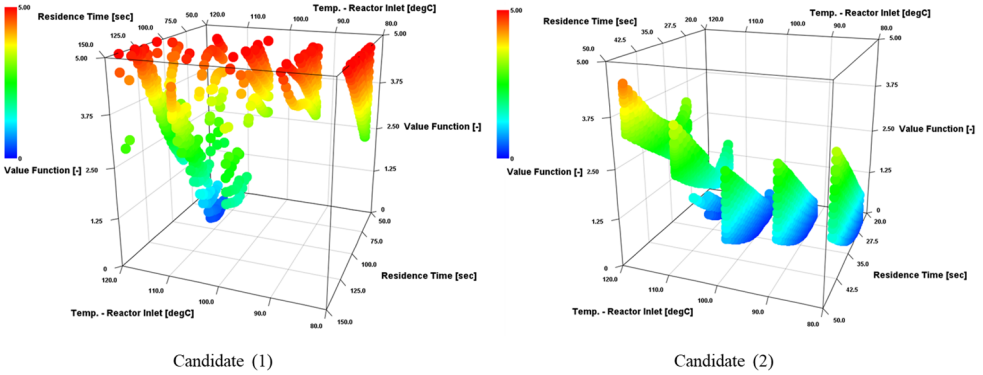


Figure 5: Narrow-range three-dimensional plot of the value function

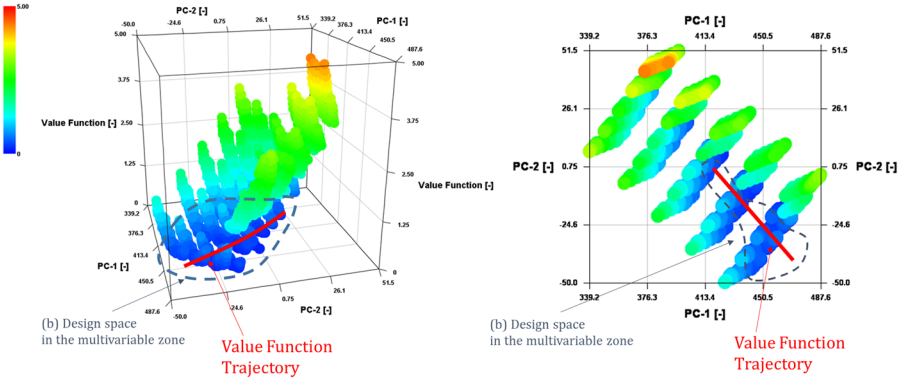


Figure 6: Trajectory of the value function and the allowable operation state

5. Trajectory Tracking Control

5.1. Determination of value function trajectory

The value function was affected not only by the residence time and inlet temperature, but also the feed flow and concentration change. To consider such a multivariable effect on the value function, the x-y axes of the candidate zone (2) in Figure 5 were transformed to the principle components, PC-1 and PC-2; and the results are shown in Figure 6.

The operation range of low-value function is clearly expressed in Figure 6, with the group of minimum values as optimized line in the design space, and the allowable operation range could be designed as the multivariable zone around the value function trajectory.

5.2. Proposed control

The trajectory and the surrounding allowable zone were considered as the potential function, and its features were quadratic. Ortega (2002) utilized passivity-based control to design robust controllers for the Euler-Lagrange equations of motion. This consideration were applied to the Hamiltonian function, described by the following equation:

$$H = V° + V + F \tag{5}$$

where, V° is the value function at the optimal condition, V is the relationship between the value function, control, and process state variables, and F is the consideration of the dynamic state equation. The stepwise change, and sensitivity of the data set among the control, process state, and value function, could be translated to each coefficient matrix. It showed that the consequence of work flow to achieve the design space could be utilized for the effective design of multivariable control structures.

6. Conclusions

Using the reactor simulation model, the design space was considered, and the trajectory of the value function and the allowable operation zone were determined. The designed control space with the trajectory of the value function could be transformed into a multivariable feedback controller based on the Hamiltonian concept. To design a control system incorporating experimental and a priori knowledge of engineering, the proposed design, that is, a design-directed approach, is useful for combining the good corroborations of the process and the control design. The different functions of the cross-sectional departments, throughout the stages of research and development of commercial operation, could be interactively connected by utilizing fundamental knowledge of mathematical modelling.

References

S. Diab and D. I. Gerogiorgis, 2020, Design Space Identification and Visualization for Continuous Pharmaceutical Manufacturing, Pharmaceutics, 12, 3, 235.

T. Larsson and S. Skogestad, 2000, Plantwide control—A review and a new design procedure, Modeling, Identification and Control, 21, 4, 209-240.

S. L. Lee et al., Modernizing Pharmaceutical Manufacturing: from Batch to Continuous Production, 2015, Journal of Pharmaceutical Innovation, 2015, 10, 3, 191-199.

A. S. Myerson et al., Control systems engineering in continuous pharmaceutical manufacturing May 20–21, 2014 continuous manufacturing symposium, Journal of pharmaceutical sciences, 2015, 104, 3, 832-839.

R. Ortega et al., Interconnection and damping assignment passivity-based control of port-controlled Hamiltonian systems, Automatica, 2002, 38, 4, 585-596.

M. B. Plutschack et al., The Hitchhiker's Guide to Flow Chemistry, Chemical reviews, 2017, 117, 18, 11796-11893.

M. Sharifzadeh, Integration of process design and control: A review, Chemical Engineering Research and Design, 2013, 91, 12, 2515-2549.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50363-8

A Thermodynamic Approach for Simultaneous Solvent, Coformer, and Process Optimization of Continuous Cocrystallization Processes

Nethrue Pramuditha Mendis, Richard Lakerveld*

Department of Chemical and Biological Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China
r.lakerveld@ust.hk

Abstract

Model-based optimization of cocrystallization processes involves the simultaneous identification of the optimal coformer and solvent types and the process operating conditions, which should suppress the formation of undesirable solid-state forms. New methods are needed for such optimization tasks. This work presents a computational framework for the optimal selection of coformers, solvents, and operating conditions for a cocrystallization process of a drug with low aqueous solubility. The method considers a cocrystal product that meets a specified target for solubility enhancement, which enhances product functionality. The proposed framework is demonstrated for the model drug carbamazepine. An optimization problem is formulated and solved with the proposed strategy, which illustrates its effectiveness for the optimization of cocrystallization processes with constraints on the pharmaceutical product performance.

Keywords: Cocrystallization, solvent selection, integrated product and process design, process optimization, PC-SAFT.

1. Introduction

Many active pharmaceutical ingredients (APIs) are orally administered as crystalline solids due to the various advantages they possess over other types of dosage forms. However, a low aqueous solubility can seriously limit the bioavailability of such an API. Pharmaceutical cocrystals (CCs) are solid-state forms that can improve the dissolution rate of APIs. They are formed when a neutrally charged API is crystallized with a neutrally charged coformer (CF). Cocrystals have a well-defined API:CF stoichiometric ratio in the crystal lattice and they can quickly dissolve into their molecular forms due to the non-covalent bonds.

Solution crystallization is a common strategy for industrial cocrystal synthesis (Lange & Sadowski, 2015). Solvents typically play a crucial role in any crystallization process. Thus, the solvent selection is a key decision. In general, systematic approaches for integrated solvent and process optimization have been reported for crystallization-based processes, for example, for the case of anti-solvent crystallization processes with recycles (Wang & Lakerveld, 2018). These problems are challenged by the strong interdependence between solvent selection and process operating conditions. Optimization of cocrystallization processes is uniquely challenging because the CF type plays a pivotal role in dictating the process and product performance. As cocrystallizing systems have the potential to form multiple solid-state forms, i.e., at least the pure API, pure CF, and the desired cocrystal, the choice of CFs, solvents, and operating conditions

should facilitate the formation of the desired cocrystal, while suppressing the formation of other solid forms to obtain a product with a high purity. Furthermore, in the case of a pharmaceutical cocrystal, the degree to which the aqueous solubility may be enhanced for a given API is determined by the CF type. Therefore, the CF selection may not only affect the process performance but also the product efficacy. Although substantial work has been reported on related topics such as cocrystal discovery (e.g., (ter Horst et al., 2009)), optimization approaches for cocrystallization processes involving simultaneous CF, solvent, and operating condition identification have not been reported.

The objective of this work is to develop a thermodynamics-based optimization framework to select CFs, solvents, and operating conditions for cocrystallization processes to minimize the process operating costs and maximize the solubility enhancement of the final cocrystal form. Our optimization framework is based on the perturbed-chain statistical associating fluid theory (PC-SAFT) (Gross & Sadowski, 2001), which can model cocrystallizing mixtures well (Lange & Sadowski, 2015). Carbamazepine (CBZ) is selected as the model API along with eight of its CFs (glutaric acid, nicotinamide, saccharin, salicylic acid, oxalic acid, malonic acid, succinic acid, 4-aminobenzoic acid). CBZ is a solubility/dissolution rate-limited API.

2. Approach

2.1. Process Model Development

The model of the process (see Figure 1) consists of material balances (omitted here for brevity) and equilibrium relations with activity coefficients obtained from the PC-SAFT. The PC-SAFT provides a means of calculating the residual Helmholtz free energy of a mixture based on hard-chain, dispersion, and association interactions, which can serve as the basis to determine other thermodynamic properties like activity coefficients (Gross & Sadowski, 2001, 2002). In this work, all the compounds involved in the process, i.e., the API, the CF, and the solvent, are characterized by five parameters as in Gross & Sadowski, 2002: segment number (m), segment diameter (σ), dispersion energy parameter (ε), association energy (ε^{AB}), and association volume (κ^{AB}). The list of 48 solvents and their PC-SAFT pure component parameters considered in this work are obtained from our previous work (Wang & Lakerveld, 2018). Pure component parameters of the API and the eight CFs are estimated from solubility data reported in the literature, which is a standard procedure for estimating PC-SAFT parameters of solid compounds and omitted here for brevity.

The solubility of a pure compound is calculated according to:

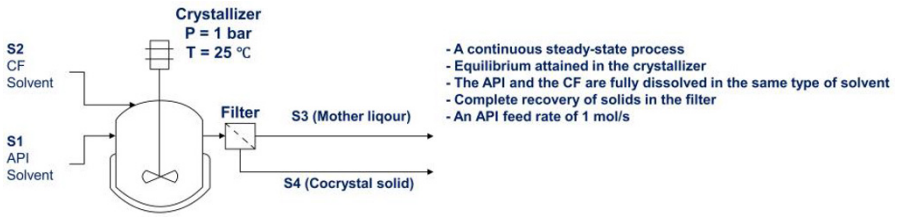


Figure 1: The proposed single-stage crystallization process configuration. The CF type, CF feed flow rate, solvent type, and the solvent flow rates in stream S1 and stream S2 are the free variables to be optimized.

$$x^{sat}\gamma^{sat} = \exp\left[\frac{\Delta H_f^m}{R}\left(\frac{1}{T_m} - \frac{1}{T}\right)\right], \tag{1}$$

where x^{sat}, γ^{sat}, ΔH_f^m, and T_m are the mole fraction, activity coefficient, enthalpy of fusion, and the melting temperature, respectively. R and T are the universal gas constant and the temperature, respectively.

The following inequality constraints are applied to ensure that both the API and the CF are fed to the process in fully dissolved form and do not crystallize in their pure forms:

$$x_{API,S1} \le x_{API}^{sat},\ x_{CF,S2} \le x_{CF}^{sat},\ x_{API,S3} \le x_{API}^{sat},\ x_{CF,S3} \le x_{CF}^{sat} \tag{2}$$

The solubility of a cocrystal is expressed as a solubility product (Good & Rodríguez-Hornedo, 2009) as follows:

$$\left(x_{API,S3}\gamma_{API,S3}\right)^a \left(x_{CF,S3}\gamma_{CF,S3}\right)^b = K_{sp} \tag{3}$$

where a and b are the stoichiometric coefficients of the API and the CF in the cocrystal lattice, and K_{sp} is the solubility product.

2.2. Cocrystal Solubility Advantage

The potential solubility enhancement of cocrystals is often quantified by the cocrystal solubility advantage (SA), which can be defined as the ratio of the maximum API concentration due to cocrystal dissolution over the solubility of the pure API (Good & Rodríguez-Hornedo, 2009). The SA for a given API is primarily determined by the CF type. The pH of the dissolution medium is critical when calculating the SA for ionizable API/CFs. In this work, a pH of 2.5 is adopted for SA calculations, which represents the stomach pH. The extents of ionization of the CFs are determined using their k_a values reported in the literature. The ionization of the API (CBZ) is neglected (Good & Rodríguez-Hornedo, 2009).

The following expressions are aided to calculate the SA for a monoprotic weak acid-CF (*HA)*,

$$x_{tot,CF} = x_{HA} + x_{A^-},\ bx_{API} = ax_{tot,CF},\ k_a = \frac{[A^-][H_3O^+]}{[HA]},\ \left(x_{API}\gamma_{API}\right)^a \left(x_{HA}\gamma_{HA}\right)^b = K_{sp} \tag{4}$$

Note that only the nonionized portion of the CF contributes to the solubility product. In case the CF is a base, a similar approach is followed with the k_a value of the conjugate acid.

2.3. Optimization

In the optimization problem, CF type is characterized by the 10-dimensional vector $\mathbf{p}_{CF}$ whose elements include the five PC-SAFT pure component parameters, and additionally ΔH_f^m, T_m, and k_a of the CF, *b/a* ratio, and K_{sp} of the cocrystal. Solvent type is characterized by the 5-dimensional vector $\mathbf{p}_{sol}$, which consists of the five PC-SAFT

pure component parameters of the solvent. The CF and solvent selection is formulated as:

$$\mathbf{p}_{CF} = \sum_i y_{CF,i}\mathbf{p}_{CF,i},\ \sum_i y_{CF,i} = 1,\ i \in \{1,2,3,...,n_{CF}\},\ y_{CF,i} \in \{0,1\} \tag{5}$$

$$\mathbf{p}_{sol} = \sum_j y_{sol,j}\mathbf{p}_{sol,j},\ \sum_j y_{sol,j} = 1,\ j \in \{1,2,3,...,n_{sol}\},\ y_{sol,j} \in \{0,1\} \tag{6}$$

where i and j are the existing CF and solvent candidates, whose choice is represented by binary variables $y_{CF,i}$ and $y_{sol,j}$. n_{CF} and n_{sol} are the total numbers of CF and solvent candidates, respectively. The above formulation ensures only one CF and solvent candidate each is selected. The normalized production cost (NPC) and the SA at pH 2.5 represent the process efficiency and the product performance, respectively, where the former needs to be minimized and the latter to be maximized. We formulate a multiobjective optimization problem as follows:

$$\min_{z_p, z_{sol}, z_{CF}} J_1 = \frac{c_{sol}\left[F_{S1,sol}\left(z_p, z_{sol}, z_{cf}\right) + F_{S2,sol}\left(z_p, z_{sol}, z_{cf}\right)\right]}{F_{S4,API}\left(z_p, z_{sol}, z_{cf}\right)},\ \max_{z_{CF}} J_2 = SA(z_{CF}) \qquad \text{(P1)}$$

s.t., process and thermodynamic models, inequality constraints in Eq.(2), Eq.(4), Eq.(5), Eq.(6)

z_p, z_{sol}, and z_{CF} stand for the operating conditions (CF feed flow rate and the solvent flow rates in stream S1 and S2), solvent type, and CF type, respectively. $F_{S1,sol}$ and $F_{S2,sol}$ are the solvent flow rates in streams S1 and S2, respectively. $F_{S4,API}$ is the API flow rate in stream S4 (product stream). The solvent cost is accounted for by a solvent cost parameter (c_{sol}) of 7.94 USD/L, which is estimated from data from Lab Alley (www.laballey.com). The CF cost is neglected as it is consumed in much smaller quantities compared to the solvent, and CFs are usually inexpensive chemicals. The optimization problem is solved by employing the epsilon-constraint method, where the lower bound for the SA (J_2) is increased from 0 to 40 to generate cases with different product performances, and the NPC (J_1) is minimized for each case.

The resulting optimization problem is an MINLP due to the discrete nature of CF/solvent selection and continuous operating conditions, which demands a relaxation strategy. The continuous mapping method (Bardow et al., 2010) is a commonly used MINLP relaxation strategy for problems involving simultaneous optimization of solvent types and operating conditions supported by the PC-SAFT. However, the optimization problem in this work is complicated by the simultaneous optimization of the CF type, which involves ten CF-related variables. A continuous mapping method may not be the most efficient solution strategy when the number of relaxation variables is large compared to the number of candidate compounds in the database. An approach similar to the traditional branch-and-bound approach is likely more efficient for such a case. Therefore, we propose a hybrid algorithm comprising the continuous mapping method for optimization of the solvent type and operating conditions and a branch-and-bound-like strategy for the CF selection. The algorithm involves two steps: 1) the optimization problem P1 is solved as an NLP by relaxing all integer variables, i.e., $y_{CF,i}$ in Eq.(5) and $y_{sol,j}$ in Eq.(6), to continuous variables, 2) P1 is solved separately for a selected set of CF candidates identified from step 1).

The solution for the relaxed problem in step 1 may not correspond to an existing CF candidate, which would be reflected by multiple nonzero values for the relaxed $y_{CF,i}$. In step 2, the optimization problem P1 is solved separately for each CF candidate i for which a nonzero $y_{CF,i}$ was found from the relaxed problem solution in step 1. As the CF type is fixed in P1 now, the solvent type and operating conditions can be optimized with an existing continuous mapping method from the literature (Wang & Lakerveld, 2018). Multiple integer solutions, i.e., solutions with an existing CF and solvent candidate, are obtained at the end of step 2, from which the optimal integer solution is identified. To verify that the integer solution obtained is the true optimal, step 1 is repeated, but without those CF candidates that have already been identified previously, i.e., n_{CF} is smaller now. If the objective function value of the new relaxed problem has not improved compared to the already identified optimal integer solution, the algorithm stops, and the identified optimal integer solution is the final solution. Otherwise, step 2 is repeated to identify more integer solutions, the optimal integer solution is updated, and step 1 is repeated again after excluding all the CF candidates already been identified. The algorithm stops when the relaxed problem objective function value does not improve compared to the already identified optimal integer solution. To solve the resulting NLPs, the CONOPT solver (Drud, 1985) was implemented in The General Algebraic Modeling System (GAMS Development Corporation).

3. Results and Discussion

When no lower bound is set for the SA, the optimal CF-solvent pair is oxalic acid (OA) and benzyl alcohol, which allows for the lowest NPC (Integer Solution 1 in Figure 2). Both the relaxed and integer solutions remain unchanged as the lower bound on the SA is increased from 0 to 6.78, which is consistent with the SA=6.78 for CBZ-OA cocrystal under the given conditions. When the lower bound of the SA is increased beyond 6.78, the NPC of the relaxed solution gradually increases. All these relaxed solutions correspond to the same integer solution (Integer Solution 2 in Figure 2), where nicotinamide (NA) and benzyl alcohol are the optimal CF and the optimal solvent type, respectively. Even though CBZ-NA cocrystal has a substantially higher SA of 41.72, the increase in NPC is substantial for the integer solution, likely because only eight CFs are considered in this work, i.e., it appears no CFs with intermediate properties are available. The optimal solvent type and required quantity for dissolving the API are the same for both optimal solutions. However, the higher total solvent consumption and the lower yield from Solution 2 cause an increase in the NPC. Finally, the optimal API:CF ratios are 1:1.46 and 1:3.16 for Solution 1 and Solution 2, respectively. This result shows that the determination of the optimal ratio at which the API and the CF have to be mixed to minimize the NPC is not trivial as it differs from the API:CF stoichiometric ratio (1:0.5 for CBZ-OA and 1:1 for CBZ-NA).

4. Conclusions

The proposed modeling and optimization framework for cocrystallization processes can simultaneously identify optimal CFs, solvent types, and operating conditions while considering both process design and product performance. The method can be used to balance the complex trade-offs that often exist in crystallization-based processes for the

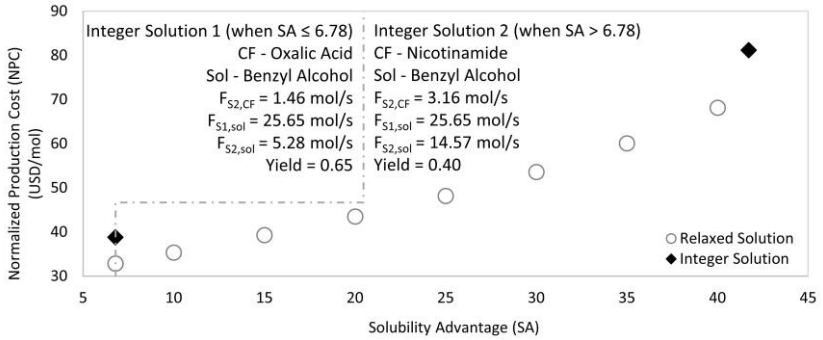


Figure 2: The Pareto fronts of the relaxed and integer solutions of the optimization problem (Eq.(P1)). The yield is defined as the fraction of the API in the feed that ends up in the cocrystal phase.

production of pharmaceutical cocrystals. This framework is particularly suitable for the early stages of process and product design, where the available experimental data are limited, and the identified set of CFs, solvents, and operating conditions can serve as the basis to launch an experimental program or more detailed simulations. Additionally, the method has the capability to recommend cocrystal property targets for efficient process and product design, which can guide future experimental efforts related to cocrystal discovery.

Acknowledgment

The work described in this paper was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, People's Republic of China (project no. 16214418).

References

Bardow, A., Steur, K., & Gross, J. (2010). Continuous-Molecular Targeting for Integrated Solvent and Process Design. Industrial & Engineering Chemistry Research, 49(6), 2834–2840.

Drud, A. (1985). CONOPT: A GRG code for large sparse dynamic nonlinear optimization problems. Mathematical Programming 1985 31:2, 31(2), 153–191.

Good, D. J., & Rodríguez-Hornedo, N. (2009). Solubility advantage of pharmaceutical cocrystals. Crystal Growth and Design, 9(5).

Gross, J., & Sadowski, G. (2001). Perturbed-Chain SAFT: An Equation of State Based on a Perturbation Theory for Chain Molecules. Industrial & Engineering Chemistry Research, 40(4), 1244–1260.

Gross, J., & Sadowski, G. (2002). Application of the Perturbed-Chain SAFT Equation of State to Associating Systems. Industrial & Engineering Chemistry Research, 41(22), 5510–5515.

Lange, L., & Sadowski, G. (2015). Thermodynamic Modeling for Efficient Cocrystal Formation. Crystal Growth & Design, 15(9), 4406–4416.

ter Horst, J. H., Deij, M. A., & Cains, P. W. (2009). Discovering New Co-Crystals. Crystal Growth & Design, 9(3), 1531–1537.

Wang, J., & Lakerveld, R. (2018). Integrated solvent and process design for continuous crystallization and solvent recycling using PC-SAFT. AIChE Journal, 64(4), 1205–1216.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50364-X

Optimizing the selection of drug-polymer-water formulations for spray-dried solid dispersions in pharmaceutical manufacturing

Suela Jonuzaj[a], Christopher L. Burcham[b], Amparo Galindo[a], George Jackson[a] and Claire S. Adjiman[a]
[a] *Department of Chemical Engineering, The Sargent Centre for Process Systems Engineering, Institute for Molecular Science and Engineering, Imperial College London, London SW7 2AZ, UK.*
[b]*Eli Lilly and Company, Indianapolis, IN, USA.*
c.adjiman@imperial.ac.uk

Abstract

In this work we present a systematic computer-aided design methodology for identifying optimal drug-polymer-water formulations with desired physical and chemical properties that are used in the spray drying of drug products. Within the proposed method, the UNIFAC model is employed to predict the solubility and miscibility of binary and ternary mixtures, whereas the Gordon-Taylor equation is used to estimate the glass transition temperature of a wide range of chemical blends. The design methodology is applied to the selection of optimal drug-polymer blends that maximize the loading of naproxen, while ensuring that stable formulations are designed. Finally, we explore the trade-offs between two competing objectives through multiobjective optimization, where the drug loading and water-content of API-polymer-water blends are maximized simultaneously. A ranked list of optimal solutions (mixtures with different chemicals and compositions) that can be used to guide experimental work is obtained by introducing integer cut inequalities into the model.

Keywords: Spray drying dispersion; optimal formulation; solubility; phase stability.

1. Introduction

Most of the new chemical entities in the drug discovery pipeline present unfavorable solubility properties, which often translates into low intrinsic bioavailability and makes the development of solid oral dosage forms challenging (Duarte et al., 2015). Spray-dried dispersion (SDD) is an effective technique to address these limitations; it has been successfully employed to improve the solubility and bioavailability of poorly soluble drugs in pharmaceutical manufacturing. SDD consists of an active pharmaceutical ingredient (API), preferably in its amorphous state to increase solubility, dispersed in a hydrophilic polymer matrix that stabilizes the amorphous form of the drug. Despite the benefits of this strategy, solid dispersions are often thermodynamically metastable, so there is a risk that the API may crystallize, leading to low-solubility behavior. To avoid such risks, it is important to select suitable API-polymer formulations that meet desired properties and ensure high solubility and bioavailability of the drug (Davis and Walker, 2018). In current industrial practice, heuristic approaches and experimental-based workflows are typically employed for pre-screening a small set of commonly used polymers for SDD. Such time-consuming and costly procedures can lead to longer development timelines and limited innovation. The development of computational

methodologies that improve the selection strategies in spray drying formulations and reduce the experimental efforts and required resources is therefore highly desirable.

In recent years, several researchers have developed model-based approaches that focus on predicting the phase behavior of given API-polymer systems using common thermodynamic tools. Among others, Bansal et al. (2016) and Tian et al. (2013) employed Flory-Huggins (F-H) theory to predict the solubility and miscibility of drug-polymer blends. Sadowski and co-authors have studied the long-term thermodynamic stability of amorphous solid dispersions (ASD) and constructed phase diagrams of binary and ternary systems using the perturbed-chain statistical associating fluid theory (PC-SAFT). In particular, Prudic et al. (2014) have investigated the impact of copolymer composition on the phase behavior of solid dispersions of different drug-polymer blends. Lehmkemper et al. (2017a & 2017b) used experimental data and thermodynamic modeling to construct binary phase diagrams and explore ASD physical stability for different API-polymer systems at various humidity levels. Dohrn et al. (2020 & 2021) studied the impact of solvents on the phase separation of ASDs during drying, computing the solubility and liquid-liquid phase diagrams of API-polymer-solvent systems using PC-SAFT.

Despite these recent modeling advances, a limited set of pre-defined blends, with fixed ingredients, have been investigated to date. Thus, the selection of suitable polymers that can be more effective for stabilizing existing and/or new drugs has not been explored systematically or via optimization methods. In this work, we present a systematic approach for identifying optimal (i) API-polymer and (ii) API-polymer-water formulations that meet desired physicochemical properties and can lead to drug products with improved bioavailability. Within the proposed methodology, property-prediction models are employed to estimate the solubility, miscibility and glass transition temperatures of a wide range of binary and ternary blends. In addition, we exploit advanced optimization techniques (Jonuzaj et al., 2016, 2019) to design improved formulations that yield high solubility and stability of the drug. As a case study, the design approach is applied to the selection of optimal polymers that maximize the drug loading of naproxen in different API-polymer and API-polymer-water mixtures, while ensuring phase stability and a sufficiently high glass transition temperature of the final blend. The proposed optimization model yields a ranked list of diverse high-performing solutions, where solid dispersion blends of the API with different polymers, polymer proportions and sorbed water content are identified.

2. Design Methodology

2.1. Problem definition

The proposed mathematical model involves the design of optimal API-polymer and API-polymer-water formulations used in spray drying dispersions. The solid dispersion blends to be designed consist of a predefined API, an optimal polymer (p_1), and sorbed water (when ternary systems are considered). The following set $I = \{\text{API}, p_1, H_2O\}$ represents all components in the designed blend, where the optimal polymer is selected from a predefined set $P = \{1, \dots, N_p\}$. Due to limited space, the mixed-integer programming problem (MINLP) presented in the next section considers the design of optimal binary API-polymer blends; it can be extended to formulating ternary API-polymer-water systems as shown in the supplementary information.

2.2. MINLP formulation for designing drug-polymer blends

The following mathematical formulation is derived to identify optimal polymers that maximize the drug loading of an API and satisfy pure and mixture property constraints:

$$\max_{w,x,y} \ \frac{w_{API}}{w_{p_1}} \tag{1}$$

$$s.t. \quad \breve{T}_{g,mix} = \frac{\breve{w}_{API} T_{g,API} + K\breve{w}_{p_1} T_{g,p_1}}{\breve{w}_{API} + K\breve{w}_{p_1}} \geq T_g^L \tag{2}$$

$$K = \frac{\rho_{API} T_{g,API}}{\rho_{p_1} T_{g,p_1}} \tag{3}$$

$$w_{API} \leq 0.95\breve{w}_{API} \tag{4}$$

$$\frac{\partial \ln \gamma_{API}(T,x)}{\partial x_{API}} + \frac{1}{x_{API}} \geq 0 \tag{5}$$

$$\ln \tilde{x}_{API}^{eq} + \ln \gamma_{API}(T, \tilde{x}^{eq}) = \frac{\Delta H_{fus,API}}{R}\left[\frac{1}{T_{m,API}} - \frac{1}{T}\right] \tag{6}$$

$$w_i = \frac{x_i MW_i}{\sum_{j\in I} x_j MW_j};\ \tilde{w}_i^{eq} = \frac{\tilde{x}_i^{eq} MW_i}{\sum_{j\in I} \tilde{x}_j^{eq} MW_j};\ i = API, p_1 \tag{7}$$

$$\sum_{p=1}^{N_p} y_{p_1,p} = 1 \tag{8}$$

$$x \in [x^L, x^U] \subset \mathbb{R}^I;\ w \in [w^L, w^U] \subset \mathbb{R}^I;\ y \in \{0,1\}^q$$

The vectors x and w represent the mole and mass fractions in the designed blend, respectively; $\tilde{w}^{eq}$ and $\breve{w}$ represent mass fractions on the solid-liquid and glass transition phase boundaries, respectively; $y_{p_1,p}$ is a binary variable for assigning a polymer p from the set P to the polymer component p_1 in the designed blend. Eq. (1) is the objective function, where the ratio of API/polymer is maximized. Eqs. (2)-(8) are property constraints for estimating the glass transition temperature, the solid-liquid equilibrium and the miscibility of API-polymer blends. A schematic phase diagram of a drug-polymer blend is shown in Figure 1, where the blend is designed to be in areas A and B at system temperature T through eqs. (2)-(7). The glass transition temperature $\breve{T}_{g,mix}$ of a binary blend (green curve) is calculated using the Gordon-Taylor equation (Gordon and Taylor, 1952) given in eqs. (2) & (3), where ρ_i and $T_{g,i}$ are the density and glass transition temperature of component i, respectively. $\breve{T}_{g,mix}$ is set to be higher than a user-specified value T_g^L. To ensure that the API has low molecular mobility and does not undergo rapid crystallization in the designed blend, an offset in mass fraction is imposed in eq. (4), so that the glass transition temperature of the blend is well below $\breve{T}_{g,mix}$. Through eq. (5), the API-polymer blend is miscible at the chosen composition x. The solubility $\tilde{x}_{API}^{eq}$ is calculated in eq. (6), where the API heat of fusion $\Delta H_{fus,API}$ and melting point $T_{m,API}$ are taken from experimental data and R is the gas constant. Activity coefficients are

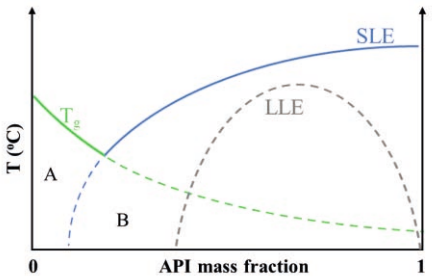


Figure 1: Schematic phase behavior of an API-polymer blend (adapted from Lehmkemper et al., 2017a). The blue curve represents solid-liquid equilibrium (SLE), the gray curve is liquid-liquid equilibrium (LLE), which may or may not be present, and the green curve the glass-transition temperature of the blend. In order to ensure the stability of the blend at room temperature, the desired operating region includes areas A & B.

calculated using the UNIFAC model (Fredenslund et al., 1975). The mass fractions of the API in the designed blend and at SLE are calculated in eq. (7), where MW_i is the molar mass of component i. Note that the binary blend can be above or below the SLE curve, so $\tilde{x}_i^{eq}$ and $\widetilde{w}_i^{eq}$ could be calculated after the solution of the optimization problem. The logic relation presented in eq. (8) ensures only one polymer is selected from the set P.

3. Case study: optimal binary and ternary blends for the SSD of naproxen

3.1. Problem description

Naproxen (NPX) is a nonsteroidal anti-inflammatory drug commonly used to treat pain in joints and muscles. We selected naproxen to demonstrate the design methodology as it is a well-studied drug for SDD, and it can be modeled with UNIFAC. The proposed systematic methodology is applied to the design of optimal binary (NPX-polymer) and ternary (NPX-polymer-water) systems, where problems of increasing complexity are formulated and solved. First, a binary NPX-polymer system is considered, where optimal polymers are selected from a set of 25 candidates to maximize the drug loading of naproxen. The number of monomers (repeated units) in each polymer is allowed to vary between 10 and 200 in order to explore a wide range of polymers with different sizes and structures. Next, a larger design problem that takes into account a ternary NPX-polymer-water solid dispersion blend is formulated via multiobjective optimization (MOO), where the drug loading and water-content of the ternary mixture are optimized simultaneously. Relevant measured naproxen data are included in Table 1 and the optimal solutions obtained with the binary and ternary models are discussed in the next section.

Table 1: Measured naproxen property data and system temperature used in this case study.

API	Mw_{API} (g/mol)	$\Delta H_{fus,API}$ (J/mol)	$T_{m,API}$ (K)	$T_{g,API}$ (K)	T (K)	T_g^L (K)
Naproxen	230.26	32673.75	427.32	277.15	298.15	338.15

3.2. Results and discussion

The design formulations are implemented and solved in GAMS version 36.2.0, using local and global algorithms. The two MINLP models for designing binary and ternary blends for the SDD of naproxen can be found at *doi.org/10.5281/zenodo.5637599* and the results are given in Table 2 and Figure 2. Some of the solutions reported are global solutions obtained with SCIP. Where global optimality is not reached with either SCIP or BARON within 1000 s, local solutions obtained with SBB are presented. For cases where both SBB and SCIP terminate successfully, the two solvers converge to the same solution (triangles in Figure 2), giving confidence in the performance of the local solver.

The ranked list of optimal solutions presented in Table 2 shows that (hydroxypropyl)methyl cellulose p55, methacrylic acid-methyl methacrylate, and polyvinylpyrrolidone K30 are promising polymer candidates that yield high loading of naproxen in the polymeric carrier. The mass fraction of naproxen (w_{API}) in each binary mixture is higher than the mass fraction of the API at solid-liquid equilibrium ($\widetilde{w}_{API}^{eq}$), and lower than the mass fraction at the glass transition boundary ($\breve{w}_{API}$). In addition, the optimization model ensures the designed binary blends are miscible at optimal compositions of naproxen and polymer in the mixture, so that phase separation is

prevented. Thus, the designed API-polymer blends are in the desired phase region, i.e., below the solubility and glass transition curve, and outside the immiscible area.

Table 2: Top 3 optimal solutions of the API-polymer model, including the optimal API/polymer ratios; the mass fraction of API in the designed blend (w_{API}); the SLE mass fraction ($\widetilde{w}_{API}^{eq}$); the mass fraction ($\breve{w}_{API}$) on the reference glass transition curve ($\breve{T}_{g,mix}$); the identity of the optimal polymers and the number of the repeated units (N_m) in each polymer.

w_{API}/w_{p1}	w_{API}	$\widetilde{w}_{API}^{eq}$	$\breve{w}_{API}$	$\breve{T}_{g,mix}(\breve{w}_{API})$	Polymers	N_m
1.601	0.616	0.001	0.648	338.15	HPMC p55	10
1.181	0.541	0.001	0.570	338.15	Eudragit L100	11
1.052	0.513	0.003	0.540	338.15	PVP K30	10

In the ternary model, the drug/polymer ratio and water content are optimized simultaneously via MOO in order to investigate the maximum amount of water that can absorbed in the designed blends while maintaining high drug loading. The set of Pareto optimum solutions given in Figure 2 is obtained using the ϵ-constraint method, in which the water mass fraction is maximized, and the API/polymer ratio is constrained by a given lower bound, ϵ. The value of ϵ is increased from 0.2 to 2 with a step-size of 0.2. All Pareto points share the same monomer structure, HPMC p55, but differ in the polymer size and water-polymer composition (cf. Table S1 in supplementary information). We note that the miscibility constraints were not included in the formulation of the ternary problem. Further investigation is required in order to ensure there is no phase separation of the optimal ternary blends designed.

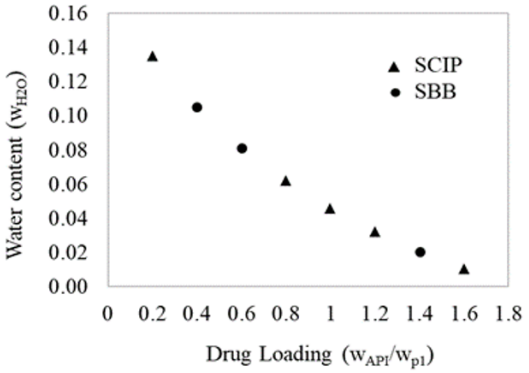


Figure 2: Optimal drug loading and water mass fraction when solving the multiobjective optimization problem for the ternary API-polymer-water system.

4. Conclusions

In this work we have developed a systematic computer-aided methodology for identifying optimal formulations for the spray drying of drug products. Within the proposed approach, two optimization models were formulated and solved. First, a two-component formulation was derived, where optimal API-polymer mixtures that meet desired property targets and satisfy given phase boundary conditions, were identified. A ranked list of optimal solutions (different binary blends with optimized polymer structure, size and compositions) was obtained by introducing integer cut inequalities into the model. Next, a ternary API-polymer-water system was formulated via multiobjective optimization, where the drug loading and water fraction were optimized simultaneously. Future work will focus on validating the modeling results with experimental data (where possible), and constructing binary and ternary phase diagrams of the optimal mixtures obtained. In addition, the models will be extended to designing API-polymer-solvent blends, taking

into account key process constraints used in spray drying solvent selection, spray solution viscosity, spray atomization and final particle size of the solid spray dried particles. Finally, the proposed models will be incorporated in a polymer and solvent selection workflow for spray drying that can be used to identify better-performing designs in a more efficient and systematic way.

Acknowledgments: The authors gratefully acknowledge financial support from Eli Lilly and Company, and the EPSRC (EP/T005556/1).
Data statement: Data underlying this article can be accessed and used under the Creative Commons Attribution license on Zenodo at *doi.org/10.5281/zenodo.5637599*.

References

Bansal, K., Baghel, U.S., Thakral, S., 2016, Construction and validation of binary phase diagram for amorphous solid dispersion using flory-huggins theory, AAPS PharmSciTech 17, 318-327.

Davis, M., Walker, G., 2018, Recent strategies in spray drying for the enhanced bioavailability of poorly water-soluble drugs, Journal of Controlled Release 269, 110-127.

Dohrn, S., Reimer, P., Luebbert, C., Lehmkemper, K., Kyerematen, S.O., Degenhardt, M., Sadowski, G., 2020, Thermodynamic modeling of solvent-impact on phase separation in amorphous solid dispersions during drying, Molecular Pharmaceutics 17, 2721-2733.

Dohrn, S., Luebbert, C., Lehmkemper, K., Kyerematen, S.O., Degenhardt, M., Sadowski, G., 2021, Solvent influence on the phase behavior and glass transition of amorphous solid dispersions, European Journal of Pharmaceutics and Biopharmaceutics 158, 132-142.

Duarte, I., Santos, J.L., Pinto, J.F., Temtem, M., 2015, Screening methodologies for the development of spray-dried amorphous solid dispersions, Pharmac. Research 32, 222-237.

Fredenslund, A., Jones, R.L., Prausnitz, J.M., 1975, Group-contribution estimation of activity coefficients in nonideal liquid mixtures, AIChE Journal 21, 1086-1099.

Gordon, M., Taylor, J.S., 1952, Ideal copolymers and the second-order transitions of synthetic rubbers. i. non-crystalline copolymers, Journal of Applied Chemistry 2, 493-500.

Jonuzaj, S., Akula, P.T., Kleniati, P.M., Adjiman, C.S., 2016, The formulation of optimal mixtures with generalized disjunctive programming: A solvent design case study, AIChE J. 62, 1616-33.

Jonuzaj, S., Cui, J., Adjiman, C.S., 2019, Computer-aided design of optimal environmentally benign solvent-based adhesive products, Computers & Chemical Engineering 130, 106518.

Lehmkemper, K., Kyerematen, S.O., Heinzerling, O., Degenhardt, M., Sadowski, G., 2017a, Long-term physical stability of PVP- and PVPVA-amorphous solid dispersions, Molecular Pharmaceutics 14, 157-171.

Lehmkemper, K., Kyerematen, S.O., Heinzerling, O., Degenhardt, M., Sadowski, G., 2017b, Impact of polymer type and relative humidity on the long-term physical stability of amorphous solid dispersions, Molecular Pharmaceutics 14, 4374-4386.

Prudic, A., Kleetz, T., Korf, M., Ji, Y., Sadowski, G., 2014, Influence of copolymer composition on the phase behavior of solid dispersions, Molecular Pharmeceutics 11, 4189-4198.

Tian, Y., Booth, J., Meehan, E., Jones, D. S., Li, S., Andrews, G.P., 2013, Construction of drug–polymer thermodynamic phase diagrams using Flory–Huggins interaction theory: identifying the relevance of temperature and drug weight fraction to phase separation within solid dispersions, Molecular Pharmaceutics 10, 236−248.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50365-1

Integrated design of injectable manufacturing processes considering characteristics of process- and discrete-manufacturing systems

Masahiro Yamada[a], Isuru A. Udugama[a], Sara Badr[a], Kenichi Zenitani[b], Kokichi Kubota[b], Hayao Nakanishi[b], Hirokazu Sugiyama[a*]

[a]*Department of Chemical System Engineering, The University of Tokyo, Tokyo 113-8656, JAPAN*
[b]*Settsu Plant, Shionogi Pharma Co Ltd., Osaka 566-0022, JAPAN*
**sugiyama@chemsys.t.u-tokyo.ac.jp*

Abstract

This work proposes a superstructure-based design approach to optimize an end-to-end injectable manufacturing process. At the core of this approach are unit operation models for batch and continuous operations and a plant-wide scheduling model that can explicitly model the semi-continuous operations of injectable manufacturing to a high level of precision. An integrated evaluation model consisting of a Net Present Value (NPV) and technology readiness modules is used to identify the optimal flowsheet. This approach can simultaneously optimize the design of a given process flowsheet alternative consisting of both batch and continuous unit operations considering product parameters, process conditions, and market characteristics. The overall approach developed was then demonstrated on an end-to-end injectable manufacturing case study with four process flowsheet alternatives (batch and continuous operations of compounding and lyophilization). Based on the techno-economic analysis, it was shown that the alternative with batch compounding and lyophilization was preferred.

Keywords: Process design; injectable manufacturing; optimization; continuous manufacturing; superstructure

1. Introduction

Injectables are an effective dosage form for many pharmaceutical products, including COVID-19 vaccines (Alharbi, 2021). Injectables are typically manufactured in a production line consisting of solution compounding, sterile filtration, filling, lyophilization, and inspection unit operations. Currently, these units are operated as batch processes. While continuous technologies are being developed for injectable manufacturing, they are still in the early stage of conceptualization and research. Examples include continuous compounding (Casola et al., 2015) and lyophilization (Bockstal et al., 2017). The main driving force for introducing these continuous technologies is the promise of reduced manufacturing cost, environmental load, and improved process flexibility.

As a result, in the future, there is the possibility to design an injectable manufacturing process where each unit can be operated in either batch or continuous modes. From a design perspective, this requires evaluating multiple potential process flowsheet alternatives considering multiple operational, design, and economic factors. As such,

there is a need to develop decision-support tools that can enable informed decision-making without entirely relying on empirical knowledge.

Superstructure optimization is such a decision-support concept. In short, the superstructure-based approach evaluates the manufacturing system's overall performance rather than focusing on a single unit operation. As a result, this approach can find a globally optimal process design (Quaglia et al., 2015). In the chemical industry, superstructure based optimization is a commonly used method for determining process sequence (Tian et al., 2020). While in pharmaceutical production processes, superstructure-based approaches have been used for choosing between batch and continuous unit operations in tablet manufacturing (Matsunami et al., 2020). Figure 1 illustrates a tentative superstructure of the injectable manufacturing process, where operation mode and equipment materials choices are listed. The set of process alternatives that need to be evaluated are all possible choice combinations.

To accurately evaluate process alternatives generated from a superstructure, there is a need for an approach that can explicitly optimize their process design and operations. In the context of injectable manufacturing, this requires the development and integration of two types of models. One model that can accurately capture the continuous operations which are “process-manufacturing systems” where the bulk solution behavior is essential. Another model that can describe batch operations which are “discrete-manufacturing systems” where process operation behavior is essential. In addition, the scheduling of the overall manufacturing process is needed as the same production line needs to process multiple products while guaranteeing sterility levels. The output of these combined models then must be techno-economically evaluated. The rest of this work is organized as follows. In section 2, the overall superstructure-based approach is introduced. This is followed by section 3, where the proposed approach is demonstrated on a case study. Finally, in section 4, the conclusions are given.

2. Methodology

For process optimization, a model-based approach using superstructure was suggested. First, a superstructure covering all possible candidates was created based on literature data and discussions with production experts. Figure 1 illustrates the overall superstructure and the multiple decision variables that can be considered, such as the operation mode (batch vs continuous), equipment material (stainless vs resin), and tasks (machine vs operator inspection). In the future, further choices can be added to the superstructure.

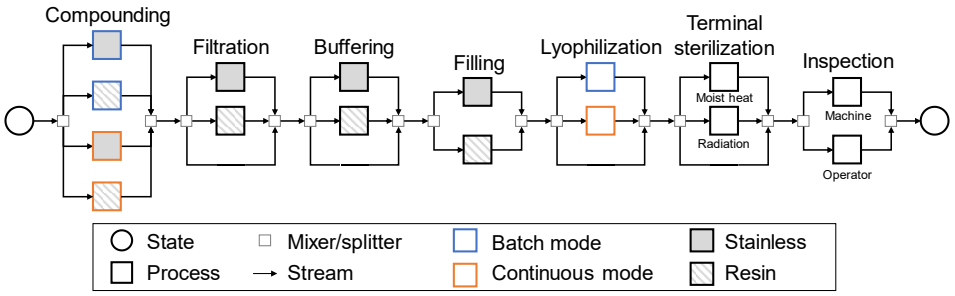


Figure 1 Superstructure of injectable manufacturing process

Figure 2 displays the overall modeling framework followed to evaluate each process flowsheet candidate systematically. The entire model can be divided into three parts: i) process flowsheet model, ii) scheduling model, and iii) evaluation model. Through these models, the optimal candidate is obtained based on NPV and Technology Readiness Level (TRL) evaluation. Due to the level of fidelity of the process flowsheet model, this framework can optimize the economic performance of each process alternative identified considering multiple design choices, market conditions and process parameters. Hence, applying the approach developed with a superstructure-based flowsheet generation enables the optimal process alternative and the corresponding process design to be found simultaneously.

2.1. Process flowsheet model

The objective of the process flowsheet model is to calculate the required equipment size, the input and output material flows, and the expected unit operation processing time for a one-lot production. To carry out this evaluation, the process flowsheet model uses an integrated process and discrete production module. The process module explicitly considers Active Pharmaceutical Ingredient (API) crystal dissolution and Residence Time Distribution (RTD) factors. On the other hand, the discrete module considers lot residue, lot disposal, and operational defects which are critical operational factors for discrete process operations. The scheduling model receives the expected processing time for each unit operation for single-lot production from the process flowsheet model.

2.2. Scheduling model

The objective of the scheduling model is to evaluate if the production line can meet product demand and inventory requirements with the given operational constraints and changeover tasks. Suppose a single product is manufactured in a production line. In that case, the repetition of the lot production can meet the demand and inventory margin if there is sufficient time allocated for production. However, when multiple products are manufactured in a production line, the production sequence must be optimized to meet the constraints and minimize the non-production time. If there is no solution to satisfy the constraints, the design alternative is evaluated as unfeasible.

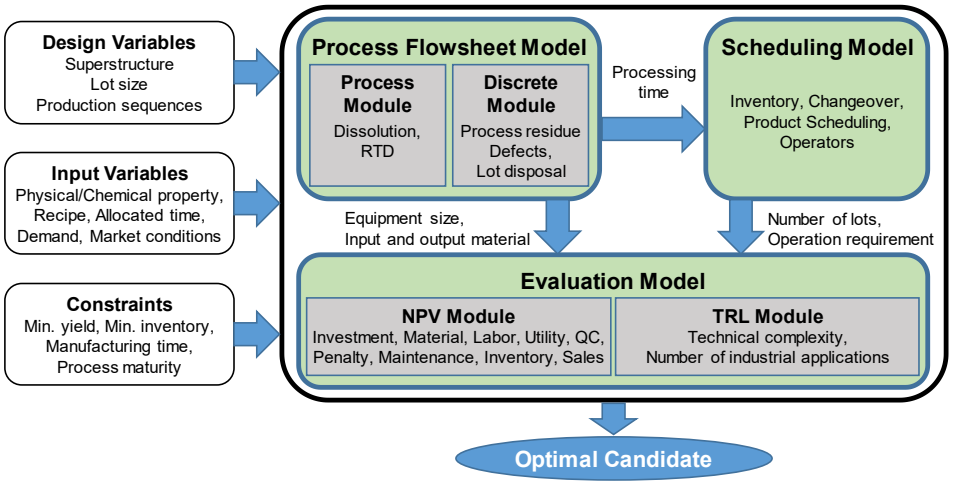


Figure 2 Framework of the constructed models

2.3. Evaluation model

The evaluation model consists of the NPV module and TRL module. The overall objective of these two evaluation modules is to calculate the economic performance of the process flowsheet (considering the intricacies of process- and discrete-manufacturing systems) and to calculate the overall technology readiness level.

In the NPV module, NPV is calculated as an economic indicator, as shown in Eq. (1).

$$\max NPV = -investment + \sum \frac{CF}{(1+r)^n} \tag{1}$$

Investment cost $investment$ [JPY] and annual cash flow CF are calculated by considering the following factors: the annual cash flow, material cost, utility cost, labor cost, quality control cost, penalty cost, maintenance cost, and inventory cost as well as the product sales. The process flowsheet model generates information related to equipment sizing and material mass per lot. The operational requirement information, e.g., the sum of production duration, required space for inventory management, and the number of lots for each product, is generated from the scheduling model. In the TRL module, the overall process flowsheet is given a score between a TRL of 1 and 9. The TRL score is based on the state of progress in research, experiment, and implementation of each unit operation that is selected. This evaluation is carried out according to the guidelines set in Li et al. (2019) and Silk et al. (2020).

3. Case study

The proposed approach was then applied to a case study examining four process alternatives, as illustrated in figure 3. Each alternative consisted of end-to-end injectable manufacturing operations where the operations of the compounding and lyophilization units were changed between batch and continuous modes. Following assumptions were made during the subsequent analysis: 1) Case study represents a grass-root design, where resin equipment and final sterilization were not considered; 2) Three products A, B, and C, with fixed demand, were manufactured in one production line. Product A and B is liquid products, and C is a freeze-dried product, which needs lyophilization; 3) Batch/continuous operation in compounding and lyophilization were examined. Inspection with a machine was chosen for liquid products, and inspection with operators was chosen for the freeze-dried product; 4) Production of 5 days per week and 24 h per day was assumed; 5) The sale period was set as 20 years; 6) The lot size is varied as a process design variable, which is shown in Eq. (2).

$$V_{\text{lot}}\,[\text{L}] \in \{50, 100, 200\} \tag{2}$$

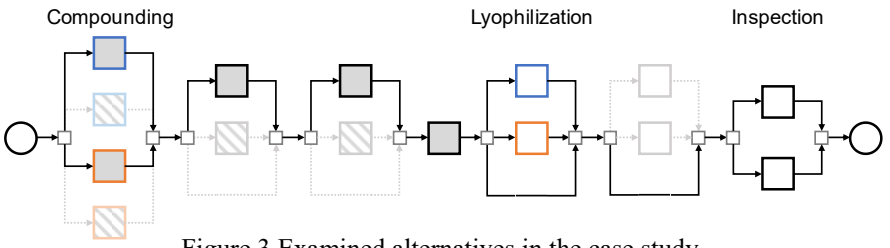


Figure 3 Examined alternatives in the case study

Over a half million design alternatives were analyzed for each of the four process alternatives. Figure 4 (a) illustrates a violin plot of production times observed for each process alternative. Analyzing Figure 4 (a), continuous operations in the lyophilization unit resulted in a noticeable reduction in the overall production time, which is in line with previous results (Pisano et al., 2019). In comparison, the mode of operation in the compounding unit had a minor effect. The design condition of each process alternative has a significant influence on the observed production duration. For instance, a good process alternative (e.g continuous compounding and lyophilization) can have a longer production time than a bad process alternative (e.g batch compounding and lyophilization) purely due to design choices. This shows the importance of simultaneous process alternatives selection and process design optimization during the decision-making processes.

Figure 4 (b) illustrates the NPV of selected design conditions for the four process alternatives where the lot size and production sequences were fixed. The production time of four alternatives is marked in red in Figure 4 (a). It can be seen that continuous lyophilization and batch compounding results in the best outcome. It can also be seen that continuous lyophilization choice has a notable positive impact on the NPV. This is mainly because of the reduction in labor costs due to the shortened production time. In comparison, the introduction of continuous compounding results in a decrease in NPV. This is because the cost of investment and initial start-up requirements outweigh other benefits. It should be noted that these designs have not been economically optimized by considering design factors such as lot size, production sequence, and other market factors. Instead, designs with a shorter overall production time were selected.

From a TRL point of view, batch compounding and lyophilization are tried and tested technologies implemented in the industry. Hence, batch compounding and lyophilization operations can be awarded a TRL of nine. In contrast, continuous operations are only implemented for slurry and liquid, and not for API crystal and water for injection. Hence continuous compounding and lyophilization can be awarded a TRL score of three, as only the fundamentals are established.

Overall the fully batch alternative is recommended over other alternatives as 1) the NPV of the fully batch alternative is only 10% lower than the most optimal process alternative and 2) the significant difference in TRL between batch and continuous operations.

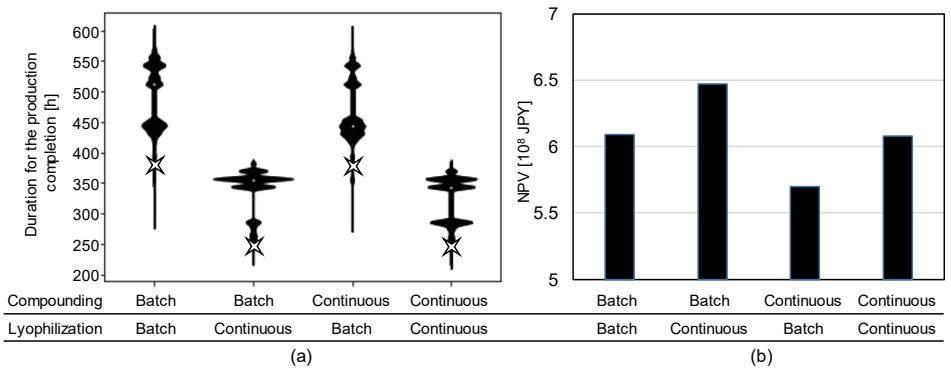


Figure 4 Evaluation result for (a) total duration for the production and (b) NPV

4. Conclusion

A model-based design approach for the injectable manufacturing process was presented. A superstructure was constructed to consider multiple process candidates while process and discrete models were developed for each unit operation to optimize the process and design conditions for each process candidate. An evaluation model consisting of an NPV and TRL analysis was developed to rate each candidate's performance. The approach was applied to a case study where batch and continuous compounding/lyophilization alternatives were examined. The case study illustrated the importance of simultaneous determination of process alternatives and design considering the product, process, market characteristics. In this instance, the alternative with batch compounding and continuous lyophilization was shown to be economically optimal. However, when considering TRL, the fully batch operations was a better alternative. As future work, the economic evaluation should be done for all alternatives, including the process parameter and production sequences. Model updates are also necessary to evaluate other process alternatives and different equipment materials.

5. References

N. Alharbi, 2021. Overview of SARS-CoV-2 and COVID-19 Vaccine, Bahrain Medical Bulletin, 43 (2), 511–515

G. Casola, S. Yoshikawa, H. Nakanishi, M. Hirao, H. Sugiyama, 2015. Systematic retrofitting methodology for pharmaceutical drug purification processes, Computers and Chemical Engineering, 80 (2), 177–188

B. Li, I. A. Udugama, S. S. Mansouri, W. Yu, S. Baroutian, K. V. Gernaey, B. R. Young, 2019. An exploration of barriers for commercializing phosphorus recovery technologies, Journalk of Cleaner Production 229 (20), 1342–1354

K., Matsunami, F. Sternal, K. Yaginuma, S. Tanabe, H. Nakagawa, H. Sugiyama, 2020. Superstructure-based process synthesis and economic assessment under uncertainty for solid drug product manufacturing, BMC Chemical Engineering, 2 (1), 1–16

R. Pisano, A. Arsiccio, L. C. Capozzi, B. L. Trout, 2019. Achieving continuous manufacturing in lyophilization: Technologies and approaches, European Journal of Pharmaceutics and Biopharmaceutics, 142, 265–279

A. Quaglia, C. L. Gargalo, S. Chairakwongsa, G. Sin, R. Gani, 2015. Systematic network synthesis and design: Problem formulation, superstructure generation, data management and solution, Computers and Chemical Engineering, 72 (2), 68–86

D. Silk, B. Mazzali, C. L. Gargalo, M. Pinelo, I. A. Udugama, S. S. Mansouri, 2020. A decision-support framework for techno-economic-sustainability assessment of resource recovery alternatives, Journal of Cleaner Production, 266 (1), 121854

G. Tian, A. Koolivand, N. S. Arden, S. Lee, T. F. O'Connor, 2019. Quality risk assessment and mitigation of pharmaceutical continuous manufacturing using flowsheet modeling approach, Computers & Chemical Engineering, 129 (4), 106508

P. Van Bockstal, S. Mortier, L. De Meyer, J. Corver, C. Vervaet, I. Nopens, T. De Beer, 2017. Mechanistic modelling of infrared mediated energy transfer during the primary drying step of a continuous freeze-drying process, European Journal of Pharmaceutics and Biopharmaceutics, 114, 11–21

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50366-3

A bi-level decomposition approach for CAR-T cell therapies supply chain optimisation

Niki Triantafyllou[a], Andrea Bernardi[a], Matthew Lakelin[b], Nilay Shah[a], Maria M. Papathanasiou[a*]

[a]*Sargent Centre for Process System Engineering, Imperial College London, SW72AZ, London, United Kingdom*
[b]*TrakCel Limited, 10/11 Raleigh Walk, Cardiff, CF10 4LN UK*
maria.papathanasiou11@imperial.ac.uk

Abstract

Autologous cell therapies are based on bespoke, patient-specific manufacturing lines and distribution channels. They present a novel category of therapies with unique features that impose scale out approaches. Chimeric Antigen Receptor (CAR) T cells are an example of such products, the manufacturing of which is based on the patient's own cells. This automatically: (a) creates dependencies between the patient and the supply chain schedules and (b) increases the associated costs, as manufacturing lines and distribution nodes are exclusive to the production and delivery of a single therapy. The lack of scale up opportunities and the tight return times required, dictate the design of agile and responsive distribution networks that are eco-efficient. From a modelling perspective, such networks are described by a large number of variables and equations, rendering the problem intractable. In this work, we present a bi-level decomposition algorithm as means to reduce the computational complexity of the original Mixed Integer Linear Programming (MILP) model. Optimal solutions for the structure and operation of the supply chain network are obtained for demands of up to 5000 therapies per year, in which case the original model contains 68 million constraints and 16 million discrete variables.

Keywords: CAR T cell therapy; supply chain optimisation; MILP; personalised medicine; bi-level decomposition.

1. Introduction

Chimeric Antigen Receptor (CAR) T cell therapy is a type of immunotherapy, where the patient's own immune system is utilised to recognise and kill cancer cells (Sadelain *et al.* 2015). The patient's T cells are removed from the bloodstream and are genetically engineered to express the CAR, rendering them capable of recognising and attacking the target tumour cells. CAR T cells can be obtained from the patient's own blood (autologous) or the lymphocytes of another healthy donor (allogeneic). Following the success in Phase 1 of clinical trials, the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA) have approved 5 of these therapies so far (UPMC, 2021). Currently, there are 6,581 active and ongoing clinical trials regarding CAR T cell treatments, with most of them being autologous, while their allogeneic counterpart is progressing as well (Caldwell *et al.* 2021). The relatively high prices of these therapies can be partially attributed to the high manufacturing, distribution and product administration costs (Spink *et al.* 2018). Time-intensive manufacturing processes, in-time delivery under hospital admission and daily monitoring of the patient for side effects are among the factors that increase the cost (Han *et al.* 2021). Another key challenge of the

CAR T cell therapy lifecycle is the minimisation of the turnaround time, which varies between 15 and 24 days for the commercially available treatments (Nucleus Biologics, 2021). The in-time delivery is of utmost importance for the patients as late administration may negatively impact the response to treatment. To address these challenges, digital tools such as mathematical models and optimisation policies are used to assist the decision-making process by coordinating the different tasks and identifying the optimal supply chain network structures (Sarkis *et al.* 2021a). The complexity of the CAR T cell supply chain can be easily observed by the product's lifecycle and autologous nature that challenge the identification of an optimal supply chain network (Sarkis *et al.* 2021b). The main steps of a typical CAR T cell therapy lifecycle are: (a) patient identification, (b) leukapheresis, (c) manufacturing, (d) Quality Control, (e) therapy administration.

There have been works in the literature focusing on the optimisation of CAR T cell therapies via Mixed Integer Linear Programming (MILP) models (Bernardi *et al.* 2021; Karakostas *et al.* 2020). The autologous nature of these therapies often results in novel supply chain formulations (Papathanasiou *et al.* 2020), where the MILP problem comprises a significantly high number of integer variables. The latter can range from 600,000 for clinical trial applications to over 16 million for commercial scales of average demand. This can render the convergence to global or sometimes even local optimality infeasible. Hence, methodologies to enable the solution of large-scale instances are required (Erdirik-Dogan and Grossmann, 2008; Terazzas-Moreno and Grossmann, 2011). In this work, we present a bi-level decomposition algorithm capable of providing candidate solutions with respect to the location, number and capacity of manufacturing sites, and the most suitable mode of transport. Optimal solutions for the structure and operation of the supply chain network are obtained for demands of up to 5000 therapies per year under three different time constraint scenarios (17, 18 and 19 days) total return time. The latter refers to the total duration of the therapy life cycle, starting from the leukapheresis procedure and ending with the delivery of the therapy at the hospital.

2. Materials and methods

The model examined in this work is an in-house MILP model that describes the CAR T cell supply chain, used for the identification of the optimal supply chain network structure for in-time delivery of the therapies (Bernardi *et al.* 2021). An overview of the model formulation is presented in Table 1. The supply chain network includes 4 nodes; namely, leukapheresis site, manufacturing site, quality control (QC), and hospital (Figure 1). More specifically, a patient is allocated to a specialised leukapheresis site, where T cells are isolated from the bloodstream. Subsequently, the leukapheresis material undergoes freezing (cryopreservation) and is shipped to the manufacturing site. After the completion of the manufacturing process, the final product is tested in the QC site, which is co-located with the manufacturing facilities. Finally, the cryopreserved CAR T cell therapy is transported to the hospital for administration.

The objective is to minimize the total cost of the therapies over a long-term planning horizon (year quarter), whilst operating the supply chain in the short-term (daily) and fulfilling several constraints. The supply chain network's performance (full space model) is assessed for different demand scenarios (200, 500, 1000, 2000, 3000 patients per year) generated by an in-house algorithm and different return times (17, 18, and 19 days). The model parameters, such as the demand profiles and cost coefficients are assumed to be deterministic. The study considers 4 leukapheresis sites and 4 hospitals in the UK and 6

manufacturing sites located in the UK and Europe. The manufacturing facilities have a capacity of 4, 10, or 31 parallel lines, and a forward-looking scenario of a manufacturing time of 7 days is considered.

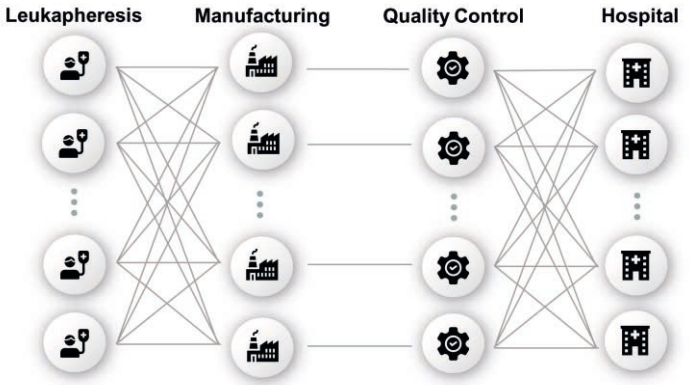


Figure 1. CAR T cell supply chain network with 4 nodes: (a) leukapheresis site, (b) manufacturing site, (c) Quality Control, and (d) hospital.The hospital and the leukapheresis centre are collocated in this case.

Table 1: Overview of the model formulation.

Index	Mathematical Formulation	Description
Objective function		
(1)	$min\, C_{total\ cost} = C_{manufacturing} + C_{transport} + C_{quality\ control}$	Total cost of therapies
Constraints		
(2)	$TRT_p = t_{delivery} - t_{start} \leq U^t$	Return time of therapy
(3)	$CAP_{m,t} = FCAP_m - \sum_p INM_{p,m,t}$	Capacity constraint
(4)	$X1_{c,m} \leq E1_m, \forall c, m, \quad X2_{m,h} \leq E1_m, \forall c, h, \quad \sum_m E1_m \leq U^M$	Network constraints
(5)	$INC_{p,c,t} = OUTC_{p,c,t+TLS}$, $INM_{p,m,t} = OUTM_{p,m,t+TMFE+TQC}$	Sample balances at each node (leukapheresis, manufacturing)
(6)	$LSR_{p,c,m,j,t} = LSA_{p,c,m,j,t+TT1_j}, \quad \forall p,c,m,j,t$, $FTD_{p,m,h,j,t} = MSO_{p,m,h,j,t+TT2_j}, \quad \forall p,m,h,j,t$	Transport constraints

An assessment of the model's complexity was conducted in order to evaluate the capabilities of the full space MILP problem. In this case, CPLEX results into global optimum solutions in less than an hour for small-scale problems. However, an increase in the number of patients and the planning horizon, makes the problem computationally intractable and leads to CPU times over 10,000s. It is shown that the CAR T cell supply chain problem is a problem with complicating constraints, and it is more sensitive to an increase in the number of therapies rather than an increase in the time horizon. The bottleneck constraints were identified via a model complexity analysis. These are the two transport constraints (Table 1) based on which the therapy is shipped from the leukapheresis site to the manufacturing site and from the manufacturing site to the hospital. Hence, the search space of the detailed model becomes very large for commercial scale problem instances, mainly because of the penta-dimensional transportation variables. The identified complicating constraints are key elements in the development of the bi-level decomposition algorithm. To keep the problem computationally tractable for a higher number of therapies, a bi-level decomposition

approach is proposed. The original detailed model is decomposed into an upper-level planning and a lower- level scheduling problem in order to decrease the computational complexity and solve for larger instances of the problem.

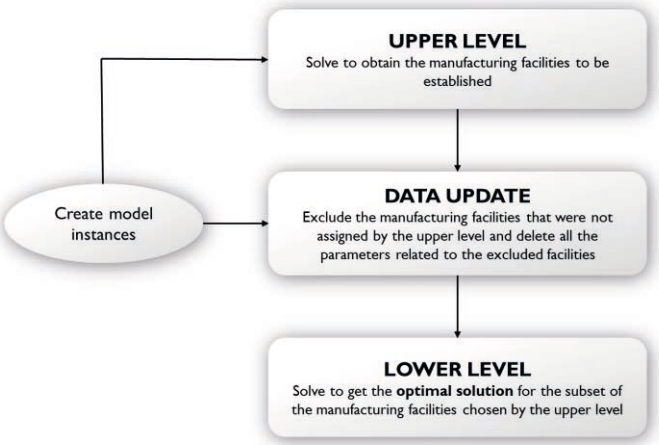


Figure 2. Flowchart for the proposed algorithm.

Figure 2 presents a flowchart of the proposed bi-level decomposition algorithm. The upper-level model is a relaxation of the original full space model, and it is responsible for strategic planning. More specifically, it chooses the number and the location of the manufacturing facilities to be established. The upper-level model is lower-dimensional compared to the original and thus the candidate transport modes and hospitals are considered fixed. The supply chain model for the upper level considers the network up until the end of manufacturing and all the constraints regarding material flows post manufacturing are ignored here. Consequently, the complicating constraint about the shipping of the therapy from the manufacturing site to the hospital is eliminated from the upper level. Almost all the variables in the upper level are identical to the ones in the original detailed model, apart from some that are redefined to fit the dimensional changes in the upper level. Given that the upper level is a relaxation of the original model, its solution provides a lower bound to the lower level subproblem. The lower-level model is a subproblem of the original detailed scheduling model, as it is solved for the subset of the manufacturing facilities chosen by the upper level. Specifically, manufacturing sites that were predicted not to be established by the upper level, are excluded from the lower level. Hence, the search space of the lower-level model is significantly reduced. The lower level can choose the same or a subset of the manufacturing sites that were predicted in the upper level. The optimal solution of the lower-level model becomes the final solution of the problem.

3. Results and discussion

All the models have been implemented in Python 3.7.1 and Pyomo 5.6.1 and solved with CPLEX 12.9. All computational experiments were performed in a 24-core Xeon E5-2697 machine with 96GB. Here we present the results of seven problem cases of increasing size. All cases were solved using both the bi-level decomposition algorithm, as well as the full-space model, aiming to assess the capabilities of the former. The bi-level algorithm was tested for cases of up to 5000 therapies per year, two different demand

profiles (A and B) for each case, and different turnaround times (17, 18, and 19 days). It should be noted that with the proposed algorithm both subproblems reach global optimality. In addition, most of the solutions obtained by the bi-level decomposition algorithm are identical to the global optimum solutions of the full space model. Specifically, the solutions of the four first cases (200, 500, 1000, and 2000 patients per year) are identical to the global optimum solutions of the full space model. In the fifth case of 3000 patients per year, the proposed algorithm arrives at a significantly improved solution and both subproblems arrive at global optimality. In the last two cases, the full space model was unable to provide a solution in contrast to the bi-level algorithm. For these cases, both the upper and lower levels arrive at global optimality. Based on the above, the first four cases are used to guarantee that the algorithm arrives at global optimality, which has been proven from the full space model. Finally, the matches between the manufacturing facilities are the same across all solutions. Both the full space model and the bi-level algorithm result in the same supply chain network structures.

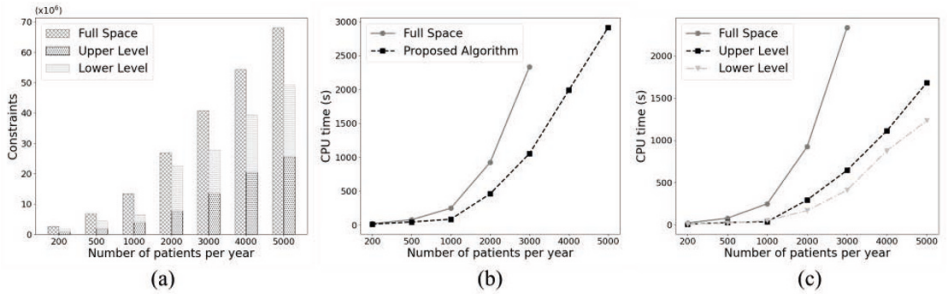


Figure 3. Comparison of the full space model and the bi-level decomposition algorithm for an increasing number of therapies in (a) the number of constraints, (b) the solution time in CPU seconds and (c) the solution time in CPU seconds of the full space model and the upper- and lower-level of the proposed algorithm.

As illustrated in Figure 3a, the proposed approach decreased the total number of constraints in the model. Additionally, the computational time reduced by more than 50% compared to the full space model (Figure 3b). The advantage of the bi-level algorithm is that the two separate subproblems are solved sequentially and thus the computational complexity of the one does not impact the other. This is clearly illustrated in Figure 3c, where the computational time for the upper- and the lower- level is displayed. For example, in the 3000 patients/year case the CPU time of the full space model is 2335 s, while the CPU time of the upper- and lower- level is 644 s and 409 s respectively. The algorithm will stop being efficient only when the two subproblems become computationally intractable. This leads to the conclusion that the bi-level algorithm can provide feasible and optimal solutions for instances higher than 5000 patients/year.

4. Conclusions

In this work, we have addressed the increasing demand in CAR T cell therapies by developing a novel bi-level decomposition algorithm. The proposed algorithm is rigorous and can provide global optimum solutions in reasonable CPU times, even when the full space model provides local or no solutions at all. The original detailed model is decomposed into an upper-level planning problem responsible for strategic planning decisions and a lower-level scheduling problem. As a result, the computational complexity is decreased and solutions for large-scale instances of the problem are

obtained. The computational analysis showed that the proposed algorithm has made significant improvements over the full space model. The bi-level algorithm can provide feasible and optimal solutions for instances of up to 5000 patients/year. Nevertheless, the efficiency of the proposed algorithm creates new possibilities to explore, such as an even higher number of therapies per year or an increased time horizon.

5. Acknowledgements

Funding from the UK Engineering & Physical Sciences Research Council (EPSRC) for the Future Targeted Healthcare Manufacturing Hub hosted at University College London with UK university partners is gratefully acknowledged (Grant Reference: EP/P006485/1). Financial and in-kind support from the consortium of industrial users and sector organisations is also acknowledged.

References

FDA-approved CAR T-cell Therapies | UPMC Hillman. https://hillman.upmc.com/mario-lemieux-center/treatment/car-t-cell-therapy/fda-approved-therapies (accessed Nov. 05, 2021).

Kymriah vs. Yescarta Nucleus Biologics." https://nucleusbiologics.com/resources/kymriah-vs-yescarta/ (accessed Feb. 10, 2021).

A. Bernardi, M. Sarkis, N. Triantafyllou, M. Lakelin, N. Shah, M. M. Papathanasiou. Assessment of intermediate storage and distribution nodes in personalised medicine (in press). Computers & Chemical Engineering. 2021

K. J. Caldwell, S. Gottschalk, A. C. Talleur. Allogeneic CAR Cell Therapy—More Than a Pipe Dream. Front. Immunol. 2021;11:1–12.

M. Erdirik-Dogan, I. E. Grossmann. Simultaneous planning and scheduling of single-stage multi-product continuous plants with parallel lines. Comput. Chem. Eng. 2008;32(11): 2664–2683.

D. Han, Z. Xu, Y. Zhuang, Q Qian. Current progress in CAR-T cell therapy for hematological malignancies. J Cancer. 2021;12(2):326–334.

P. Karakostas, N. Panoskaltsis, A. Mantalaris, and M. C. Georgiadis. Optimization of CAR T-cell therapies supply chains. Comput. Chem. Eng. 2020;139:106913.

M. M. Papathanasiou, C. Stamatis, M. Lakelin, S. Farid, N. Titchener-Hooker, N. Shah. Autologous CAR T-cell therapies supply chain: challenges and opportunities?. Cancer Gene Ther. 2020;27:799–809.

M. Sadelain, R. Brentjens, I. Rivière, J. Park. CD19 CAR Therapy for Acute Lymphoblastic Leukemia. Am. Soc. Clin. Oncol. Educ. B. 2015;35:e360–e363.

M. Sarkis, A. Bernardi, N. Shah, M. M. Papathanasiou. Decision support tools for next-generation vaccines and advanced therapy medicinal products: present and future. Curr. Opin. Chem. Eng. 2021a;32(1):100689.

M. Sarkis, A. Bernardi, N. Shah, M. M. Papathanasiou. Emerging Challenges and Opportunities in Pharmaceutical Manufacturing and Distribution. Processes. 2021b;9(457):267–277.

K. Spink, A. Steinsapir. The long road to affordability: a cost of goods analysis for an autologous CAR-T process. Cell Gene Ther. Insights. 2018;4(11):1105–1116.

S. Terrazas-Moreno, I. E. Grossmann. A multiscale decomposition method for the optimal planning and scheduling of multi-site continuous multiproduct plants. Chem. Eng. Sci. 2011;66(19):4307–4318.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50367-5

An agent-based model for cost-effectiveness analysis in the manufacture of allogeneic human induced pluripotent cells in Japan

Yusuke Hayashi[a*], Kota Oishi[a], Hirokazu Sugiyama[a]

[a]*Department of Chemical system Engineering, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*
y-hayashi@pse.t.u-tokyo.ac.jp

Abstract

This work proposes an agent-based model for cost-effectiveness analysis in the manufacture of allogeneic human induced pluripotent (hiPS) cells in Japan. The agent-based model was developed that can estimate the disability-adjusted life years (DALYs) of each patient and the total required manufacturing cost for allogeneic hiPS cells. The DALYs were defined as the effectiveness indicator, while the total required cost for manufacturing was applied as the cost indicator. Given the disease, the annual number of treated patients, and the treatment mode, the agent-based model can calculate these two indicators. The model was applied to analyze allogeneic hiPS cell therapy for two diseases that are under clinical studies in Japan. A case study demonstrated that the treatment mode would have a significant impact on the cost-effectiveness.

Keywords: Regenerative medicine, Healthcare, Disability-adjusted life year, Therapeutic effect, Kernel density estimation.

1. Introduction

Japan is one of the most advanced countries of research and development on allogeneic human induced pluripotent stem (hiPS) cells. The cells were first produced at Kyoto University (Takahashi et al., 2007), and are one of the most promising sources of regenerative medicine products. Along with successful clinical studies, e.g., Parkinson's disease, implementation of allogeneic hiPS cell therapy is in progress.

Recently in Japan, approval of several products by Pharmaceuticals and Medical Devices Agency (PMDA) received public attention, e.g., nivolumab and tisagenlecleucel. This was because both the high therapeutic effect and the price set by PDMA (e.g., JPY ca. 700,000 for nivolumab in 2014, the year of new introduction). In the national health insurance system of Japan, generally 70 % of the treatment cost is paid by public health system. Further approval of expensive products could lead concerns regarding budget deficits. The products based on allogeneic hiPS cells, once marketed, would also be given ultra-high prices considering the high therapeutic effects. Therefore, the balance between the therapeutic effect and the financial impacts should be considered, in order to pursue sustainability of the healthcare system in Japan. However, detailed investigations considering the individual patient's conditions, e.g., age, have yet to be performed.

Agent-based modeling is useful to deal with complex systems based upon agents, which enables consideration of the individual person's condition. Various topics have been analyzed by agent-based models, e.g., emissions trading considering exchange rates by

Peng et al. (2019), integrated energy systems planning and operation by Zhang et al. (2020), and market acceptance of electric vehicles in China by Huang et al. (2021).

This work proposes an agent-based model for cost-effectiveness analysis in the manufacture of allogeneic hiPS cells in Japan. The agent-based model was developed that can quantify the disability-adjusted life years (DALYs) of each patient and the required manufacturing cost for allogeneic hiPS cells in Japanese society. The DALYs were applied as the effectiveness indicator, while the total required cost was used as the cost indicator. Given the disease, the annual number of treated patients, and the treatment mode, the agent-based model can evaluate these two indicators. The model was applied to analyze allogeneic hiPS cell therapy for two diseases that are in clinical studies in Japan.

2. Methods

2.1. Effectiveness indicators of medical treatments

Generally, there are two indicators used for cost-effectiveness analyses of medical treatments: DALYs (Murray and Lopez, 1997) and quality-adjusted life years (QALYs; Zeckhauser and Shepard, 1976). DALYs indicate overall disease burden expressed as the number of life years lost. The coefficients used for calculating DALYs were provided by World Health Organization (World Health Organization, 2004). On the other hand, QALYs represent overall health condition expressed as the number of life years considering quality of life. The coefficients used for calculating QALYs require to be defined by questionnaires for patients. Allogeneic hiPS cell therapy is still in the middle of clinical studies, and it is quite difficult to perform questionnaires to patients. Hence, in this work DALYs is adopted as the effectiveness indicator.

2.2. Model overview

Figure 1 shows an overview of the developed agent-based model. The overall inputs are defined as the disease, θ [–] (e.g., Parkinson's disease), the annual number of treated patients, $N_{\mathrm{patient}}^{\mathrm{treat}}$ [person year^{-1}], the treatment mode, γ [–] (e.g., from youngest to oldest), and the treatment period, $t_{\mathrm{treat}}^{\mathrm{final}}$ [year]. The outputs are defined as the total DALYs of all patients at the end of the treatment period, $DALY$ [year], and the total required manufacturing cost for allogeneic hiPS cells, C_{total} [JPY]. The output variables of $DALY$ and C_{total} are used as the effectiveness indicator and the cost indicator, respectively.

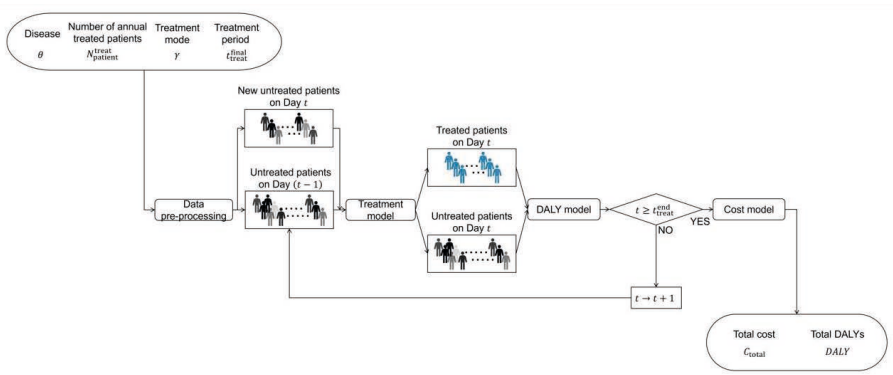


Figure 1. Model overview.

2.3. Data pre-processing

Generally, the data of patient's age is collected as discreate data. For example, the patient's age data in Japan is collected in five-year increments. However, continuous data of patient's age is needed for DALY estimation. Thus, the discrete data of patient's age is converted into continuous data as follows:

$$D_{\mathrm{age}}(Y) = \frac{1}{N_{\mathrm{patient}}^{\mathrm{initial}} h} \sum_{i=1}^{N_{\mathrm{patient}}^{\mathrm{initial}}} K\left(\frac{Y - Y^i}{h}\right) \tag{1}$$

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \tag{2}$$

where D [–] is the distribution, Y [year] is the patient age, $N_{\mathrm{patient}}^{\mathrm{initial}}$ [person] is the initial number of patients, h [–] is the smoothing parameter, and K [–] is the kernel function.

2.4. Treatment model

Figure 2 shows the details of the treatment model. The inputs of the model are defined as the untreated patients on Day $(t-1)$ and the new untreated patients on Day t. The treatment model is classified into reorder and cure models. In the reorder model, all untreated patients on Day t are sorted in the order of treatment. In the cure model, a fixed number of untreated patients based on the annual number of treated patients, $N_{\mathrm{patient}}^{\mathrm{treat}}$, is treated. Moreover, the treatment model is defined, based on the following assumptions.

- For diseases causing sudden death, it occurs with a probability of β [–].
- The effect of treatment is 100%.
- The required time to manufacture allogeneic hiPS cells can be ignored.

2.5. DALY model

The DALYs of a patient i, $DALY^i$ [year], and the total DALYs are estimated as follows (Murray and Lopez, 1997):

$$DALY^i = YLL^i + YLD^i \tag{3}$$

$$YLL^i = \int_{Y_{\mathrm{dead}}^i}^{Y_{\mathrm{life}}} \alpha \, \mathrm{d}Y \tag{4}$$

$$YLD^i = \int_{Y_{\mathrm{sick}}^i}^{Y_{\mathrm{cure}}^i} \alpha DW \, \mathrm{d}Y \tag{5}$$

$$DALY = \sum_{i=1}^{N_{\mathrm{patient}}^{\mathrm{final}}} DALY^i \tag{6}$$

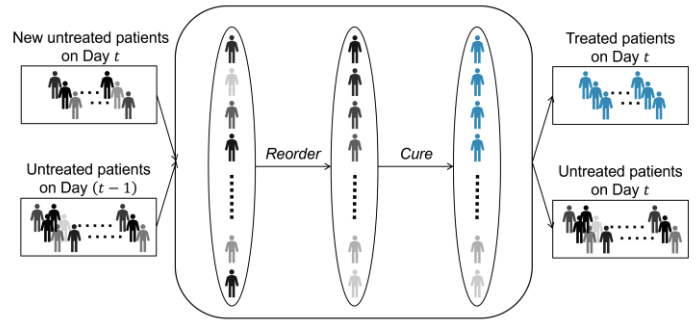


Figure 2. Details of the treatment model.

where YYL [year] is the years of life lost, Y_{life} [year] is the average life span, Y_{dead} [year] is the age of death, α [–] is the disability adjusted life year weight, YLD [year] is the years lost due to disability, Y_{cure} [year] is the age of complete recovery, Y_{sick} [year] is the age of onset, W [–] is the disability weight, and $N_{\mathrm{patient}}^{\mathrm{final}}$ [person] is the final number of all patients including completely cured patients.

2.6. Cost model

The total required manufacturing cost of allogeneic hiPS cells, C_{total}, is calculated using the following equations:

$$C_{\mathrm{total}} = C_{\mathrm{const}} + C_{\mathrm{man}} + C_{\mathrm{trans}} \tag{7}$$

where C_{const} [JPY] is the cost for construction, C_{man} [JPY] is the cost for manufacturing, and C_{trans} [JPY] is the cost for transportation. The value used in the calculation of C_{man} were defined with reference to Sugiyama et al. (2020).

3. Results and discussion

3.1. Data pre-processing

Figure 3 shows the relationship between the patient age, Y, and the probability density, φ [–], for the two diseases in Japan. For both diseases, the parks of the probability density are around 80 years old.

3.2. Case study: Investigation of treatment mode

The impact of the treatment mode on the cost-effectiveness of allogenic hiPS cell therapy for the two diseases was investigated in this section. For the cost-effectiveness analysis, the total required manufacturing cost of allogenic hiPS cells per one DALY, $C_{\mathrm{total}}^{\mathrm{DALY}}$ [JPY year^{-1}], was defined as the evaluation indicator of cost-effectiveness, as shown in the following equations:

$$C_{\mathrm{total}}^{\mathrm{DALY}} = \frac{C_{\mathrm{total}}}{\Delta DALY} \quad \left(N_{\mathrm{patient}}^{\mathrm{treat}} \geq 1\right) \tag{8}$$

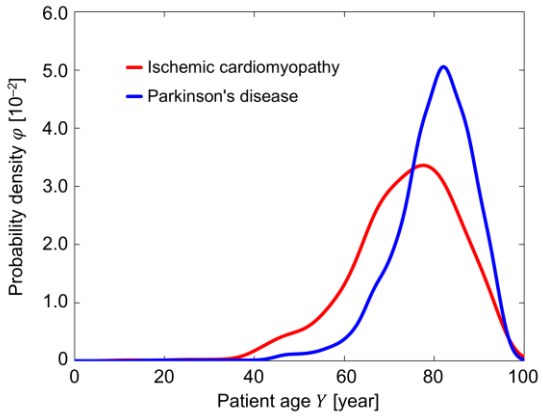


Figure 3. Distribution of the patient age.

$$\Delta DALY = -\{DALY(N_{\text{patient}}^{\text{treat}}) - DALY(0)\} \tag{9}$$

where $\Delta DALY$ [year] is the disability-adjusted life years that can be reduced by the treatment. The optimization problem was formulated as the following equation:

$$\min C_{\text{total}}^{\text{DALY}}(\gamma, N_{\text{patient}}^{\text{treat}}) \tag{10}$$

subject to

$$\begin{array}{lll} \gamma & \in & \{\text{From youngest to oldest, From oldest to youngest}\} \\ N_{\text{patient}}^{\text{treat}} & \leq & 6.0\times10^{3} \\ \theta & \in & \{\text{Ischemic cardiomyopathy, Parkinson's disease}\} \end{array}$$

Figure 4 shows the relationship between $N_{\text{patient}}^{\text{treat}}$, $C_{\text{total}}^{\text{DALY}}$, γ, and θ. For both diseases, in the range where $N_{\text{patient}}^{\text{treat}}$ was small, the values of $C_{\text{total}}^{\text{DALY}}$ differed greatly depending on the treatment mode. On the other hand, in the range where $N_{\text{patient}}^{\text{treat}}$ was large, the

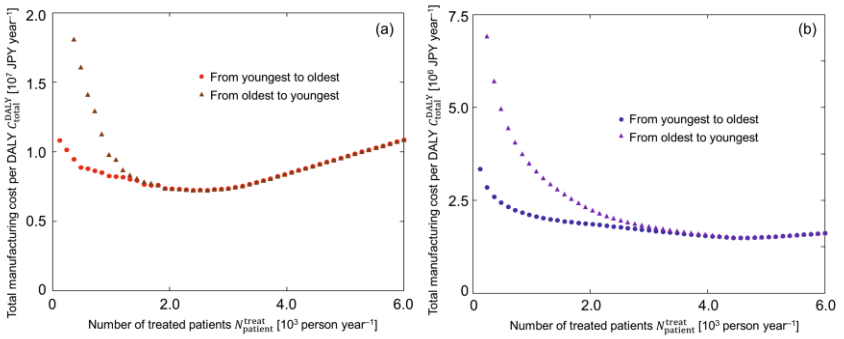


Figure 4. Impact of the treatment mode and the number of treated patients on the cost-effectiveness for (a) ischemic cardiomyopathy and (b) Parkinson's disease.

difference in the values of $C_{\text{total}}^{\text{DALY}}$ was small. Thus, in the range where $N_{\text{patient}}^{\text{treat}}$ was small, the treatment mode needs to be discussed carefully. Moreover, in the range where $N_{\text{patient}}^{\text{treat}}$ is small, $C_{\text{total}}^{\text{DALY}}$ decreases as $N_{\text{patient}}^{\text{treat}}$ increases because the construction cost per one patient decreases. On the other hand, in the range where $N_{\text{patient}}^{\text{treat}}$ is large, $C_{\text{total}}^{\text{DALY}}$ increases as $N_{\text{patient}}^{\text{treat}}$ increases because the overproduction of hiPS cells happens. Hence, careful selection of $N_{\text{patient}}^{\text{treat}}$ should be needed considering the situation of the disease.

4. Conclusions and outlook

In this work, we presented an agent-based model for cost-effectiveness analysis in the manufacture of allogeneic hiPS cells in Japan. The agent-based model was developed that can quantify the DALYs and the total required manufacturing cost of allogeneic hiPS cells. The case study showed that the treatment mode needs to be discussed carefully when the number of annual treated patients is small. In the field of computer-aided process engineering, cell therapy related studies are becoming relevant, e.g., Moschou et al. (2020) and Hayashi et al. (2020). Further model-based studies in this area are encouraged.

Acknowledgements

Discussions with Prof. Masahiro Kino-oka at Osaka University are appreciated. H. S. is thankful for financial support by a Grant-in-Aid for Challenging Research (Exploratory) No. 20K21102 from the Japan Society for the Promotion of Science.

References

Y. Hayashi, I. Horiguchi, M. Kino-oka, and H. Sugiyama, 2020, Slow freezing process design for human induced pluripotent stem cells by modeling intracontainer variation, Comput. Chem. Eng., 132, 106597.

X. Huang, Y. Lin, F. Zhou, M. Lim, and S. Chen, 2021, Agent-based modelling for market acceptance of electric vehicles: Evidence from China, Sustain. Prod. Consum., 28, 206.

D. Moschou, M. Papathanasiou, M. Lakelin, and N. Shah, 2020, Investment planning in personalised medicine, Comput. Aided Chem. Eng., 48, 49.

C. Murray and A. Lopez, 1997, Global mortality, disability, and the contribution of risk factors: global burden of disease study, Lancet, 349, 1436.

Z. Peng, Y. Zhang, G. Shi, and X. Chen, 2019, Cost and effectiveness of emissions trading considering exchange rates based on an agent-based model analysis, J. Clean. Prod., 219, 75.

H. Sugiyama, M. Shiokaramatsu, M. Kagihiro, K. Fukumori, I. Horiguchi, and M. Kino-oka, 2020, Apoptosis-based method for determining lot sizes in the filling of human-induced pluripotent stem cells, J. Tissue Eng. Regen. Med., 14, 1641.

K. Takahashi, K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda, and S. Yamanaka, 2007, Induction of pluripotent stem cells from adult human fibroblasts by defined factors, Cell, 131, 861.

World Health Organization, 2004, Disability weights for diseases and conditions.

R. Zeckhauser and D. Shepard, 1976, Where now for saving lives?, Law Contemp. Probl., 40, 5.

Z. Zhang, R. Jing, J. Lin, X. Wang, K. Dam, M. Wang, C. Meng, S. Xie, and Y. Zhao, 2020, Combining agent-based residential demand modeling with design optimization for integrated energy systems planning and operation, Appl. Energy, 263, 114623.

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50368-7

Design and operation of healthcare facilities using batch-lines: the COVID-19 case in Qatar

Brenno C. Menezes[a*], Mohamed Sawaly[a,b], Mohammed Yaqot[a], Robert E. Franzoi[a], Jeffrey D. Kelly[c]

[a]*Division of Engineering Management and Decision Sciences, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar*
[b]*Department of Supply Chain, Hamad Medical Corporation, Doha, Qatar*
[c]*Industrial Algorithms Ltd., 15 St. Andrews Road, Toronto M1P 4C3, Canada*
bmenezes@hbku.edu.qa

Abstract

In the wake of the COVID-19 pandemic, hospitals worldwide have been overwhelmed and deprived of valuable resources such as bed capacities, medical equipment, personal protection equipment (PPE) stocks, and personnel. These factors imposed unforeseen challenges in the healthcare treatment systems. Mitigating inefficiencies by learning from COVID-19 is necessary to be better prepared to save lives and conserve resources. The main goal of this study is the development of an optimized healthcare treatment network by using predicted epidemiology curves to determine influxes of patients and bed capacities in a hospital facility for both in-patient (IP) wards (oxygen outlets) and intensive care units (ICU). Our model considers flows of patients by distinguishing them in terms of medical severity for their optimal allocation in an existing or installed healthcare facility treated as batch-lines (batch-processes in lines) with time-varying yields of a number of patients per day of treatment. Considering the hospital's admission and discharge of patients from 2020's 1st wave of COVID-19 in Qatar, we determine the bed space availability at any given future date for a hospital facility. This enables the prescription of engineered solutions to increase the capacity, responsiveness, and preparedness of healthcare systems infrastructure and management.

Keywords: Healthcare systems, supply chain resilience, optimization, COVID-19.

1. Introduction

The rapid spread of COVID-19 cases demonstrated the challenges of containing a pandemic whilst providing adequate care (Murthy et al., 2020). Design and operational inefficiencies are among the biggest reasons healthcare systems fail to minimize death rates and spreads of pandemics. Given the inevitable occurrence of future pandemics, healthcare systems must predict the growth and spread of the virus, implement strategies to contain it, and prepare their facilities and resources accordingly.

Several works address predictions on epidemiology curves (Santosh, 2020; Jewell et al., 2020) and their respective effects on resources such as personal protective equipment (PPE) (Tosh et al., 2014). Stübinger and Schneider (2020) propose a forecast of the future COVID-19 spread by addressing identified lead-lag effects using dynamic time warping from batch process monitoring and analysis. Garbey et al. (2020) use data obtained from the French Government during COVID-19 in a computational model to anticipate the patient load of each care unit, and the amount of PPE required by these units, as well as

other key parameters that measure the performance of a healthcare system. Goodarzian et al. (2021) introduce a sustainable-resilience healthcare network for handling COVID-19 pandemic using meta-heuristics for allocation of medicine, resources, and staff throughout the supply chain elements considering capacities and flows among warehouses, distribution centers, pharmacies, hospitals, etc.

The proposition of this work is to develop prescriptive analytics for the optimal healthcare treatment systems in the planning, scheduling, and coordination of the disease treatment networks. With the utilization of the epidemiology curves, decisions can be made to determine optimal bed capacities needed during the COVID-19 pandemic, enabling the design and operation for a resilient medical supply chain to the COVID-19 pandemic.

2. Problem statement

The epidemiology data obtained for this study provide a daily prediction of positive cases from February 1st to May 31st, 2021 in Qatar. From the total number of suspected cases, it is assumed that 1% ends up in to the national healthcare systems' triage facility. We develop a mixed-integer linear programming (MILP) model for 120 days as time-horizon with 1-day time-step, in which 30% of the suspected patients at the triage result in negative diagnostic, and the remaining 70% result in positive. Among the admitted in hospital, 70% of the patients went as an in-patient (IP) and 30% in an intensive care unit (ICU). The model considers actual distributions (in terms of medical severity) of approximately 7,800 patients admitted into a hospital from March 2020 for one year, including the daily inflows and outflows of patients among the facility networks. It also considers the hospital's capacity as 335 IP and 230 ICU beds, with initial occupancies of 30% for each. The field hospital to be opened has a capacity of 160 IP and 80 ICU beds.

For the design and operation optimization of healthcare treatment systems, the network in Figure 1 shows a flowsheet of existing and future facilities and connections constructed in the unit-operation-port-state superstructure (UOPSS) from Kelly (2005) built-in in the Industrial Modeling and Programming Language (IMPL) (Kelly and Menezes, 2019). The shapes are considered as: a) unit-operations m for sources and sinks (◇), tanks or inventories (Δ), batch-processes (□) and b) the connectivity involving arrows (→), inlet-port i (○) and outlet-port j. Unit-operations and arrows are modeled by binary y and continuous x variables and the ports as yields of patients.

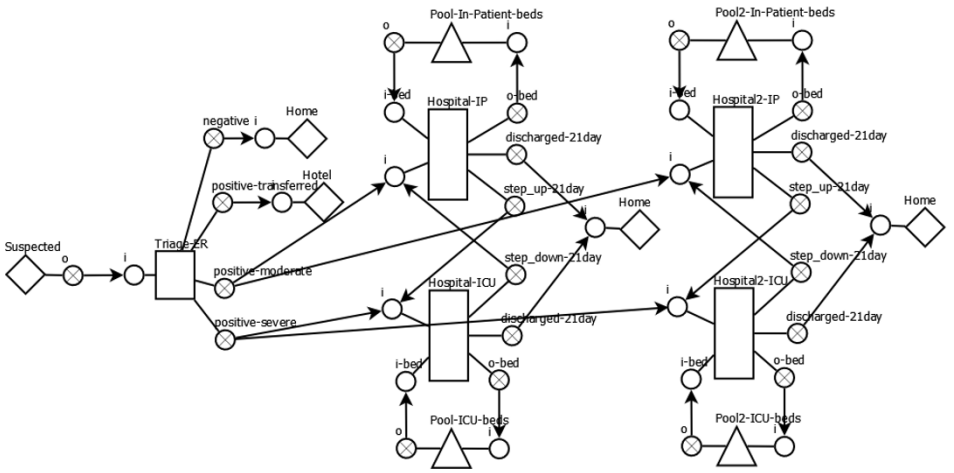


Figure 1. Base flowsheet of the healthcare system.

The model includes predicted yields of patient step-downs (from ICU to IP – meaning the patients' medical status improved), patient step-ups (from IP to ICU – meaning the patients' medical status deteriorated), patient admission yields, patient discharge yields, death yields, and transfer yields. The maximum batch-time is 21 days (which captures approximately 85 to 90% of hospitalized patients' lengths of stay), although there are time-varying yields from the actual distribution of patients in- and out-fluxes both IP and ICU wards from the observed cases. The 1- to 20-days yields and their connections are not represented in Figure 1 for simplicity.

3. Mathematical modeling

The objective function in Eq.(1) maximizes the pre-treatment of the suspected cases in the triage emergency room (ER), where $x_{j,i,t}$ represent number of cases for flows from the outlet port set j to inlet set i at time t. The variable $xh_{m,t}$ defines batch-processes' or hospital-units' holdups and pools of bed capacity in the model. All flows and holdups are governed by semi-continuous constraints of the shapes to themselves, such as $\bar{x}^L_{j,i,t}\, y_{j,i,t} \leq x_{j,i,t} \leq \bar{x}^U_{j,i,t}\, y_{j,i,t}\ \forall\ (j,i) \in JI, t$. The sets I and J represent in- and out-ports, respectively, while the set JI defines connecting patient flows between out- and in-ports. For the batch-processes (triage and hospitals), the holdup $xh_{m,t}$ is taken when they are starting up ($zu_{m,t}$=1) constrained by the respective bounds of the hospital facility capacities. The UOPSS formulation in Eq.(2) establishes that the holdup or inventory level bounds ($\overline{xh}^L_{m,t}$ and $\overline{xh}^U_{m,t}$) of the hospital facilities respect the sum of the flows arriving in and leaving from ports (in- and out-ports) whenever the respective startup variable $zsu_{m,t}$ is active. The sets M_{BATCH} include triage-ER, IP, and ICU facilities and M_{POOL} the IP/ICU bed's pools. In the indices in the summations from Eq.(1) to (5), the subsets of the I, J, and JI follow the flowsheet in Figure 1. For $x_{j,i,t}, xh_{m,t} \geq 0;\ y_{j,i,t}, y_{m,t} = \{0,1\};\ zsu_{m,t} = (0,1)$:

$$Max\ Z = \sum_t \sum_{JI_{Suspected}} x_{j,i,t} \tag{1}$$

$$\overline{xh}^L_{m,t}\, zsu_{m,t} \leq \sum_{i \in I} x_{j,i,t} \leq \overline{xh}^U_{m,t}\, zsu_{m,t} \quad \forall\ (m,j) \in M_{BATCH}, t \tag{2}$$

$$\sum_{j \in J_{up}} x_{j,i,t} = xh_{m,t} \quad \forall\ (i,m) \in M_{BATCH}, t \tag{3}$$

$$x_{j,i \in I_{do},t+delay} = \bar{r}_{j,t+delay}\, xh_{m,t} \quad \forall\ (m,j) \in M_{BATCH}, t \tag{4}$$

$$xh_{m,t} = xh_{m,t-1} + \sum_{j_{up} \in J} x_{j_{up},i,t} - \sum_{i_{do} \in I} x_{j,i_{do},t} \quad \forall\ (i,m,j) \in M_{POOL}, t \tag{5}$$

$$y_{m_{up},t} + y_{m,t} \geq 2y_{j_{up},i,t} \quad \forall\ \left(m_{up}, j_{up}, i, m\right), t \tag{6}$$

$$\sum_{tt<t} zsu_{m',tt} + y_{m,t} \leq y_{j,i,t} \quad \forall\ (m',j,i,m), t \tag{7}$$

$$y_{m,t} - y_{m,t-1} - zsu_{m,t} + zsd_{m,t} = 0 \quad \forall\ m \in M_{BATCH}, t \tag{8}$$

$$y_{m,t} + y_{m,t-1} - zsu_{m,t} - zsd_{m,t} - 2zsw_{m,t} = 0\ \forall\ m \in M_{BATCH}\ , t \tag{9}$$

$$zsu_{m,t} + zsd_{m,t} + zsw_{m,t} \leq 1\ \forall\ m \in M_{BATCH}, t \tag{10}$$

Equations (3) and (4) are related to the modeling of hospitals as batch-processes and in a special case called batch-lines. This is applied in Menezes et al. (2020) for livestock planning to determine the initial procreation of the animal batches, in which there is no accumulation of amounts of batches at each time step, as in Eq.(3). Instead, balances of batch amounts at a single time-window and the delaying and yield of amounts leaving the facility are modeled, as in Eq.(4). In the hospital facilities, new batches of patients arrive in the Triage-ER, Hospital IP, and Hospital ICU every day. The unit-operations' inventory or holdup quantity balance of pools are determined in Eq.(5) for both IP and ICU bed capacities. These constraints manage the availability of beds (holdup) to be utilized in the hospitals by controlling the a) inlet flow, when patients are dispatched outside the system or change their status to step-up or step-down; and b) the outlet flow, when the beds are needed in the hospital facilities.

Equations (5) and (6) represent the constraints for the structural transitions that allow the setup $y_{m,t}$ or startup $zsu_{m,t}$ of connected out-port-states j and in-port-states i unit-operations. When the setup of unit-operations m and m' is equal to the unitary in Eq.(6), by implication, the setup variable the arrow stream $y_{j,i,t}$ between the neighbor unit-operations must be true. In Eq.(7), addressing the hospital facilities as batch-processes, as the setup variable of m' is changed by the summation of the startups. These logic valid cuts reduce the tree search in branch-and-bound methods. The temporal transition in Equations (8) and (9) control the operations for semi-continuous blenders from Kelly and Zyngier (2007). The binary variable $y_{m,t}$ manages the start-up ($zsu_{m,t}$) switch-over ($zsw_{m,t}$) and shut-down variables ($zsd_{m,t}$), which are relaxed in the interval [0,1]. Equation (10) guarantees the integrality of the relaxed variables.

4. Results

The optimization for the proposed MILP in Figure 1 for 120 days as time-horizon with 1-day time-step is solved in 63 seconds with GUROBI 9.1.1 and 256 seconds with CPLEX 20.1.0 both at 1.0% of MILP relaxation gap using an Intel Core i7 machine at 3.4 GHz (8 threads) with 64 GB of RAM. There are 68,721 constraints for 23,765 continuous variables and 15,248 binary variables in the problem. The results in Figure 2 show that the existing facility's IP capacity could not sustain the surge of patients caused by the new strain of COVID-19 (initiated on day 63), which triggered the opening of the new field hospital by day 85.

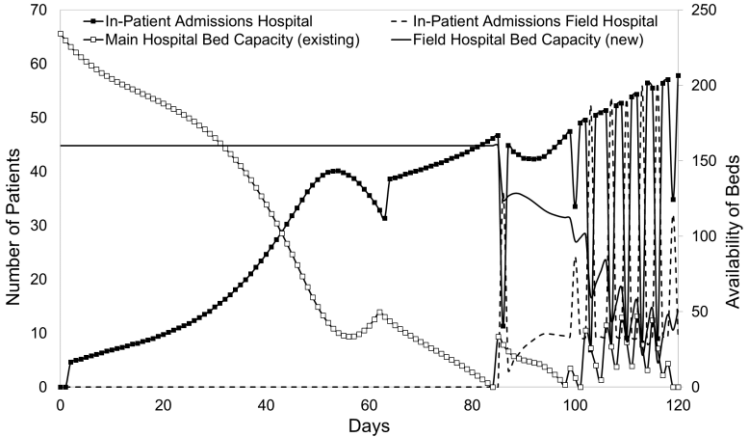


Figure 2. Design and operation of the IP facility.

Based on Figure 2, the existing hospital was only able to sustain the increase in patients for exactly 21 days (day 84), where the number of In-Patient admissions was 47 patients per day. From day 85 onwards, both facilities simultaneously received patients, which relieved the pressure on the existing facility and made the design feasible. Days 98, 101, and 119 demonstrate how the new field hospital's capacity helped sustain the operation when the main existing facility's capacity depleted. By the end of the time horizon, there were still 51 vacant IP beds in the field hospital (i.e., new facility). The lines in the plot are symmetrical as each facility worked hand in hand to handle the influx of patients.

Figure 3 demonstrates how the opening of 80 more ICU beds aided in sustaining the hospitalization of patients on days 83, 92, and onwards. The phenomenon observed from days 92 to 107, where the capacity of the field hospital remains zero, is illustrated in Figure 4.

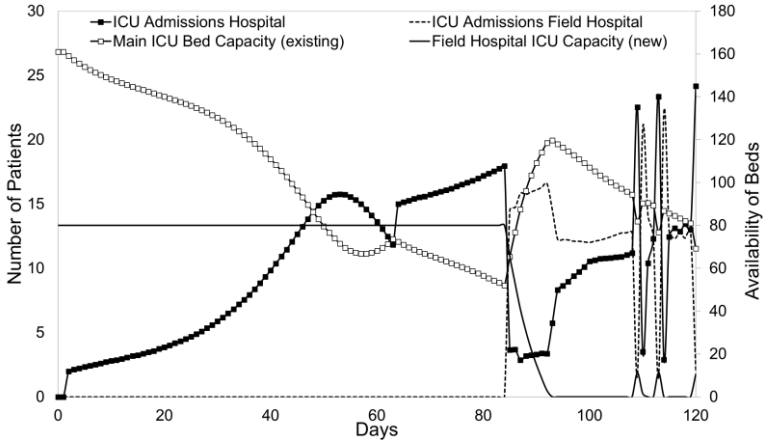


Figure 3. Design and operation of the ICU facility.

Figure 4 explains why the Field Hospital ICU capacity curve remains flat at zero from day 92 to 107, in which the capacity pool freeing up (by people being discharged) is occupied by other patients at the same rate. This phenomenon shows extreme efficiency since utilizes the limit opened capacity available, although it does not consider real-world factors such as the disinfection or preparation of ICU space.

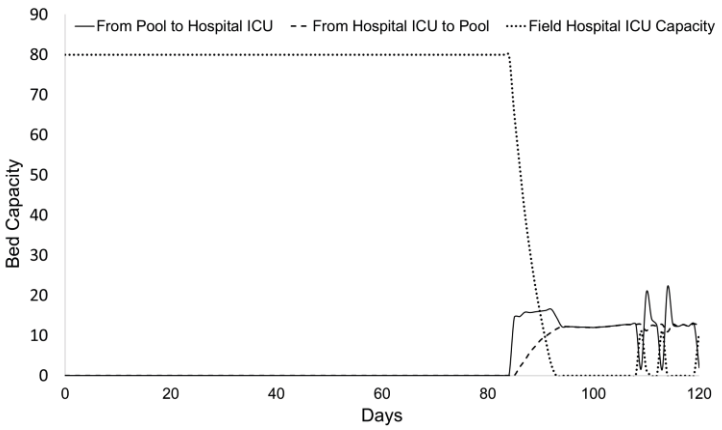


Figure 4. Pool of patients in and out causing the capacity to remain flat at zero.

5. Conclusions

It was evident from the events of COVID-19 that our world is interconnected in a way that virus outbreaks in a region can easily spread and cause impacts in a global sphere. Protecting the lives of humans entails that we must have a proper number of resources allocated in a timely and efficient manner. This work demonstrates how to be better prepared by designing and operating a healthcare treatment network with different facilities as the triage-ER, in-patient (IP), and intensive care unit (ICU) wards. The predicted epidemiology curves and the time-varying yields of the distribution of patients throughout the network served as inputs for the modeling and solving of batch-lines of patients interconnected among the facilities and their outlets. With such proposition, availability of bed capacities has been determined along the time-horizon, and installations of new field facilities for IP and ICU were necessary to handle the increased number of moderate and severe patients. Future work can implement procurement planning to ensure continuous availability of PPE and medical equipment; model staff scheduling models to ensure that no sick person is left unattended; design entire health system networks to ensure fully and optimal utilization of bed space; provide better utilization of quarantine and hotel facilities; and develop more accurate epidemiology curves to provide more reliable predictions for the potential of strains of viruses in further pandemic events.

References

S. Murthy, C. D. Gomersall, R. A. Fowler, 2020, Care for critically ill patients with COVID-19. Journal of the American Medical Association 323(15), 1499–1500.

M. Garbey, G. Joerger, S. Furr, V. Fikfak, 2020, A model of workflow in the hospital during a pandemic to assist management. PLOS ONE, 15(11), e0242183.

F. Goodarzian, P. Ghasemi, A. Gunasekaren, A.A. Taleizadeh, A. Abraham, 2021, A sustainable-resilience healthcare network for handling COVID-19 pandemic, Annals of Operations Research.

N. Jewell, J. Lewnard, B. Jewell, 2020, Caution warranted: using the Institute for Health Metrics and Evaluation model for predicting the course of the COVID-19 pandemic. Ann Intern Med, Epub ahead of print.

J.D. Kelly, 2005, The Unit-Operation-Stock Superstructure (UOSS) and the Quantity-Logic-Quality Paradigm (QLQP) for Production Scheduling in The Process Industries, In Multidisciplinary International Scheduling Conference Proceedings: New York, United States, 327-333.

J.D. Kelly, D. Zyngier, 2007, An improved MILP modeling of sequence-dependent switchovers for discrete-time scheduling problems, Industrial and Engineering Chemistry Research, 46, 4964-4973.

J.D. Kelly, B.C. Menezes, 2019, Industrial Modeling and Programming Language (IMPL) for Off- and On-Line Optimization and Estimation Applications. In: Fathi M., Khakifirooz M., Pardalos P. (eds) Optimization in Large Scale Problems. Springer Optimization and Its Applications, 152, 75-96.

B.C. Menezes, J.D. Kelly, T. Al-Ansari, 2020, Livestock production planning with batch-lines in the agriculture industry, Computer Aided Chemical Engineering, 48, 465-470.

K. Santosh, 2020, COVID-19 Prediction Models and Unexploited Data. Journal of Medical Systems, 44(9), 170.

J. Stubinger, L. Schneider, 2020, Epidemiology of coronavirus COVID-19: Forecasting the future incidence in different countries. Healthcare 8(2), 99.

P.K. Tosh, H. Feldman, M. D. Christian, 2014, Business and continuity of operations: Care of the critically ill and injured during pandemics and disasters: CHEST Consensus Statement. Chest, 146(4).

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50369-9

Application of PSE Methods on Monoclonal Antibody Productivity Improvement and Quality Control

Ou Yang[a], and Marianthi Ierapetritou[b*]

[a] *Department of Chemical and Biochemical Engineering, Rutger, the State University of New Jersey, Piscataway, NJ, 08854, USA*
[b] *Department of Chemical Engineering and Biomolecular Engineering, University of Delaware, Newark, DE, 19716, USA*
mgi@udel.edu

Abstract

With a high demand for monoclonal Antibodies (mAbs) in the current biopharmaceutical market, there is a need to improve overall process productivity while maintaining product quality. This work introduces three case studies that apply process systems engineering (PSE) methods including flowsheet modeling, mechanistic modeling, and process optimization to provide strategies on production mode selection, process prediction, and operating design space determination. These approaches can be utilized in process development and ultimately improve the productivity during biopharmaceutical manufacturing.

Keywords: Process simulation; Dynamic modeling; Biopharmaceutical manufacturing

1. Introduction

MAbs is one of the most promising therapeutic products with its wide applications in cancers, infections, and autoimmune disorders treatment. It has been reported that the growth rate of mAb market ranges between 7.2 and 18.3% since 2016 and would reach $130-200 billion in the year 2022 (Grilo and Mantalaris, 2019). Most of mAbs belong to IgG class which contains two regions: antigen-binding (Fab) and crystallizable regions (Fc). Protein glycosylation in Fc region affects protein stability and efficacy (Xu et al. 2011, Zheng et al., 2011). Thus, maintaining operating conditions within a feasible region to ensure the required level of glycosylation is also critical and challenging in mAb production. Process systems engineering (PSE) methods such as process simulation, sensitivity analysis, and process optimization allow early-stage process design, system analysis, and process improvement, which has great potential to address the above challenges.

In this work, flowsheet modeling is used as a process decision-making tool to evaluate cost-effectiveness of fully integrated continuous operation over the conventional batch operation. Then mechanistic and surrogate models are used to capture nonlinear bioreactor dynamics under different operating conditions. The models are able to correlate operating parameters with productivity and quality. Furthermore, feasibility analysis is applied to process model to determine the design space for desired protein production and drug quality.

2. Flowsheet modeling and techno-economic analysis

In this case study, fed-batch and continuous mAb production lines are designed to evaluate the benefits of continuous operations. Different analysis approaches including deterministic cost analysis, and sensitivity analysis are used to discover the possibilities and challenges of continuous applications in biopharmaceutical manufacturing.

2.1. Process Description

Biopharmaceutical manufacturing for mAbs production includes inoculation, cell culture protein production, clarification, primary capture, polishing, and final formulation steps. The inoculation is a series of cell culture passages that contain test tubes, T flasks, shake flasks, and seed culture bioreactor. Fed-batch bioreactor and perfusion bioreactor are used as production bioreactors for batch process and continuous process, respectively. After production in batch operating mode, proteins are purified by centrifuge followed by microfilters, dead-end filters, Protein A chromatography, acid-based virus inactivation tanks, AEX chromatography, nano filter, ultrafiltration, and diafiltration. For continuous process, tangential filtration is used to harvest protein and send it to filters and periodic counter-current (PCC) protein A column for primary capture. Two virus inactivation tanks and AEX chromatography work alternatively to achieve continuous virus inactivation. For the final formulation staged single-pass tangential flow filtration and counter-current staged diafiltration are used for buffer exchange and protein concentration.

2.2. Simulation software and techno-economic analysis

SuperPro Designer (Intelligen, Scotch Plains, NJ) is a recipe-driven simulator that can be used to simulate both batch and continuous biopharmaceutical manufacturing. Material balances and process scheduling can be captured through the whole process. Process analysis including economic analysis with breakdown cost categories, throughput analysis, and sensitivity analysis are used to evaluate the two operating modes under different scales and process operating parameters.

Economic analysis is performed using SuperPro Designer with customer inputted equipment and material costs. The costs are referenced from literature and online vendor resources. The calculation includes capital investment and operating costs. The capital investment is based on equipment expenditure. The installation, piping, insulation costs are calculated based on the equipment cost. The operating cost contains material, consumables, utilities, labor-depended, quality control and quality assurance, and facility cost. Cost of goods per gram (COG/g) is used to represent cost per unit of production.

2.3. Results

Two integrated lines, one with batch operating mode and another with continuous operating mode are simulated. The base case scenario is adjusted to 620 kg/yr production rate and the selling price of mAb is assumed at $20/mg. The result shows that upstream takes the highest percent of the overall cost in both operating modes. Comparing batch and continuous processes, the overall capital cost in the fed-batch process is $165 million, which is 2 times higher than that of the continuous process. The operating cost of fed-batch and continuous processes are $61 million/yr, and $32 million/yr, respectively.

Figure 1 clearly shows that the main benefit of the continuous process stems from the capital cost investment and facility dependent cost, mainly due to smaller footprint of the continuous process. The consumables cost is also reduced due to resin cost savings in the PCC process.

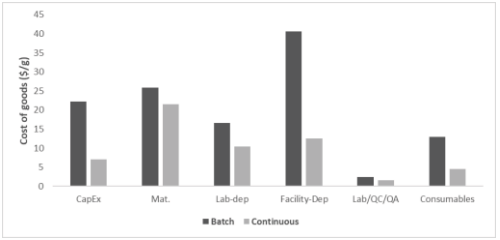


Figure 1 Cost of goods analysis between fed-batch and continuous

Sensitivity analysis is used to further investigate the impact of throughput and upstream parameters on the cost-effectiveness of the two operating modes. The manufacturing scale changes from 50 kg/yr to 1200 kg/yr as shown in Figure 2. Results show a decreased trend of unit operating cost with plant capacity changes, which is mainly attributed to the cost savings of labor dependent cost. The continuous process is more cost-effective than the batch process through the whole range of capacity. Similar results are also found whe titer changes as shown in Figure 3, where titer varies from 1.5 g/L to 5.5 g/L.

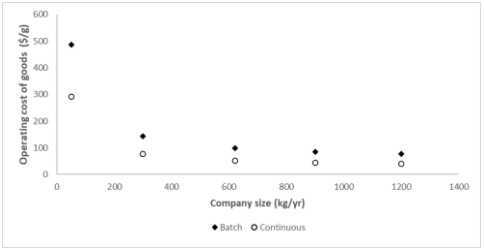


Figure 2 Manufacturing scales change vs. COG/g

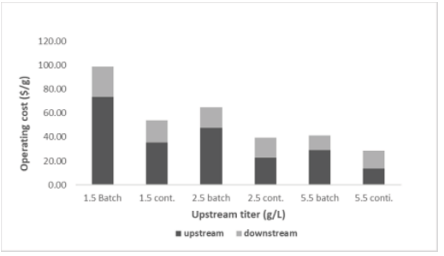


Figure 3 Upstream titer vs operating cost

3. Predictive modeling of cell culture and protein glycosylation processes

From the previous case study, it has been found that the upstream bioreactor takes the highest percentage of the overall cost and is usually identified as a bottleneck of the overall process production. In this section experimental data are used to develop a mechanistic model to understand cell culture and protein glycosylation process in order to improve production and achieve target product quality.

3.1. Background

In mAbs production, the protein glycosylation process is a post-translation modification process that affects product potency and efficacy. Operating conditions such as temperature, pH, and metabolites concentrations all affect glycosylation. To understand the effect of temperature on the glycosylation process, Chinese hamster ovary (CHO)-K1 cell is first cultured under 37 ℃ and switched to 35 ℃, 37 ℃ and 39 ℃ on day 4. Different metabolite concentrations, protein titer and glycan fractions are measured at different time points and used to train a mechanistic model.

3.2. Equations

The mechanistic model includes two parts, the cell unstructured model and the structured glycosylation model. The unstructured model considers cells as a black box and only captures the mass balance of critical components in a bioreactor. Equation (1) represents the mass balance of cell density.

$$\frac{d(V[X_v])}{dt} = F_{in}\,[X_{v0}\,] + \mu V[X_v] - \mu_d V[X_v] - F_{out}X_v \quad (1)$$

where X_v is cell density, F_{in}, F_{out} represent feed addition and sampling during the cell culture; μ is cell growth rate, which can be represented by empirical equations correlated to other metabolites/nutrients in the solution; and μ_d is cell death rate. The single cell model considers Golgi Apparatus, where major glycosylation reaction happens, as a plug flow reactor and the mass balance as shown in equation (2).

$$\frac{\partial[G_m]}{\partial t} = -V_1\frac{\partial[G_m]}{\partial z} + \sum_n^{Enzyme} v_{m,n}r_n \quad (2)$$

where G_m is glycan fractions; r_n represents kinetic rate for enzyme reaction n; $v_{m,n}$ is the reaction coefficient of glycan m that catalyzed by enzyme n; V_1 represents the linear velocity that protein glycan transfers through the Golgi apparatus.

3.3. Results

Least-squares parameter estimation is used to fit the model to experimental data. Experimental data and simulation fitting results for product titer and G0 fraction (one of the glycan fractions) under different temperatures are shown in Figure 4, and 5. The result shows that the mechanistic model is able to capture the general trend of both titer and G0 fraction and the relative trends under different temperature can also be obtained. Cell density, glucose, ammonia concentrations and other glycan fractions can also be captured.

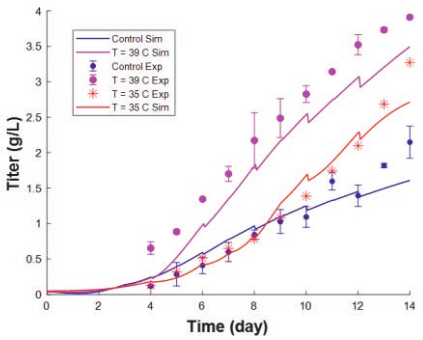


Figure 4 Experimental data and simulation fitting results for titer

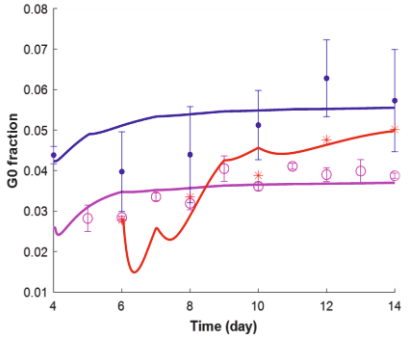


Figure 5 Experimental data and simulation fitting results for G0 fraction

4. Fed-batch bioreactor modeling and design space identification

In this case study, dynamic kriging together with a surrogate-based adaptive sampling approach is used to capture the effects of temperature and pH on productivity (product titer) and product quality (glycan fraction in glycosylation process) and furthermore determine the design space for upstream bioreactor operation (Yang and Ierapetritou 2021).

4.1. Kriging and dynamic kriging

As an interpolation method, kriging equation is shown in Equation (3).

$$\hat{f}(\mathrm{x}^{\mathrm{i}}) = \beta f(\mathrm{x}^{\mathrm{i}}) + \varepsilon(\mathrm{x}^{\mathrm{i}}) \quad (3)$$

$\beta\mathrm{f}(\mathrm{x}^{\mathrm{i}})$ indicates a known regression model that defines the global trend of the data $f(\mathrm{x}^{\mathrm{i}})$, β is unknown parameter; $\varepsilon(\mathrm{x}^{\mathrm{i}})$ is a residual term that represents the error at location x^{i}

which is usually normally distributed with zero mean and variance σ^2. Dynamic kriging is a modification of kriging model as shown in equation (4).

$$\hat{f}(\mathrm{x}_k^{\mathrm{i}}) = \beta f(\mathrm{x}_k^{\mathrm{i}}, \hat{f}(\mathrm{x}_{k-1}^{\mathrm{i}})\ ) + \varepsilon(\mathrm{x}^{\mathrm{i}}, \hat{f}(\mathrm{x}_{k-1}^{\mathrm{i}})) \quad (4)$$

The dynamic system is first discretized into different time points k, and the kriging model is used as an autoregressive model that collects the predicted results $\hat{f}(\mathrm{x}_{k-1}^{\mathrm{i}})$ from the previous time point (k-1) and combines with the state variables or control input $\mathrm{x}_k^{\mathrm{i}}$ to estimate the future time point $\hat{f}(\mathrm{x}_k^{\mathrm{i}})$(Hernandez and Grover 2010). In this work, both models are built using DACE toolbox in MATLAB.

4.2. Feasibility analysis

Feasibility function is defined in equation (5).

$$\varphi(x) = \max_{j \in J} \mathrm{g}_{\mathrm{i}}(x) \quad (5)$$

where $\mathrm{g}_{\mathrm{i}}(x)$ represent different constraints including productivity and product quality, x includes temperature, pH and other operating parameters which are total operating time, initial conditions of cell density, glucose and mAb concentrations. Initial sample points can be generated by space filing sampling and a feasible region ($\varphi(x)$<0) can be obtained by calculating the feasibility function as shown in Equation (5). Adaptive sampling method is used to improve the accuracy of the feasible boundary by maximizing the modified EI function, shown in equation (6). The new sample points that are close to the boundary of the feasible region are used to update the kriging model.

$$EI_{feas}(x) = \hat{s}(x)\phi\left(-\frac{\hat{y}(x)}{\hat{s}(x)}\right) = \hat{s}(x)\frac{1}{\sqrt{2\pi}}e^{-0.5(\frac{\hat{y}(x)^2}{\hat{s}(x)^2})} \quad (6)$$

Standard error $\hat{s}(x)$ can be obtained from kriging prediction at location x, and $\hat{y}(x)$ is the predicted value. The detailed explanation of the modified EI function can be found in (Boukouvala and Ierapetritou 2014).

4.3. Results

The mechanistic model for mAbs production outlined in Section 3 is first used to generate training datasets to build kriging models. A two-level full factorial design is applied to generate data under different temperature and pH. Viable cell density, glucose concentration, protein titer and glycan fractions are obtained. The prediction from dynamic kriging and regular kriging are compared in Figure 6. Dynamic kriging provides higher prediction accuracy than regular kriging, because dynamic kriging considers more sample points (the previous time points) and correlations during the model prediction.

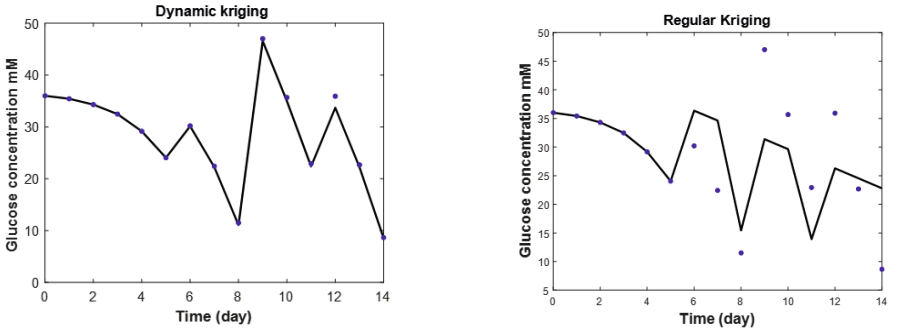


Figure 6 Comparison between dynamic kriging and regular kriging

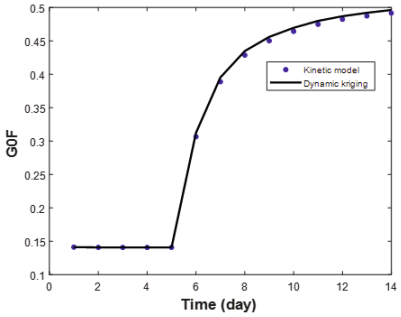


Figure 7 Prediction of glycan fraction using dynamic kriging

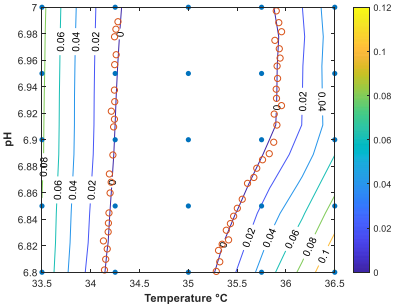


Figure 8 Design space obtained from adaptive sampling.

Input parameters (temperature and pH) together with viable cell concentration, protein titers, and glycan fractions at (t−1) time points are used to predict the glycan fractions at time t. Figure 7 provides the prediction of one of the glycan fractions from dynamic kriging which shows that dynamic kriging is able to predict the dynamic trend of glycan fractions with high accuracy. Feasibility analysis is used to ensure high product titer while maintaining glycan fractions within the required range. Figure 8 shows the contour plot for feasibility function values under different pH and temperatures. The feasible operating region is demonstrated inside the zero line.

5. Conclusions

This work shows the application of process modeling and system analysis methods on the improvement of biopharmaceutical manufacturing process. Three approaches are provided, including flowsheet simulation and single unit operation modeling and optimization. Flowsheet modeling is applied to process design and integrated process evaluation. Mechanistic modeling is used to correlate process parameters with cell growth and critical product quality attributes. A framework is built to improve process understanding as well as finding an optimum design space to satisfy required productivity and quality. For the future work, the mechanic model can be integrated to the flowsheet model to obtain an end-to-end biopharmaceutical manufacturing simulation system. Operation of single unit and integrated line can be optimized by adaptive sampling and surrogate based optimization. Validating and training the model with real data would further improve the robustness and reliability of the model.

References

F. Boukouvala, M. Ierapetritou, 2014. Derivative-free optimization for expensive constrained problems using a novel expected improvement objective function. AIChE Journal **60**(7): 2462-2474.

X. Xu, H. Nagarajan, S. Lewis, et al. 2011, The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nat Biotechnol. 29 (8), 735-741

A. Grilo, A. Mantalaris, 2019, The Increasingly Human and Profitable Monoclonal Antibody Market. Trends Biotechnol. 37, 9–16

K. Zheng, C. Bantog, R. Bayer, 2011, The Impact of Glycoyslation on Monoclonal Antibody Conformation and Stability, MAbs, 3, 568-576

O. Yang, M. Ierapetritou, 2021, mAb Production Modeling and Design Space Evaluation Including Glycosylation Process. Processes

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19-23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50370-5

Image classification of experimental yields for cardiomyocyte cells differentiated from human induced pluripotent stem cells

Samira Mohammadi, Ferdous Finklea, Mohammadjafar Hashemi, Elizabeth Lipke, and Selen Cremaschi*

Department of Chemical Engineering, Auburn University, Auburn, Alabama, USA
selen-cremaschi@auburn.edu

Abstract

Cardiovascular diseases (CVDs) are the number one cause of death worldwide. Mass production of engineered heart tissue using differentiation of human-induced pluripotent stem cells (hiPSCs) can substitute a large number of the lost heart muscle cells in patients with CVDs. However, the scale-up of the differentiation systems for heart tissue, i.e., cardiomyocyte (CM), production is challenging because many parameters affect the process. Machine learning (ML) techniques can be employed to identify critical process parameters for differentiation systems and build models to elucidate the impact of these parameters on process outcomes. Here, we present a ML model to predict CM content on day 10 of the differentiation. Phase-contrast images of microspheroid tissues on differentiation day 5 are the inputs of the ML model, and the output is CM content on 10 of differentiation, classified as either sufficient and insufficient. Support vector machines are used as the classifier models. We utilized feature extraction and selection methods. The best classifier had an accuracy of 77% in predicting the sufficient CM content class.

Keywords: cardiac differentiation, machine learning, support vector machines

1. Introduction

Heart muscle cells (cardiomyocytes (CMs)) are one of the least regenerative cells in the body. Cardiovascular diseases (CVDs) can lead to heart failure and loss of in the order of billion CMs (Kempf et al., 2016). Few viable treatments are present for patients with CVD and post-heart attack problems. Production of CMs via differentiation from human-induced pluripotent stem cells (hiPSCs) may contribute to developing and testing therapeutics for CVDs, e.g., in fields such as drug monitoring and cell therapy (Denning et al., 2016). Mass production of CMs and their implementation in cell therapy of CVD patients is another potential application of hiPSC-derived CMs (hiPSC-CMs).

The production of CMs by differentiation of hiPSCs in a 3D platform is a complex, expensive process, and a high number of parameters impact the system performance (Gaspari et al., 2018). The 3D platforms are promising for the scale-up of CM production, and identifying critical process parameters and their optimal ranges for 3D platforms is the first step towards scale-up. More specifically, distinguishing an unsuccessful batch from a successful one at an earlier time point of the differentiation would significantly reduce the expense and time required for CM production.

In recent years, machine learning (ML) techniques have been successfully used to study complex systems where fundamental understanding is limited. These techniques use

information from data sets to infer the relationships between process parameters (inputs) and outcomes (outputs). With the progress in ML algorithms and computational power, many studies exploited the information contained in images to build models to study different systems, such as quantification of CM contraction using image correlation analysis (Kamgoué et al., 2009) and plant disease detection (Vishnoi et al., 2021).

This study investigates the ability to classify CM content on day 10 of hiPSC-laden microspheroid differentiation using images taken on day 5. The CM content is defined as the percentage of the cells which are CMs on the specific differentiation day. We hypothesize that the phase-contrast images of the cells taken during differentiation include information regarding differentiation progress and that a classifier model can capture this information to distinguish batches with sufficient CM content from those with insufficient. Support vector machines are trained using different extracted feature sets of the phase-contrast images to predict the CM content class. The best model had an accuracy of 77% and an MCC of 0.53.

2. Methods and Materials

2.1. Experiments

HiPSCs were encapsulated within PEG-fibrinogen (PF) by using a novel microfluidic system (Tian and Lipke, 2020) in microspheroids with different sizes and axial ratios (AR). After culturing the hiPSC-laden microspheroids in E8 or mTeSR-1 media for 3 days, the CM differentiation is carried out by supplemented CDM3 or RPMI/B27 minus insulin with CHIR on day 0 and IWP2 on days 1 and 3, respectively. Fresh CDM3 was added on days 3, 5, 7, and fresh RPMI/B27 minus insulin media was added on days 1 and 5. Following day 7 or 10, the microspheroids were cultured with RPMI/B27 (Gibco), and the media was exchanged every 3–4 days. (Figure 1). Phase-contrast images were taken throughout the differentiation timeline on days 0, 1, 3, and 5, shown in Figure 1.

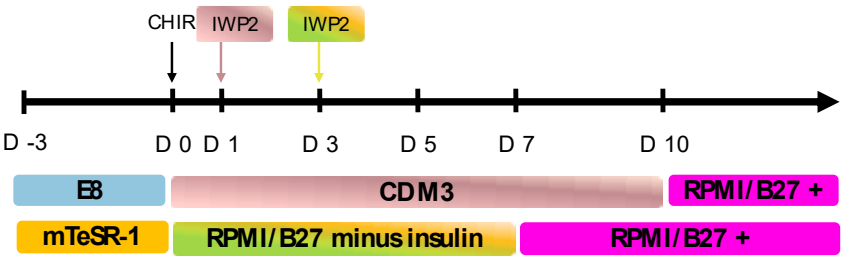


Figure 1. Differentiation protocol of hiPSC-laden microspheroids

2.2. Data Used to Build the Classifier Model

The initial training data set included 301 phase-contrast images, from day 5 of differentiation, with their corresponding CM content on day 10. Images on day 5 were used because day 5 is the earliest time point without any external stimuli or changes to the system with image availability. Each image contained 496 × 658 pixels. Figure 2 shows two representative images. The images were augmented to increase the number of training data points to improve the model's generalization. Each image was flipped and rotated (180°), increasing the number to 903.

CM content above 70% on the 10th differentiation day was defined as the *Sufficient* class, and batched with CM content below 70% belonged to the *Insufficient* class. The data was

split into test and train sets using 20% and 80% of data ratios, respectively. Different classifier models were compared based on their performance on the test set.

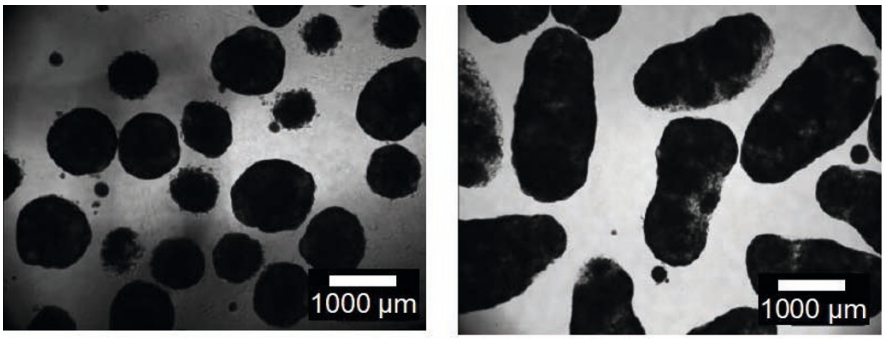


Figure 2. Representative phase-contrast images of microspheroids on day 5

2.3. Feature Extraction

The RGB color space features (color features) of the image pixels formed the initial input feature set. We used two techniques to extract additional features from the images, Histogram of Oriented Gradients (HOGs) (Freeman and Roth, 1994) and texture transformations (Haralick et al., 1973). The HOG feature descriptor is used for object detection and utilizes the local intensity gradient distributions to identify object edges in the images. In the texture transformation method, the grey level co-occurrence matrix (GLCM) is used to calculate six different statistical attributes to explain the image texture patterns. Four different directions, 0°, 90°, 45°, and 135°, were used to calculate the GLCM matrices. The six attributes derived from the co-occurrence matrix (Aborisade et al., 2014; Haralick et al., 1973) includes

1) **Contrast**, which is a measure of the local intensity variations,

$$Contrast = \sum_i \sum_j |i-j|^2 p(i,j) \quad \text{Eq. 1}$$

2) **Dissimilarity**, which is a localized measure of distance for a pair of pixels,

$$Dissimilarity = \sum_i \sum_j |i-j|\, p(i,j) \quad \text{Eq. 2}$$

3) **Angular Second Moment (ASM)**, which represents the orderliness of each window of the image,

$$ASM = \sum_{i,j} p(i,j)^2 \quad \text{Eq. 3}$$

4) **Energy**, which is the square root of the ASM,

$$Energy = \sqrt{ASM} \quad \text{Eq. 4}$$

5) **Homogeneity**, which represents the local homogeneity within the image by comparing the elements to the diagonal value of the GLCM matrix, and

$$Homogenity = \sum_i \sum_j \frac{1}{1 + |i - j|^2} \, p(i,j) \quad \text{Eq. 5}$$

6) **Correlation**, which is a measure of the linear correlation between the grey-level values of neighbouring pixels.

$$Correlation = \sum_i \sum_j \frac{(i - \mu_i)\,(j - \mu_j)\,p(i,j)}{\sigma_i \sigma_j} \quad \text{Eq. 6}$$

In Eqs. (1) – (6), $p(i,j)$ is the normalized value of the GLCM matrix element at row i and column j, and $\mu_i$and σ_i are mean and variance for each row of the GLCM matrix components.

We constructed five feature sets as potential inputs for the classifier model using color features, HOG features, and texture transformation features. The first set includes all features (color+HOG+texture), the second color and HOG features (color+HOG), the third color and texture features (color+texture), the fourth HOG and texture features (HOG+texture), and the last one only texture (texture) features. Principal Component Analysis (PCA) (Hotelling, 1933) was used to reduce feature set dimensions. PCA uses orthogonal transformations to build components with a linear combination of the original input features to convert a set of possibly correlated features into uncorrelated ones. The principal components (PCs) explaining 95% of the variance in the input data were considered as classifier inputs.

2.4. Classifier Model Construction and Evaluation

Support Vector Machines (SVMs) (Drucker et al., 2002) were used as the classification models. Linear, radial basis function, and second and third-order polynomials, were evaluated as potential kernels for the SVMs. Kernel selection and regularization parameter tuning were carried out using five-fold cross-validation. Accuracy (Guyon and Elisseeff, 2003), recall (Sokolova and Lapalme, 2009), precision (Sokolova and Lapalme, 2009), and Mathew's correlation coefficient (MCC) (Matthews, 1975) were the metrics used for comparing the performance of the classifiers.

3. Results and Discussion

The performance of classification models in predicting the CM content class for the test points is shown in Figure 3. Figure 3 includes a plot of the performance metrics of the classifiers trained using each feature set. The classifiers were trained using the original data set and the augmented data set, and the performance metrics are plotted separately for these classifiers. The plots only include performance metrics calculated using the test data. Figure 3 reveals that the SVM employing the texture transformation features yielded the best performance with an accuracy of 77%, a recall of 92%, a precision of 75%, and an MCC of 0.53. The data augmentation improved the classifier model performance for the ones employing features other than textures transformations. Because texture features, except for those in which the GLCM matrix was calculated in 45° and 135° directions, are obtained using global transformations, their values are both rotation and flip invariant. As a result, the models that employ texture transformation features perform similarly when trained using the original data set or the augmented one.

The performance of classification models trained using PCs is given in Figure 4. The classifiers that employ the texture features had the best performance with an accuracy of

74% and an MCC of 0.51. The classifier model trained only using HOG and texture features for constructing the PCs, eliminating all color features, had the worst performance with recall, precision, and MCC of zero. Data augmentation, in general, improved the performance of the classifiers that used PCs as input sets. However, the performance metrics of the classifier models using PCs as inputs were lower (worse) than those of classifier models built using raw texture, color, and HOG features.

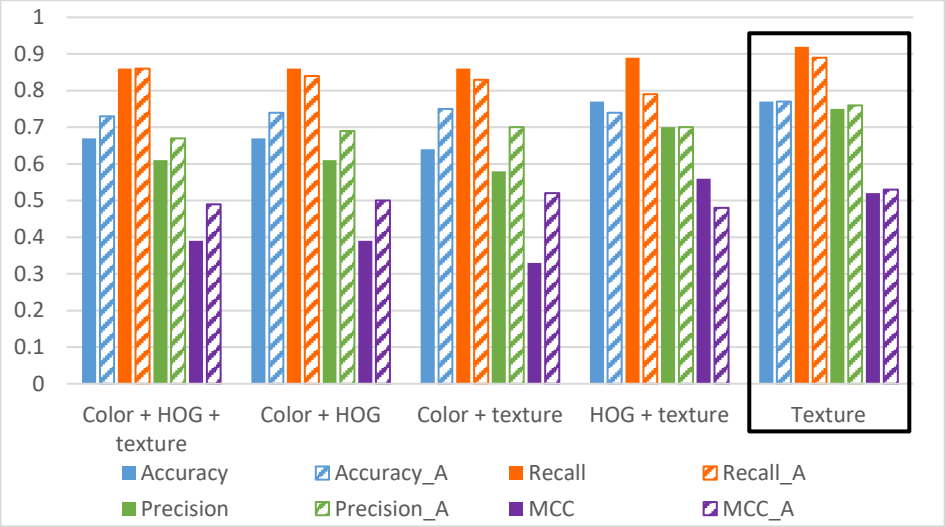


Figure 3. Bar plots of SVM classifier performance metrics trained using different feature sets for the original data set (solid bars) and augmented data set (dashed bars).

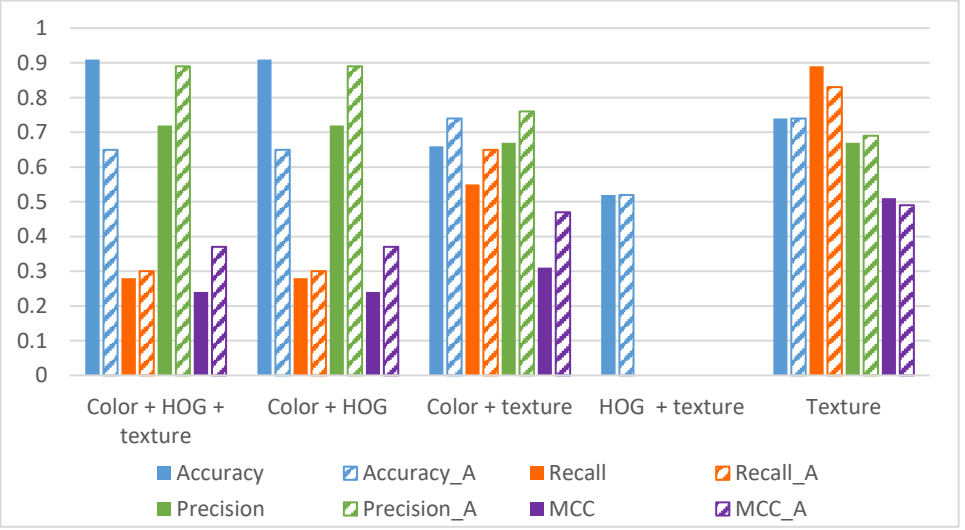


Figure 4. Bar plots of SVM classifier performance metrics trained using PC and different feature sets for the original data set (solid bars) and augmented data set (dashed bars).

4. Conclusions

Imaging is commonly used for tracking human induced pluripotent stem cell (hiPSC) differentiation. Using image-based classification, we built binary classification models to predict *Sufficient*/*Insufficient* classes of cardiomyocyte (CM) content in cells differentiated from hiPSCs. Feature extraction methods were implemented to identify and use the significant features from images to build the classifier. The experimental batched with a CM content above 70% was labeled as the *Sufficient* class, and our classifier was able to predict the classes with 77% accuracy. Future work will include consideration of mixed data from experimental variables and images and consideration of convolutional neural networks as the ML technique to improve the performance of the classifier models.

5. Acknowledgments

This work was funded by NSF grants #1743445 and #2135059.

References

Aborisade, D.O., Ojo, J.A., Amole, A.O., Durodola, A.O., 2014. Comparative Analysis of Textural Features Derived from GLCM for Ultrasound Liver Image Classification. Int. J. Comput. Trends Technol. 11, 239–244. https://doi.org/10.14445/22312803/ijctt-v11p151

Drucker, H., Shahrary, B., Gibbon, D.C., 2002. Support vector machines: Relevance feedback and information retrieval. Inf. Process. Manag. 38, 305–323. https://doi.org/10.1016/S0306-4573(01)00037-1

Freeman, W.T., Roth, M., 1994. Orientation Histograms for Hand Gesture Recognition. Gesture.

Gaspari, E., Franke, A., Robles-Diaz, D., Zweigerdt, R., Roeder, I., Zerjatke, T., Kempf, H., 2018. Paracrine mechanisms in early differentiation of human pluripotent stem cells: Insights from a mathematical model. Stem Cell Res. 32, 1–7. https://doi.org/10.1016/j.scr.2018.07.025

Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182.

Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural Features for Image Classification. IEEE Trans. Syst. Man. Cybern. SMC-3, 610–621.

Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. J. Educ. Psychol. 24, 417.

Kamgoué, A., Ohayon, J., Usson, Y., Riou, L., Tracqui, P., 2009. Quantification of cardiomyocyte contraction based on image correlation analysis. Cytom. Part A 75, 298–308. https://doi.org/10.1002/cyto.a.20700

Kempf, H., Andree, B., Zweigerdt, R., 2016. Large-scale production of human pluripotent stem cell derived cardiomyocytes. Adv. Drug Deliv. Rev. 96, 18–30. https://doi.org/10.1016/j.addr.2015.11.016

Matthews, B.W., 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim. Biophys. Acta - Protein Struct. 405, 442–451. https://doi.org/10.1016/0005-2795(75)90109-9

Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. Inf. Process. Manag. 45, 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Tian, Y., Lipke, E.A., 2020. Microfluidic Production of Cell-Laden Microspheroidal Hydrogels with Different Geometric Shapes. ACS Biomater. Sci. Eng. 6, 6435–6444. https://doi.org/10.1021/acsbiomaterials.0c00980

Vishnoi, V.K., Kumar, K., Kumar, B., 2021. Plant disease detection using computational intelligence and image processing, Journal of Plant Diseases and Protection. Springer Berlin Heidelberg. https://doi.org/10.1007/s41348-020-00368-0

Proceedings of the 14th International Symposium on Process Systems Engineering – PSE 2021+
June 19–23, 2022, Kyoto, Japan
http://dx.doi.org/10.1016/B978-0-323-85159-6.50371-7

Prediction of API concentration using NIRS measured off-line and in-line instruments

Norihiko Fukuoka[a], Sanghong Kim[b*], Takuya Oishi[c], Ken-Ichiro Sotowa[a]

[a]*Department of Chemical Engineering, Kyoto University, Kyoto 615-8510, Japan*
[b]*Department of Applied Physics and Chemical Engineering, Tokyo University of Agriculture and Technology, Tokyo 184-8588, Japan*
[c]*Powrex, Hyogo 664-0837, Japan*

sanghong@go.tuat.ac.jp

Abstract

In the powder mixing process, it is important to uniformly mix components. Since it is difficult to directly measure the concentration in the equipment, a statistical model to predict concentration using NIRS (near infrared spectrum) has been studied in this work. Statistical models can be divided into two types: those constructed from data obtained by the in-line sensor (in-line models), and those constructed from data obtained by the off-line sensor (off-line models). In the off-line data acquisition, the amount of powder used in the experiment can be reduced, and mixing experiments using actual equipment are not required. Thus it would be better if data could be collected off-line. In this study, the prediction accuracies of the two models were compared, and it was found the prediction accuracy of the model tends to be higher when the data was measured in-line. Off-line model cannot predict in-line data because the measurement environment of off-line data is different from that of in-line data. And variable selection was used to improve the in-line prediction accuracy of the off-line model.

Keywords: NIR, Solid dosage forms, Mixing, PLS, PAT, Off-line, In-line

1. Introduction

In the powder mixing process, it is important to ensure the active pharmaceutical ingredients (API) and excipients are sufficiently mixed. However, it is difficult to directly measure API concentration in the mixing equipment. One of the ordinaly method to solve the problem is to construct a statistical model to predict the API concentration from the near infrared spectrum (NIRS). Statistical models can be divided into two types: those constructed from data obtained by the in-line sensor (in-line models), and those constructed from data obtained by the off-line sensor (off-line models). In the off-line data acquisition, the amount of powder and the scale of equipment are smaller than the in-line data acquisition. However, since off-line and in-line data are measured under different conditions, off-line model cannot accurately predict API concentration. There are many researches to construct a prediction model for API concentration in the mixing process as shown in Table 1. However, in previous studies, few studies have compared off-line models with in-line models.

Table 1: Prior studies

Number	Author	Year	Sensor type	Reference
1	Wee Beng Lee et al.	2019	In-line	[W. B. Lee et al. , 2019]
2	Barbara Bakri et al.	2015	In-line	[B. Bakri et al. , 2015]
3	Leonel Quinones et al.	2014	In-line	[L. Quinones et al. 2014]
4	Yleana M. Colon et al.	2014	In-line	[Y. M.Colon et al. , 2014]
5	Aditya U.Vanarase et al.	2013	In-line	[A. U.Vanarase et al. , 2010]
6	Otto Scheibelhofer et al.	2013	Off-line	[O.Scheibelhofer et al., 2013]
7	Sanghong Kim et al.	2011	Off-line	[S. Kim et al. , 2011]
8	Brian M. Zacour et al.	2011	In-line	[B. M. Zacour et al.2011]
9	Aditya U.Vanarase et al.	2010	In-line	[A. U. Vanarase et al., 2013]
10	Otto Berntsson et al.	2002	Off-line	[O. Berntsson et al., 2013]

The prediction accuracy of in-line data tends to be worse because off-line data was taken using an apparatus different from that of in-line data. In order to quantify the effect of the data acquisition method on the prediction accuracyand improve the in-line prediction accuracy of the off-line model, the prediction accuracy of the in-line and the off-lline models are calcurated and wavelength selection are used in this work.

2. Experimental

Mixing experiments were conducted to mix acetaminophen(normal grade; Iwaki Seiyaku Co., Ltd) with lactose(Pharmatose 200M; DFE pharma). In the experiment, API and the excipient were put into the mixer and NIRS were measured through the sapphire window on a side of the equipment while the mixer was in operation. The equipment used in the experiment is shown in Figure 1(Powrex; MG-200 mixer). API concentration was changed from 1% to 40% as shown in Table 2. In the mixing experiment, the center blade rotation speed r_1 and scraper rotation speed r_2 (Figure 2) were set to the values in Table 3. The NIRS are combined with the API concentration calculated from the mass of feed powders. This data is called in-line data. Since mixing is not completed immediately, the spectra

Table 2: API concentration

Experiment number	API concentration [%]
1	1
2	10
3	15
4	20
5	30
6	40

Table 3: Operation condition of the mixer in each experiment

Number	r_1 [rpm]	r_2 [rpm]
1	0	0
2	500	20
3	0	0
4	500	100
5	0	0
6	1500	20
7	0	0
8	1500	100
9	0	0

measured after 150 seconds from the start of mixing operation were used to construct and validate the model. Therefore, the data measured in operation condition 1 and the first half of operation condition 2 were excluded from the analysis.

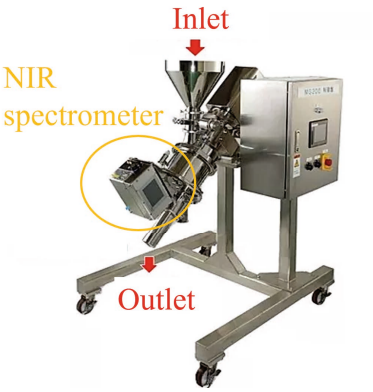


Figure 1: Powder mixer

Figure 2: Inside the mixer

After the powder was removed from the mixer and NIRS of the powder was measured off-line, the API concentration of the powder was measured by high performance liquid chromatography (HPLC). The data which consists of NIRS measured off-line and API concentration measured by HPLC is called off-line data.

3. Model development and validation

The in-line and off-line data were split into two halves respectively, and one half was used for model construction and the other for validation. The measurement condition and usage of the data are shown in Table 4. The prediction models were constructed using partial

Table 4: Data overview

Data number	Sensor type	Usage
1-1	Off-line	Model construction
1-2	Off-line	Model validation
2-1	In-line	Model construction
2-2	In-line	Model validation

least squares (PLS). Standard normal variate was applied to NIRS, and the number of latent variables in PLS was set from 1 to 5. The models were validated with data 1-2 and 2-2 as shown in Table 5.

Correlation coefficients between absorbance and API concentration were calculated using data 1-1 and data 2-1 in order to select variables which have similar relationship between absorbance and API concentration in off-line and in-line. The procedure of variable selection can be divided into two steps. First, To calculate correlation coefficients between

Table 5: Condition of model construction and validation

Case	Model construction	Model validation	Variable selection
1	1-1	1-2	all variables
2	1-1	2-2	all variables
3	2-1	2-2	all variables
4	1-1	2-2	selected variables

absorbance and API concentration on each wavelength using off-line data. Second, To select variables whose absolute value of correlation coefficient is greater than a threshold. The threshold value is 0.98, 0.97, 0.96, 0.93, 0.9, 0.7, 0.5, 0.3, 0.1, 0. The model was constructed using a absorbance at the wavelength selected based on the absolute value of the correlation coefficient between absorbance and API concentration. The result of selected variables that minimizes MAE is case 4 in Table 5.

The accuracy of the models was evaluated by using the mean absolute error (MAE), where N is the number of samples in the validation data, y_n is the API concentration of the nth sample, and $\hat{y}_n$ is the predicted API concentration for the nth sample.

$$\mathrm{MAE} = \frac{1}{N}\sum_{n=1}^{N}|\hat{y}_n - y_n| \tag{1}$$

4. Results

Figure 3 shows MAE obtained through validation for a model constructed using all variables. When the number of latent variables is 3, the MAE is 0.28%, 2.64% and 0.67% for the case 1∼3, respectively. The minimum MAE was obtained for case1, in which

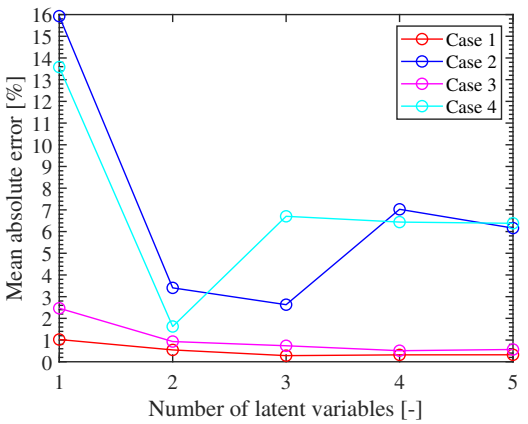


Figure 3: Validation result

off-line data was predicted with a model built with offline data. The MAE of case 2 are significantly larger than those in case 2 and 3.

In this study, the model construction with variable selection was also conducted. Figure 4 shows the relationship between the threshold used for variable selection and the minimum MAE of the model built with the chosen variables. The vertical axis is the MAE computed with the data 2-2, and the horizontal axis is the threshold value of variable selection. In the variable selection, two threshold values of 0.93 and 0.5 are taken as minima.

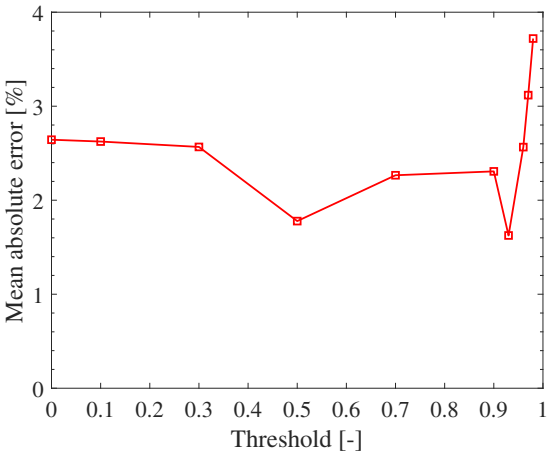


Figure 4: Variable selection result

MAE for the threshold value of 0.93 is smaller than that for the threshold value of 0.5. By comparing this result with case2, the prediction error reduces from 2.64% to 1.63% by using the threshold value of 0.93 for limiting the variables used for model building.

In Figure 4, the MAE is smaller when the thershold is between 0.5-0.93. When it is less than 0.5 the MAE becomes worse because useless input variables are inculded in the model. When the threshold is larger than 0.93 the MAE becomes worse because important input variables are not used for the prediction.

5. Conclusions

In this study, the prediction accuracy of the in-line and the off-line models was discussed. The results showed that there was a large difference in their prediction accuracy. Variable selection successully reduced improving MAE of the off-line model from 2.64% to 1.63%. There could be some improvement in prediction accuracy since the best MAE of 1.63% for the offline model is slightly larger than the MAE of the inline model.

The reason for the difference in prediction accuracy between the two models is that there are some variables that are noisy in the inline measurements and these variables are adopted in the off-line model construction.

References

B. Bakri et al., 2015, European Journal of Pharmaceutical Biopharmaceutical, 97, pp. 78-89

O. Berntsson et al., 2002, Powder Technology, 123, pp. 185-193

Y. M.Colon et al., 2014, International Journal of Pharmaceutics, 417, pp. 32-47

S. Kim et al., 2011, International Journal of Pharmaceutics, 421, pp. 269-274

W. B. Lee et al., 2019, International Journal of Pharmaceutics, 566, pp. 454-462

L. Quinones et al., 2014, AIChE Journal, 60, pp.3123-3131

O. Scheibelhofer et al., 2013, Pharmaceutical Science Technology, 14, pp. 234-243

A. U. Vanarase et al., 2010, Chemical Engineering Science, 65, pp. 5728-5733

A. U. Vanarase et al., 2013, Powder Technology, 241, pp. 263-271

B. M. Zacour, et al., 2011, Journal of Pharmaceutical Innovation, 6, pp. 10-23

AUTHOR INDEX